Investigaciones Geográficas, Boletín del Instituto de Geografía, UNAM
Núm. 51, 2003, pp. 53-72

Evaluación de la confiabilidad temática de mapas o de imágenes clasificadas: una revisión

Jean François Mas*
José Reyes Díaz-Gallegos*
Azucena Pérez Vega*

Recibido: 7 de mayo de 2003
Aceptado en versión final: 11 de febrero de 2003

Resumen. Antes de ser utilizados para tomar decisiones, los mapas temáticos, las bases de datos cartográficos y las imágenes clasificadas, deben ser evaluados para conocer su confiabilidad. Este artículo presenta una revisión de la literatura especializada sobre el proceso de evaluación de la confiabilidad temática y puede ser utilizado como guía práctica para llevar a cabo este tipo de evaluaciones.

Palabras claves: Evaluación de la confiabilidad, cartografía, clasificación de imágenes, SIG.

Assessment of the thematic reliability of maps and classified images: a review

Abstract. Before their use as a decision-making tool, thematic maps, geographical databases and classified images should be assessed for accuracy. This paper presents a review of the specialized literature on thematic accuracy assessment, which can be used as a practical guide for carrying out this type of assessments.

Key words: Reliability assessment, cartography, image classification, GIS

INTRODUCCIÓN

Uno de los insumos más importantes para definir políticas de aprovechamiento y de conservación de los recursos naturales son los mapas temáticos, tales como los mapas de uso del suelo y vegetación y de suelos, entre otros. Estos mapas se presentan en formato impreso o integrados en un sistema de información geográfica (SIG). Durante las tres últimas décadas, se incrementó notablemente el uso de imágenes de satélite para generar este tipo de información (Millington y Alexander, 2000).

En muchos casos se aceptaba que los mapas temáticos eran confiables y no se cuestionaba la calidad de su información. Por ejemplo, las principales bases de datos cartográficas del país no se someten a una evaluación de su confiabilidad. Sin embargo, toda base de datos geográfica presenta un grado de incertidumbre que depende, principalmente, de la calidad de los insumos y de la metodología adoptada para su elaboración. Se pueden producir varios errores en las diferentes etapas del proceso de elaboración de un mapa: *a)* la corrección geométrica de las imágenes; *b)* el análisis de las imágenes, que depende de la experiencia del intérprete, de la calidad de los insumos utilizados (fotografías aéreas, imágenes de satélite, observaciones de campo, entre otros) y del sistema clasificatorio; *c)* la captura (digitalización por ejemplo), y d*)* la representación de los datos en el mapa.

Generalmente se considera que existen dos tipos de error en los mapas o en las imágenes clasificadas (Chrisman, 1989; Goodchild *et al,* 1992; Janssen y Van der Wel, 1994; Pontius, 2000 y 2002; Carmel *et al.,* 2000); los errores temáticos, que se refieren a errores de atributo (etiqueta), y los errores geométricos (de posición) en la delimitación de los polígonos o la ubicación de

*Instituto de Geografía-UNAM, Sede Morelia, Aquiles Serdán 382, Colonia Centro Histórico, CP 58000, Morelia, Michoacán. E-mail: jfmas@igiris.igeograf.unam.mx

los pixeles. Estos dos tipos de error están estrechamente ligados y es difícil separarlos (Chrtsman, 1989). Aspinall y Pearson (1995) distinguen un tercer componente de error potencial en los mapas temáticos, el cual se atribuye a la heterogeneidad dentro de un polígono.

La evaluación de la confiabilidad de los mapas y de las bases digitales geográficas es un tema que está cobrando mucho interés; en gran medida, por el rápido desarrollo y aplicación de los SIG. Cuantificar la confiabilidad de un producto cartográfico, permite a los usuarios del mapa valorar su ajuste con la realidad y, así, asumir el riesgo de tomar decisiones con base en esta información cartográfica; además, ayuda también a conocer y modelar el error que resulte del cruce de varias capas con cierto grado de error en un SIG (Burrough, 1994; Goodchild *et al.,* 1992; Luneta *et al.,* 1991, Walsh *et al.,* 1987).

La evaluación de la confiabilidad temática consiste en comparar la información del mapa con información de referencia consi derada muy confiable. Generalmente se basa en un muestreo de sitios de verificación, cuya clasificación se obtiene a partir de observaciones de campo o del análisis de imágenes más detalladas (con mejor resolución), que aquellas utilizadas para generar el mapa. Por ejemplo, se utilizan fotografías aéreas para verificar mapas generados a partir de imágenes de satélite de alta resolución como Landsat o SPOT (Peralta-Higuera ef *al.,* 2001; Mas *et al.,* 2001; Vogelmann *et al.,* 2001).

La confrontación entre las clases cartografiadas y las clases determinadas en las fotografías aéreas o en el campo para los sitios de verificación se basa en el supuesto de que la información de referencia es altamente confiable y representa "la verdad"; por lo que esta confrontación permite evaluar la confiabilidad del mapa y conocer las confusiones que presenta (Congalton y Green, 1993). El proceso de evaluación de la confiabilidad temática se divide en tres etapas (Stehman y Czaplewski,1998).

a) El diseño del muestreo que consiste en la selección de las unidades de muestreo.

b) La evaluación del sitio de verificación, que permite obtener la clase correspondiente a cada unidad de muestreo.

c) El análisis de los datos, que consiste generalmente en la elaboración de una matriz de confusión y et cálculo de índices de confiabilidad.

Existe un gran número de publicaciones sobre la evaluación de la confiabilidad de mapas temáticos o de imágenes clasificadas en revistas especializadas en percepción remota y sistemas de información geográfica, muchas veces expresando opiniones contradictorias. En este artículo se hace una revisión de esta literatura para proporcionar, a los lectores y especialistas en cartografía, un panorama sintético de la aplicación e importancia de la evaluación de la confiabilidad de mapas, que pueda servirles de guía práctica para llevar a cabo este tipo de análisis. La primera parte del artículo se organiza con base en las tres etapas de la evaluación descritas por Stehman y Czaplewski (1998); *a)* diseño del muestreo, *b)* evaluación de los sitios de verificación y *c)* análisis de los datos. Luego, se discuten algunos aspectos operativos de la evaluación. Se puede encontrar material adicional sobre este tema a través de Internet, en la dirección http://indy2.igeograf.unam.mx/dote/confiabilidad.htm.

DISEÑO DEL MUESTREO

El diseño de muestreo contempla la determinación del tipo de unidades de muestreo, del método de selección de las mismas, así como del número de unidades de muestreo

necesarias (tamaño de muestra).

Las unidades de muestreo

La unidad de muestreo permite relacionar la localización de la información del mapa y del terreno. Puede ser un punto, un píxel, un grupo de píxeles, un polígono del mapa o una unidad de superficie con formas predeterminadas, por ejemplo, un cuadro o un círculo de una hectárea. No existe un consenso definitivo sobre la unidad de muestreo más adecuada (Chuvieco, 1996); su elección depende en mucho de los objetivos de la evaluación, del proceso de mapeo, de la estructura del paisaje y de las categorías que más le interesan al usuario. Si la unidad de muestreo es un punto, se compara la clasificación del mapa con relación a este punto con la misma localización en el terreno; en la práctica, lo que se evalúa es una superficie alrededor del punto. Janssen y Van der Wel (1994) recomiendan el uso de pixeles individuales como unidades de muestreo para las clasificaciones digitales píxel a píxel.

En el caso de mapas en formato vectorial, el uso de los polígonos como unidades de muestreo permite una correspondencia directa entre éstas y el mapa. Sin embargo, al modificar el mapa (actualización o agregación de clases de un sistema clasificatorio jerárquico) esta correspondencia desaparece.

En el caso de unidades de superficie predeterminadas, la superficie que debe cubrir el sitio de muestreo es también delicada de determinar, un sitio de verificación grande puede incluir varias porciones de polígonos en el mapa y varios tipos de cubierta en el terreno o en la imagen de referencia, lo que genera ambigüedades al confrontar la información del sitio de verificación con la del mapa. Al contrario, un sitio de verificación pequeño puede coincidir con una unidad del paisaje no representada en el mapa por ser más pequeña que el mínimo cartografiable del mismo (Figura 1).

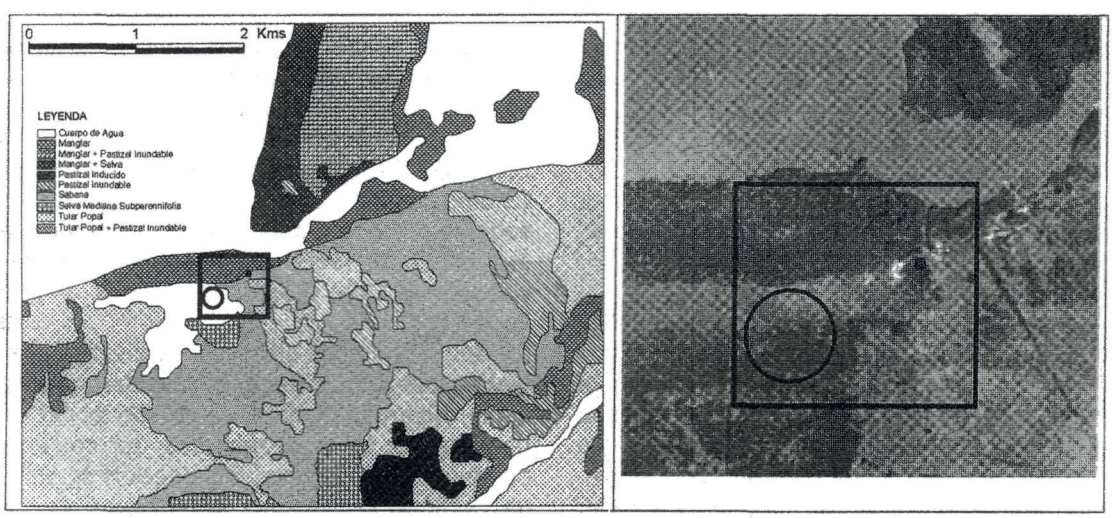


Figura 1. Unidades de muestreo puntual de diferentes superficies y formas.

Durante el análisis de la unidad de muestreo, se sugiere tener en cuenta su entorno.

El método de muestreo

Sirve para seleccionar una pequeña muestra del área cartografiada, de tal forma que sea representativa de la totalidad del mapa. En un diseño de muestreo probabilista, todas las unidades de muestreo presentes en el mapa tienen una probabilidad conocida superior a cero de ser seleccionadas, a esta probabilidad de selección se le denomina probabilidad de inclusión. Asimismo, durante la selección de las unidades de muestreo no se deben descartar sitios que presenten ciertas características; como por ejemplo, estar localizados en regiones con poca accesibilidad o en terrenos privados. Las técnicas de muestreo más empleadas en el proceso de evaluación de la conftabilidad temática son: aleatorio simple, aleatorio estratificado, sistemático, sistemático no alineado y por conglomerados (Figura 2).

Aleatorio simple. Los sitios de verificación se eligen de tal forma, que todos tienen la misma probabilidad de ser seleccionados. El problema con este tipo de muestreo es que las categorías del mapa que presentan una superficie reducida son muy poco representadas o inclusive ausente de la muestra. Esta selección genera sitios de muestreo dispersos en todo el territorio, lo que implica asumir los costos de traslado (Congalton, 1988b, Fitzpatrick-Lins, 1981).

Aleatorio estratificado. La muestra se realiza dividiendo a la población en estratos, con base en una variable auxiliar (altitud, región ecológica, división administrativa, facilidad de acceso, clase en el mapa, entre otros), lo que permite tener cierto control sobre la distribución de los sitios de muestreo y obtener información sobre subconjuntos de la población.

Sistemático. La muestra se distribuye a intervalos regulares a partir de un punto seleccionado de manera aleatoria, pero puede originar algún error cuando existe algún patrón periódico en el área estudiada (Chuviéco, 1996).

Sistemático no alineado. La muestra se distribuye de manera regular, pero con un cierto grado de libertad y permite representar todo el territorio.

Por conglomerados. Se selecciona un sitio aleatoriamente y se toman varias muestras vecinas de acuerdo con un esquema predeteminado. Por ejemplo, se seleccionan otros dos sitios, siguiendo una forma de L a cierta distancia del sitio seleccionado aleatoriamente (Figura 2).

Los muéstreos aleatorio simple, sistemático, sistemático no alineado y por conglomerados son probabilistas y resultan en probabilidades de inclusión, iguales para todas las unidades de muestreo. Los muéstreos estratificados, como el estratificado aleatorio, con un número igual de unidades de muestreo por estrato, conduce a probabilidades de inclusión diferentes según el estrato. Eso no constituye ningún problema en el análisis de los resultados, siempre y cuando estas probabilidades de inclusión sean conocidas y utilizadas para ponderar las observaciones derivadas de cada estrato (Stehman, 2000). Estudios comparativos de estos diferentes esquemas de muestreo pueden encontrarse en Congalton (1988b); Fítzpatrick-Lins (1981) y Stehman (1992 y 1999).

Existen numerosos ejemplos de diseños sesgados que no se pueden considerar como estadísticamente robustos, debido a que la muestra no es representativa del conjunto del mapa. Por ejemplo, la selección de sitios de verificación ubicados en el centro de los polígonos de los mapas conduce a una evaluación optimista de la confiabilidad del mapa, ya que los errores son más frecuentes en las zonas de transición entre diferentes

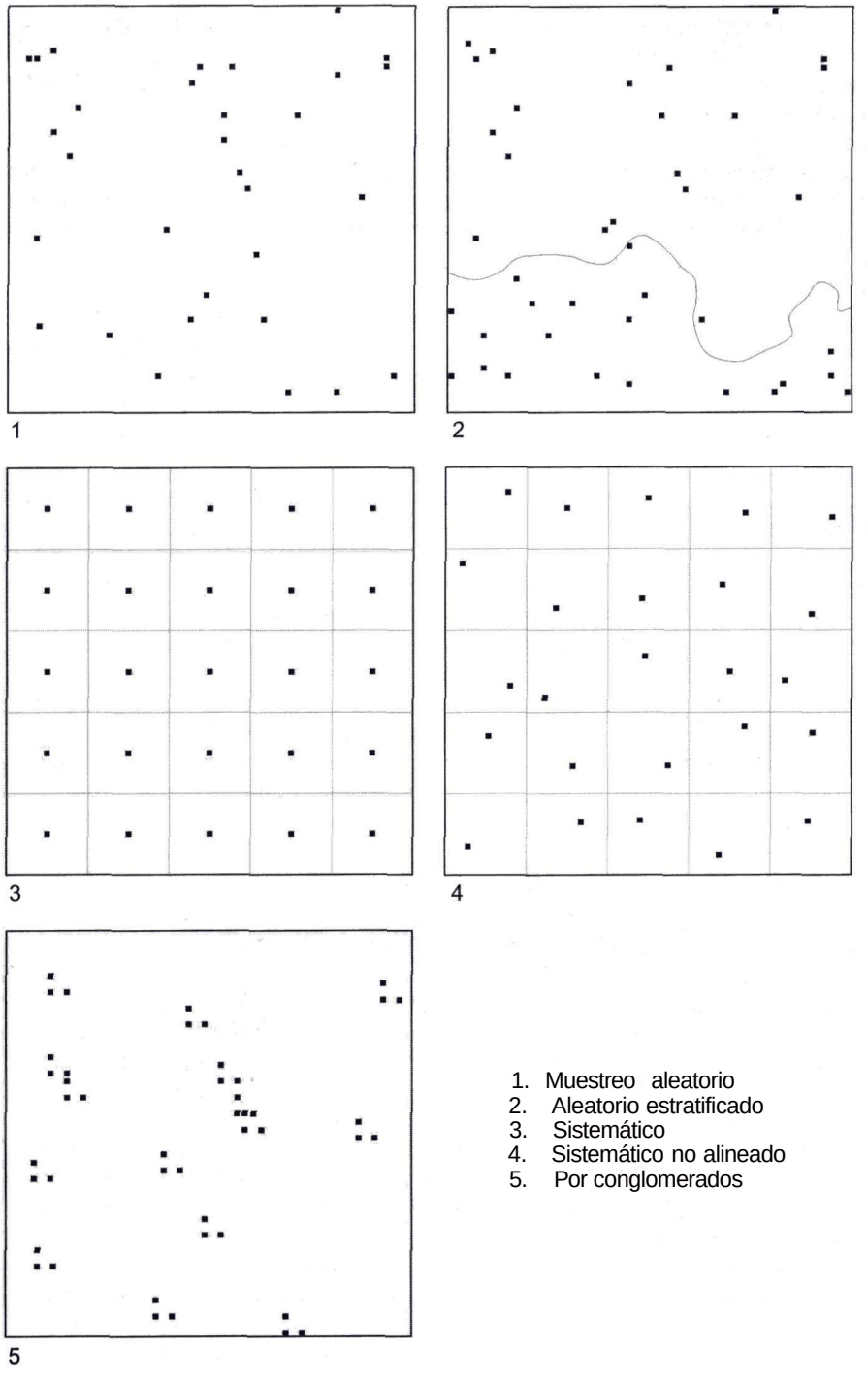


Figura 2. Esquemas de muestreo probabilistas más aplicados en la evaluación de la confíabilidad de mapas temáticos (modificado de Chuvieco, 1996).

tipos de cobertura (límites entre polígonos; Hammond y Verbyla, 1996). De la misma manera, la selección de sitios de muestreo, ubicados cerca de carreteras para facilitar el acceso durante la verificación de campo, tiende a seleccionar sitios de verificación localizados en regiones mejor conocidas {por ejemplo, por los foto-intérpretes y, en consecuencia, mejor interpretadas) y por lo regular corresponde a tipos de vegetación más perturbados. Otro ejemplo de evaluación sesgada, es la utilización de los campos de entrenamiento o de pixeles correlacionados con estos campos para evaluar la confiabilidad de clasificaciones digitales (Friedl *et al.,* 2000; Stehman y Czaplewski, 1998).

El tamaño de la muestra

El tamaño de la muestra se refiere al número de sitios de verificación utilizados para estimar la confiabilidad del mapa. Entre más grande sea el tamaño de la muestra, más precisa será la evaluación; sin embargo, por razones de costo y tiempo, es conveniente determinar el tamaño de muestra mínimo, para alcanzar los objetivos de la evaluación.

Congalton (1988b) sugiere muestrear una superficie aproximada al 1% de la superficie cartografiada. En otra publicación (1991), el mismo autor recomienda verificar por lo menos 50 sitios por categoría, y de 75 a 100 si el área en estudio es superior a 400 000 ha o si hay más de 12 categorías.

La confiabilidad pues la proporción de sitios de verificación correctamente identificados en el mapa. En estadística tradicional la desviación estándar de la estimación de una proporción depende del tamaño de la muestra, del tamaño de la población estudiada y de la proporción (Cochran, 1980; Wonnacott y Wonnacott, 1991; Stehman, 2001; ecuación 1).

$$\sigma_p = \sqrt{\frac{N-n}{N}}\sqrt{\frac{p(1-p)}{n}} \qquad (1)$$

donde CT_P es la desviación estándar de la estimación de la confiabilidad, N es el tamaño de la población, *n* es el tamaño de la muestra (número de unidades de muestreo) y p la contabilidad de la muestra.

En la Figura 3 se representa el tamaño de la muestra *n* en función del tamaño de la población *N* con base en la ecuación (1) para el caso en el cual la confabilidad p es de 0.5 (la mitad de los sitios de verificación está correctamente identificada en el mapa) y la desviación estándar σ_p es de 0.05 (es decir se evalúa la confiabilidad del mapa con un error razonablemente pequeño). Se puede observar que el tamaño de muestra necesario aumenta con el tamaño de la población y alcanza un máximo de 100 para una población de 10 000 aproximadamente. Las poblaciones superiores a 10 000 no necesitan un tamaño de muestra más importante para alcanzar una estimación de la confiabilidad con una desviación estándar inferior o igual a 0.05.

En la evaluación de la confiabilidad de mapas o de imágenes de satélite clasificadas se manejan, generalmente, poblaciones muy grandes; por ejemplo, un mapa escala 1:250 000 del INEGI, cubre una superficie de más de 20 000 km^2 y contiene, por lo tanto, más de 2 000 000 de unidades de muestreo de 100 x 100 m. Una imagen de satélite Landsat TM tiene decenas de millones de pixeles. Como se mostró en la Figura 3, las poblaciones grandes no necesitan un tamaño de muestra más grande para obtener una evaluación de la confiabilidad precisa. En consecuencia, no es necesario que el tamaño de la muestra sea un porcentaje de la población total como lo proponen Congalton (1988b) y Stehman (1997).

La aproximación normal permite determinar que tanto la confiabilidad *p* medida en la muestra permite una estimación precisa de la confiabilidad del mapa *P.* Con base en esta aproximación, se puede emplear la ecuación siguiente, derivada de la ecuación (1), para relacionar la confiabilidad p, la precisión con la cual *p* estima la confiabilidad del mapa (medio-intervalo de confianza *) y el tamaño de la muestra y de la población (Cochran, 1980; Wannacott y Wannacott, 1991).

$$\delta = t\sigma_p = \sqrt{\frac{N-n}{N}}\sqrt{\frac{p(1-p)}{n}} \qquad (2)$$

donde *p* es la confiabilidad, * el error (medio intervalo de confianza) y *t* = 1.96 para α = 0.05 (en otras palabras, la probabilidad de que el valor real de la confiabilidad del mapa *P* esté fuera del intervalo de confianza es de 5%), *n* es el número de unidades de muestreo y *N* el tamaño de la población.

Para fines prácticos, cuando *N* es grande, una primera aproximación de * y de *n* es como sigue (Cochran, 1980; Fitzpatrick-Lins, 1981;DicksyLo, 1990):

$$\delta = t\sqrt{\frac{p(1-p)}{n}}$$

que es equivalente a:

$$n = \frac{t^2 p(1-p)}{\delta^2} \qquad (3a \text{ y } 3b)$$

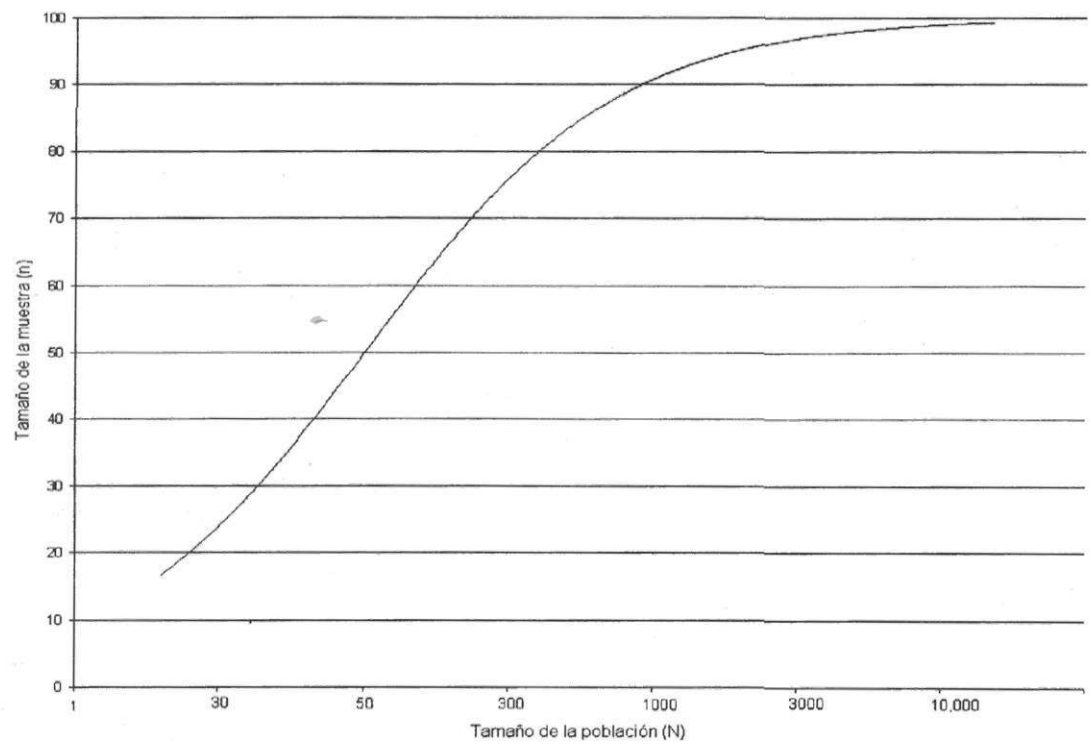


Figura 3. Tamaño de muestra necesario para alcanzar un error estándar de 0.05 en función del tamaño de la población (tomado de Stehman, 2001). El tamaño de la población está representado en escala logarítmica $\log_{10}(N)$.

Como se puede observar en la ecuación (3b), el intervalo de confianza del estimado de la confiabilidad depende de:

- *El número de unidades de muestreo.* Entre más grande sea el tamaño de la muestra, más pequeño es el intervalo de confianza; por lo tanto, más precisa es la estimación de la confiabilidad; por ejemplo, con 50 unidades de muestreo, una confiabilidad de 80% (p=0.8) presenta un medio intervalo de confianza * de 11%, es decir que, en realidad, la confiabilidad puede variar entre 69 y 91%, pero con 250 unidades de muestreo el intervalo de confianza se reduce a ±5%.

- *La confiabilidad.* Con el mismo número de unidades de muestreo, la confiabilidad se estima con menos precisión si se acerca a 50%. Por ejemplo, con 100 unidades de muestreo, una confiabilidad de 50% tiene un intervalo de confianza de ± 9.8% y una de 90%, ± 5.9%.

La tabla presentada a continuación indica el tamaño de muestra necesario para diferentes valores de confiabilidad y precisión de la evaluación.

Tabla 1. Tamaño de la muestra por clase en función de p y d

	p				
*	90%	80%	70%	60%	50%
2.5%	553	983	1291	1475	1537
5.0%	138	246	323	369	384
10.0%	35	61	81	92	96

P : confiabilidad estimada de la clase
* : medio intervalo de confianza

Para los muéstreos aleatorios estratificados con un número de unidades igual y pequeño en cada estrato, la ecuación (3a) no aplica cuando se calcula el intervalo para toda la población, reagrupando los estratos (Stehman, 2000). La autocorrelación espacial, que se puede definir por el hecho de que el error no se distribuye de manera homogénea en el mapa, pero que tiende a agregarse, también afecta las estimaciones del intervalo de confianza de la confiabilidad (pero no influye en la confiabilidad). El tipo de muestreo más afectado es por conglomerados, los muéstreos sistemáticos y aleatorios estratificados son los menos sensibles. En este tipo de muéstreos se utilizan las mismas ecuaciones para el cálculo de la confiabilidad y la precisión de la estimación, pero se debe tomar en cuenta que el intervalo de confianza es en realidad más grande que el calculado (Stehman, 2000).

LA EVALUACIÓN DE LOS SITIOS DE VERIFICACIÓN

Este paso consiste en la caracterización del sitio de verificación para asociarlo a una o varias clases de la leyenda del mapa que se evalúa. En la práctica, la evaluación de la unidad de muestreo, en particular si es un punto o un pixel, se lleva a cabo con base en el análisis de una cierta área alrededor del mismo.

Comúnmente, esta evaluación conduce a asociar el sitio de verificación a una sola categoría de la leyenda del mapa. Sin embargo, no es siempre posible ni conveniente limitarse a una clase única para caracterizar el sitio de verificación, porque este ejercicio puede ser muy subjetivo (Hord y Brooner, 1976). Esta subjetividad se debe a que el sitio puede localizarse en una zona de transición progresiva entre dos tipos de vegetación (por ejemplo, un bosque de pino y un bosque de pino-encino) o en un área fragmentada donde se encuentran varias clases. Puede también corresponder a un estadio de transición temporal entre tipos de vegetación (por ejemplo, la vegetación se-

cundaria, particularmente en los trópicos). Otra fuente de ambigüedad son los errores en la localización del sitio de verificación en el mapa (Khorram *et al.,* 2000).

Todas estas fuentes de errores llevan generalmente a subestimar la confiabilidad del mapa y varios autores han propuesto diversos mecanismos para aminorarlos. Khorram *et al.* (2000) caracterizan el sitio de verificación con una clase principal y una adicional. En la confrontación entre el mapa y la información de referencia; en caso de que la clase principal no corresponda con el mapa se da una "segunda oportunidad" con la clase adicional. Otros autores caracterizan de manera cuantitativa el sitio de verificación, tanto en el mapa como en la información de referencia, utilizando las proporciones de la superficie representada por cada clase de cubierta.

Woodcock y Gopal (2000) utilizan un enfoque difuso para calificar los sitios de verificación. En este enfoque, la pertenencia de un elemento a una clase se expresa a través de un grado de pertenencia, que es una varíable que toma cualquier valor entre 0 y 1 para expresar la pertenencia parcial a diferentes conjuntos. Por ejemplo, en el caso de dos categorías basadas en la cubierta de la copa de los árboles (vegetación "abierta" si la cubierta es entre 10 y 40%; vegetación "cerrada" cuando la cubierta de las copas es superior a 40%), un sitio de verificación, con 40% de cubierta, presenta 0.5 de pertenencia en ambas categorías (cerrada y abierta). En el enfoque booleano, el intérprete tendría que clasificarlo en una de las dos categorías. Woodcock y Gopal (2000) desarrollaron un método de caracterización del sitio de verificación basado en una escala lingüística que asocia cada sitio con una categoría a través de una calificación que expresa la adecuación de la clase con el sitio, como por ejemplo, "esta categoría define perfectamente lo que se observa en la fotografía", "esta categoría no es la más adecuada para definir lo que se observa en la fotografía, pero es aceptable", "esta categoría no define correctamente la fotografía" o bien "esta categoría no puede aplicarse a lo que se observa en la fotografía". Cada expresión se asocia con un grado de pertenencia. Estos autores proponen un método para el cálculo de índices de confiabilidad utilizando este enfoque difuso (Gopal y Woodcock, 1994; Gopal *et al.,*1999).

ANÁLISIS DE LOS DATOS

El análisis de los datos de confiabilidad se hace generalmente a través de una matriz de confusión, que permite confrontar la información de los sitios de verificación con aquella de la base cartográfica que se pretende evaluar. En la matriz de confusión, las filas representan generalmente las clases de referencia y las columnas las clases del mapa. La diagonal de la matriz expresa el número de sitios de verificación para los cuales hay concordancia entre el mapa y los datos de referencia, mientras los marginales indican errores de asignación.

La proporción de puntos correctamente asignados (diagonal) expresa la confiabilidad del mapa. Se distinguen dos tipos de error según si la lectura de la matriz se hace con base en las líneas o en las columnas. El error de comisión representa la proporción de sitios de verificación cartografíada en una cierta clase, pero que en realidad pertenece a otra categoría. El error de omisión se refie re a la proporción de sitios de verificación correspondiente a una categoría que fue cartografiada en otra (Aronoff, 1982; Chuvieco, 1996).

Congalton (1991) sugiere llevar a cabo una normalización o estandarización de la matriz de confusión. Ésta se realiza llevando a cabo un ajuste iterativo para que las columnas y las filas de la matriz sumen 1 y argumenta que esta normalización permite ajustar los valores de la matriz tomando en cuenta las

filas y las columnas. Este proceso modifica los valores de la diagonal y permite calcular un valor de confiabilidad global normalizado que representa mejor la calidad del mapa, porque incorpora información de los elementos fuera de la diagonal. Según Congalton, la normalización de la matriz permite una comparación más objetiva entre matrices derivadas de diferentes procesos de evaluación. Sin embargo, Stehman (1997) está en desacuerdo con esta normalización, porque no permite obtener estimaciones consistentes. Se entiende por consistencia que la estimación de algún parámetro con base en una muestra se vuelve exactamente igual al valor del parámetro en la población cuando la muestra es igual a la población (n = N). Stehman (1997) demuestra que las estimaciones de confiabilidad derivadas de una matriz normalizada no son consistentes con las derivadas de una matriz basada en toda la población y recomienda no usar matrices normalizadas.

Con base en la matriz de confusión, se desarrollaron varios índices de confiabilidad (Stehman y Czaplewski, 1998) que se describen a continuación:

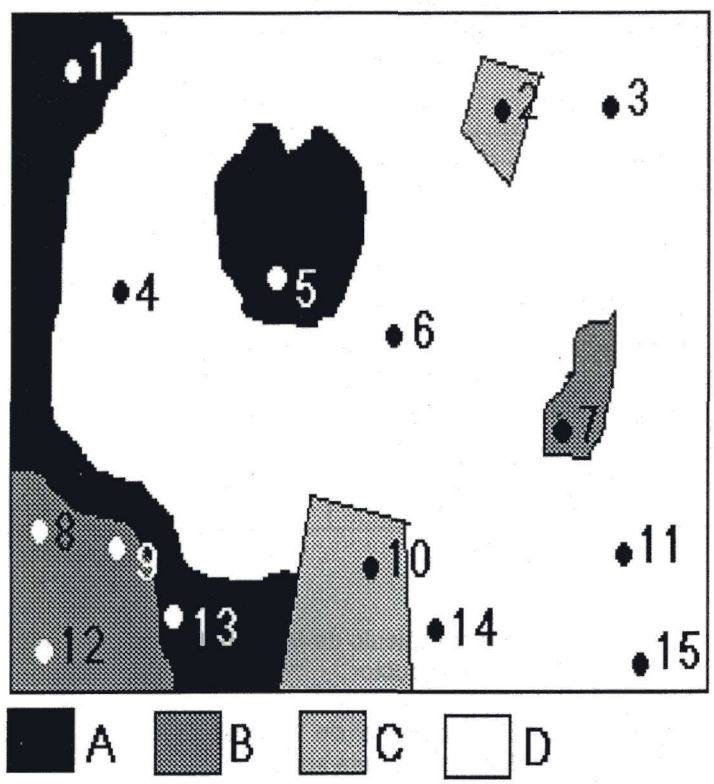
Si el diseño de muestreo es aleatorio simple, la confiabilidad global {proporción de los sitios de verificación correctamente clasificados en la base cartográfica) se obtiene dividiendo los elementos de la diagonal con el número total de sitios de verificación. Este valor representa la probabilidad para cualquier sitio en el mapa de ser correctamente clasificado. En el caso de muestreo aleatorio estratificado, el número de sitios por categoría ya no es proporcional a la superficie cubierta por cada categoría y el valor obtenido no se puede interpretar de esta manera. Por ello se tienen que hacer las correcciones necesarias con base en las probabilidades de inclusión del muestreo (Card, 1982).

Existen también índices que dan cuenta de la confiabilidad de cada una de las clases de la leyenda como: *a)* la confiabilidad del usuario, que puede interpretarse como la probabilidad que un sitio clasificado como *A* y aleatoriamente seleccionado sea realmente *A* en el terreno, y *b)* la confiabilidad del productor, que es la proporción de sitios de verificación de la clase A que están representados en el mapa o en la base de datos como tal (Janssen y van der Wel, 1994). Los valores de confiabilidad del usuario y del productor están relacionadas respectivamente con los errores de comisión y omisión. Aronoff (1982) define el riesgo del productor como el hecho de rechazar un mapa aceptable y el riesgo del usuario, aceptar un mapa no confiable.

La confiabilidad promedio es el promedio de los valores de confianza de cada categoría. Es, por lo tanto, un índice de confiabilidad que da el mismo peso a todas las categorías independientemente de la superficie que cubren.

A continuación se presenta un mapa para el cual se verificaron en campo 15 sitios de verificación seleccionados con base en un muestreo aleatorio (Figura 4). La comparación entre los mapas y la información de campo permitió elaborar la matriz de confusión y calcular los índices de confiabilidad (Tabla 2), en este ejemplo, la confiabilidad global es de 67% (10 sitios correctamente identificados de un total de 15).

Los índices de confiabilidad descritos anteriormente no toman en cuenta los elementos fuera de la diagonal de la matriz (Rosenfield y Fitzpatrick-Lins, 1986). Por esta razón, se generalizó el uso del coeficiente de Kappa, que utiliza las sumas marginales de la matriz y da cuenta de la contribución del azar en la confiabilidad del mapa. Para los cálculos que se describen a continuación, es conveniente expresar los valores de la matriz en proporción del número total de sitios (Tabla 3).



Clases determinadas para los 15 sitios de verificación.

1: A	9: A
2: C	10: C
3: D	11: D
4: C	12: B
5: A	13: B
6: D	14: C
7: A	15: D
8: B	

Figura 4. Mapa temático (4 clases A, B, C y D) y sitios de verificación.

Tabla 2. Matriz de confusión derivada de la figura 4

	Mapa					
Ref.	A	B	C	D	*Error Omis.*	*Conf.prod.*
A	2	2	0	0	*50.0*	*50.0*
B	1	2	0	0	*33.3*	*66.7*
C	0	0	2	2	*50.0*	*50.0*
D	0	0	0	4	*0.0*	*100.0*
Error Com. %	*33.3*	50.0	*0.0*	*33.3*		
Conf.us.	*66.7*	*50.0*	*100.0*	*66.7*		

Error Omis.: error de omisión, Error Com.: error de comisión, Conf. prod. confiabilidad del productor, Conf. us.: confiabilidad del usuario.

Tabla 3. Matriz de confusión expresada en proporción. Las mismas anotaciones se utilizan en las ecuaciones a continuación

	Mapa				
Referencia	1	2	...	q	Total
1	p_{11}	p_{12}	...	p_{1q}	p_{1+}
2	p_{21}	p_{22}	...	p_{2q}	p_{2+}
...	...	...	...	...	...
q	p_{q1}	p_{q2}	...	p_{qq}	p_{q+}
Total	p_{+1}	p_{+2}	...	p_{+q}	

$$K = \frac{P_o - P_c}{1 - P_c} \qquad (4)$$

donde K es el índice de Kappa, P_o la proporción de área correctamente clasificada (confiabilidad global) y P_c la confiabilidad resultante del azar.

P_o se obtiene sumando los elementos de la diagonal:

$$P_o = \sum_{k=1}^{q} p_{kk} \qquad (5)$$

P_c se calcula sumando el producto de las sumas marginales :

$$P_c = \sum_{k=1}^{q} p_{k+} p_{+k} \qquad (6)$$

El coeficiente de Kappa se puede calcular para cada categoría, con base en las filas o en las columnas de la matriz, de manera similar al cálculo de la confiabilidad del usuario y del productor.

El coeficiente de Kappa para la categoría (fila) /se calcula según;

$$K_i = \frac{p_{ii} - p_{i+}p_{+i}}{p_{i+} - p_{i+}p_{+i}} \qquad (7)$$

El coeficiente de Kappa para la categoría (columna); se obtiene según:

$$K_j = \frac{p_{jj} - p_{j+}p_{+j}}{p_{j+} - p_{j+}p_{+j}} \qquad (8)$$

En la Tabla 4 se presentan la matriz expresada en proporción y los valores de estos índices de confiabilidad calculados a partir de la matriz presentada en la Tabla 2.

La suma de la diagonal que expresa la confiabilidad global del mapa (proporción de sitios correctamente clasificados) es igual a 0.67 (67%). Para calcular el valor del coefi-

cíente de Kappa, se evalúa primero la aportación del azar P_c que es igual a la suma de los productos de las sumas marginales de la matriz. Para la matriz de la tabla 4, P_c = 0.2 * 0.27 + 0.27 * 0.20 + 0.13 * 0.27 + 0.4 * 0.27 = 0.25 y el coeficiente de Kappa K - (0.67-0.25)/(1-0.25) = 0.56.

Un coeficiente de Kappa de 0.56 significa que la clasificación es 56% mejor que la confiabilidad esperada, asignando aleatoriamente una categoría de cobertura a los polígonos (Dicks y Lo, 1990). Muchos autores no están de acuerdo con la manera de estimar la contribución del azar y existen algunas variantes en el cálculo del índice de Kappa, que utilizan otras formas de evaluarla (Foody, 1992; Ma y Redmond, 1995; Stehman, 1997). El coeficiente de Kappa es válido para un muestreo aleatorio simple (Congalton, 1988b). Stehman (1996) propone un método para calcular el valor de Kappa y su varianza para los muestreos aleatorios estratificados. El índice de Kappa resta la contribución teórica del azar a la confiabilidad y da un resultado siempre inferior o igual a la confiabilidad global. Sin embargo, para el usuario de un mapa no importa que cierta proporción del mapa esté correcta debido al azar (Turk, 1979). Por tanto, no es lógico restar la contribución del azar a la calidad del mapa, por lo cual Stehman (1997) considera que el índice de Kappa no es un buen índice de la calidad de un mapa y es más apropiado cuando se pretende evaluar diferentes métodos de clasificación.

Todos los índices descritos anteriormente, menos los derivados del enfoque difuso, consideran de la misma manera todos los errores. Sin embargo, ciertos errores son más perjudiciales que otros según el objetivo de esfuerzo cartográfico; por ejemplo, si se utiliza un mapa de uso del suelo y vegetación para calcular la superficie boscosa de una región, la confusión entre bosque de encino y bosque mesófilo es menos grave que entre bosque de encino y pastizal. Si el objetivo es cartografiar el hábitat de un organismo que se circunscribe al bosque mesófilo, este mismo error se vuelve más crítico. Naesset (1996) desarrolla índices que toman en cuenta la gravedad del error en el cálculo de los índices de confiabilidad.

Tabla 4. Matriz de confusión expresada en proporción derivada de la Tabla 2

	Mapa						
Ref.	A	B	C	D	*Total*	*Conf.prod.*	Kappa
A	*0.13*	0.13	0.00	0.00	*0.27*	*0.50*	0.38
B	0.07	*0.13*	0.00	0.00	*0.20*	*0.67*	0.58
C	0.00	0.00	*0.13*	0.13	*0.27*	*0.50*	0.38
D	0.00	0.00	0.00	*0.27*	*0.27*	*1.00*	1.00
Total	*0-20*	*0.27*	*0.13*	*0.40*			
Conf.us.o	*0.67*	*0.50*	*1.00*	*0.67*			
Kappa	*0.55*	*0.32*	*1.00*	*0.55*			

Conf. prod.: confiabilidad del productor; Conf. us.: confiabilidad del usuario.

DISCUSIÓN

Se presentaron los métodos para llevar a cabo la evaluación de la confiabilidad de mapas o de imágenes clasificadas. A continuación se discuten algunos aspectos prácticos de estas evaluaciones, como la conciliación de los requisitos estadísticos y consideraciones logísticas, así como el problema de la veracidad de los datos de referencia. Por último, se aborda el caso de la evaluación de mapas de cambio.

Algunos mecanismos para reducir costos

En la evaluación de la confiabilidad cartográfica, un problema crucial es la conciliación de los requisitos estadísticos, que permiten una evaluación objetiva y científicamente defendible con las consideraciones logísticas que toman en cuenta los problemas de costo y de acceso para recolectar la información de los sitios de verificación. Existen diferentes mecanismos que permiten aminorar los problemas prácticos relacionados con el número y la selección de los sitios de verificación sin perder rigor estadístico.

La estratificación consiste en repartir el esfuerzo de muestreo con base en estratos geográficos como regiones ecológicas o entidades administrativas, entre otros. Permite distribuir el esfuerzo de muestreo, con probabilidades de muestreo más bajas, por ejemplo, en regiones menos accesibles. La estratificación con base en las categorías del mapa permite garantizar que no se sobremuestreen las categorías con mayor extensión a expensas de las de menor superficie (Card, 1982; Figura 5).

La evaluación de la confiabilidad se puede llevar a cabo para una parte del área cartografiada con base en la hipótesis de que ésta es representativa del conjunto del mapa. Por ejemplo, el mapa de uso del suelo y vegetación de un estado de la República se podría evaluar únicamente en algunos municipios que representan la variedad de tipos de vegetación que se pueden encontrar en el estado y los resultados se generalizan para toda la entidad. Sin embargo, es necesario estar seguro que otros parámetros no influyeron en la calidad del mapa para estos municipios y que realmente los resultados obtenidos son extrapolares a todo el esta-

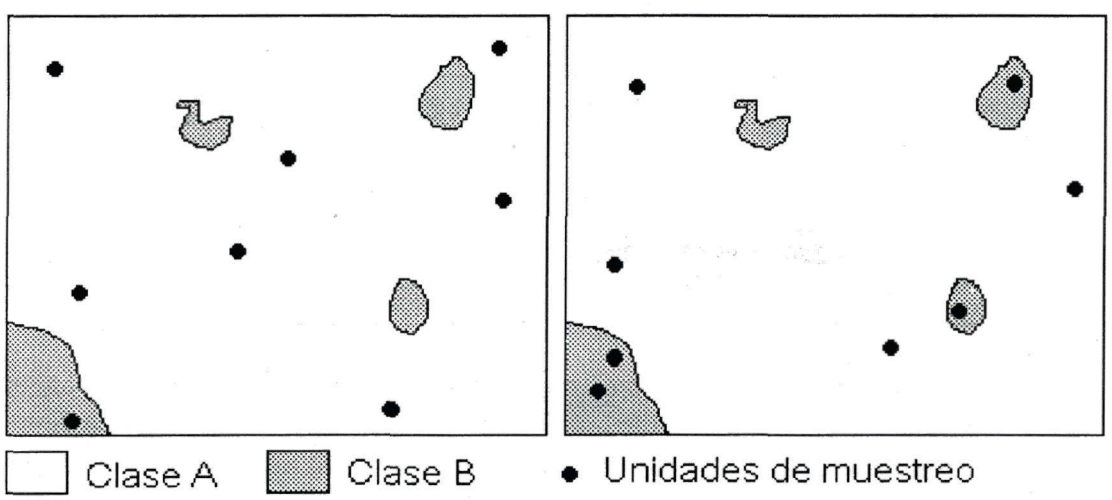


Figura 5. Muestreo aleatorio simple (ocho sitios al azar) y aleatorio estratificado (cuatro sitios al azar por categoría). Este último permite una mejor repartición de los sitios de verificación entre las diferentes categorías del mapa.

do. Otro ejemplo de evaluación basada en el supuesto anterior es la utilización de líneas de vuelo de fotografías aéreas o de videografía que cruzan el área cartografiada y conforman una cobertura parcial de esta última, como fue el caso del Inventario Forestal Nacional 2000 (Peralta-Higuera *et al.,* 2001; Mas *et al.,* 2001; Figura 6). En todo rigor, la evaluación de la confiabilidad que se deriva de estas fotografías es únicamente válida para el área cubierta por las líneas de vuelo. Sin embargo, estas líneas de vuelo cubren el territorio nacional de manera homogénea y sistemática, por lo que se puede formular la hipótesis de que no existe sesgo, esto es, que la porción del territorio que corresponde a las líneas de vuelo es representativa del conjunto del territorio y, por tanto, los resultados obtenidos son también válidos para la parte no cubierta por las fotos aéreas (Mas *et al.,* 2002).

Con el fin de reducir el número de sitios de verificación, se puede manejar una precisión variable según el interés que el usuario tenga para las diferentes categorías. Por ejemplo, la evaluación de un mapa de uso del suelo y vegetación para un usuario forestal se debe llevar a cabo con un esfuerzo de muestreo más importante en las clases forestales, lo que permite una precisión mayor y un esfuerzo menor en las demás categorías, resultando en una precisión menor en la estimación de la confiabilidad de estas categorías.

Con el fin de utilizar el número de sitios de muestreo mínimo para obtener la precisión deseada, se puede utilizar un método iterativo (Sánchez-Colón, comunicación personal). En una primera etapa se determina el número de sitios de muestreo para tener un nivel de confianza aceptable (por ejemplo * < 10%) con un nivel de confiabilidad alto (80 o 90%). Para las clases que presentan un nivel de confiabilidad bajo, el intervalo de confianza será muy grande y se analizarán sitios adicionales para reducir este intervalo. Por ejemplo, se evalúa la confiabilidad de 3 clases con 50 unidades de muestreo por clase. Para 2 de ellas se obtiene 96% ± 5.4 y 10% ± 8.3 respectivamente, para la tercera se obtiene 48% ± 13.8. Se necesita un esfuerzo de muestreo adicional para esta última clase, con el fin de reducir el intervalo de confianza por debajo de 10%. Con 50 unidades de muestreo adicionales se obtiene finalmente 56% ± 9.7 (Tabla 5).

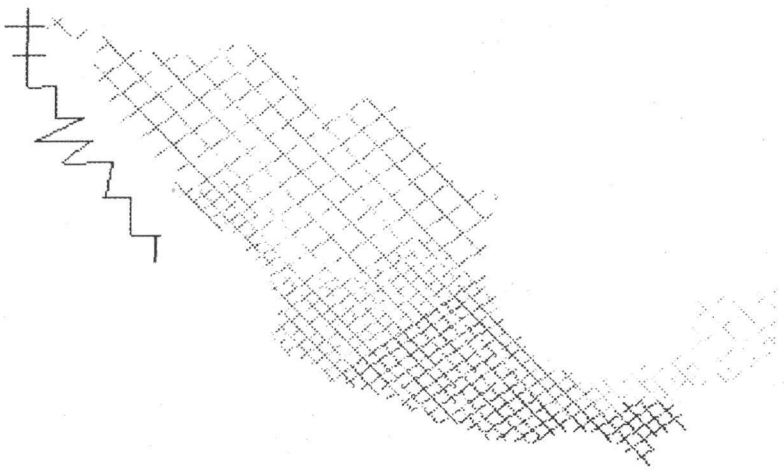


Figura 6. Líneas de vuelo para la evaluación de la confiabilidad de la cartografía del Inventario Forestal Nacional 2000 (tomado de Peralta-Higuera, *et al.*, 2001).

Tabla 5. Ejemplo numérico del método iterativo Para la clase C, se necesitaron 50 sitios adicionales para obtener un intervalo de confianza más pequeño (C*)

	Clase			
	A	6	C	C*
Núm. de sitios de verificación	50	50	50	100
Núm. de aciertos	48	5	24	56
% aciertos	96	10	48	56
Medio intervalo de confianza	5.4	8.3	13.8	9.7

El problema de la confiabilidad de los datos "verdad"

La evaluación de la confiabilidad se hace bajo el supuesto de que los datos de referencia son totalmente correctos, lo que no resulta cierto en muchos casos, en particular cuando la clase correspondiente a los sitios de verificación se determina con base en el análisis de fotografías aéreas (Biging *et al.,* 1994; Congalton, 1991). Por ejemplo, Congalton y Green (1993) reportan que en el ejercicio de la evaluación de un mapa de bosques con base en la interpretación de fotografías aéreas, en 41% de los sitios de verificación, la fotointerpretación proporcionó una clase distinta según el fotointérprete. Además, las verificaciones en campo permitieron determinar que en el 20% de los casos, la categoría obtenida por fotointerpretación era errónea y el 23% de los sitios no presentaba el mismo tipo de cobertura, debido al cambio de uso del suelo entre las fechas de toma de las imágenes que se usaron para la elaboración del mapa y su evaluación.

La evaluación de mapas de cambio

La evaluación de la confiabilidad de mapas o de imágenes de cambio es un caso un poco particular de la evaluación de un mapa temático. Existe una gran variedad de métodos de detección de cambio, que permiten generar información sobre los cambios de cobertura con base en el análisis de mapas o de imágenes de diferentes fechas. Algunos métodos permiten solamente generar mapas binarios que indican si hubo cambio o no (categorías "cambio" y "no cambio"), otros permiten caracterizar más precisamente el tipo de cambio e indican la combinación de clases para las dos fechas involucradas (por ejemplo, "selva/pastizal" o "zona urbana/ zona urbana"; Singh, 1989; Mas, 2000). Generalmente, los mapas de cambio son muy sensibles a la calidad geométrica de los insumos (corrección geométrica de las imágenes, precisión en la delimitación de los polígonos; Aspinall y Pearson, 1995; Green y Hartley, 2000). Otras características que deben tomarse en cuenta para la evaluación de estos mapas de cambio son: a; el gran número de categorías, ya que resultan de la combinación entre las categorías de las diferentes fechas; *b)* la imposibilidad de realizar verificación de campo, por lo menos para la fecha más antigua, y *c)* el hecho de que las categorías de cambio representan generalmente una pequeña proporción del área cartografiada. Debido a esta última característica, Khorram (1999) recomienda estrate-

gias y esfuerzos de muestreo diferentes para las categorías de cambio y de no cambio.

CONCLUSIÓN

Con base en los resultados de la evaluación de la confiabilidad, es necesario considerar las diferentes medidas, que permiten mejorar la información contenida en un mapa temático y los productos que se deriven del mismo. En el proceso de evaluación de la confiabilidad, las categorías que más se confunden pueden ser objeto de un esfuerzo adicional, con base en un exhaustivo trabajo de campo, por ejemplo, que permita mejorar la calidad de su representatividad en la cartografía. Otra opción más económica que ayuda a resolver esta disyuntiva consiste en reagrupar las categorías confundidas en una sola, con el fin de absorber las confusiones. En este caso, se pierde detalle en el mapa, pero se incrementa su confiabilidad (Congalton, 1988a; Janssen y Van der Wel, 1994).

La matriz de confusión indica las clases que tienden a ser sub o sobre-representadas en el mapa y, por consecuencia, cuya superficie está sub o sobre-estimada. Existen diversos métodos para llevar a cabo correcciones de las estadísticas de superficie con base en el análisis de esta matriz (Hay, 1988; Jupp, 1989; Dymond, 1992; Czaplewski y Catts, 1992; Card, 1982 y Conese y Maselli, 1992).

Los resultados de la evaluación de la confiabilidad están reportados de manera muy heterogénea en la literatura. Además, los autores utilizan diferentes índices que vuelven las comparaciones entre estos estudios aún más difíciles. En muchos casos, faltan datos importantes, como una descripción del diseño de muestreo, lo que puede modificar la interpretación de los resultados. Varios autores abogan por una estandarización del método de evaluación de la confiabilidad; otros argumentan que sería perjudicial imponer un esquema único de evaluación y de reporte de los resultados; ya que los objetivos de la cartografía y de la evaluación de la confiabilidad, así como las condiciones para llevarlos a cabo, pueden ser muy distintas (Foody, 2002). Sin embargo, todos concuerdan en que la evaluación de la confiabilidad no se puede resumir en un solo índice y que es muy importante incorporar en el reporte información sobre el muestreo, así como la matriz "bruta", para que se pueda llevar a cabo el cálculo de otros índices que no fueron contemplados por los autores (Stehman, 1997).

El valor de los datos geográficos depende de la calidad de su sistema de clasificación y su confiabilidad. Con base en los insumos comúnmente utilizados para la cartografía (imágenes de percepción remota) y por la naturaleza misma del objeto de estudio, como por ejemplo, la vegetación que presenta una distribución a veces difusa, es difícil obtener una representación cartográfica totalmente libre de error o de ambigüedades. Es, por tanto, muy importante realizar una evaluación de la confiabilidad de estos datos, ya que, de manera cada vez más frecuente son utilizados como insumo para la evaluación de tierras, el ordenamiento territorial, el cálculo de emisión o secuestro de CO_2, el diseño de políticas para la conservación de la biodiversidad, entre otros.

AGRADECIMIENTOS

Se agradece a S. Sánchez-Cotón, a S. Stehman y a dos revisores anónimos por sus comentarios. Este artículo se elaboró en el ámbito del proyecto *"Desarrollo de un método de evaluación de la confiabilidad de mapas de vegetación y uso del suelo, mediante el enfoque difuso (Fuzzy)"* apoyado por el CONACyT (proyecto 38965-T). El segundo autor agradece al CONACyT, por otorgarle la beca-crédito de doctorado periodo 2001-2003, con el número de cuenta 124650.

REFERENCIAS

Aronoff, S. (1982), "Classification accuracy: a user approach", *Photogrammetric Engineering and Remote Sensing,* 48(8):1299-1307.

Aspinall, R. y D. M. Pearson (1995), "Describing and managing uncertainty of categorical maps in GIS", Fisher, P. (ed.), *Innovations in GIS* 2, Taylor & Francis, London, pp. 71-83.

Biging, G. S., R. Congalton y E. Murphy (1994), "A comparison of photointerpretation and ground measurements of forest structure", Fenstermaker, L. (ed), *Remote Sensing Thematic Accuracy Assessment: a Compendíum,* American Society for Photogrammetry and Remote Sensing, pp. 148-157.

Burrough, P. A. (1994), "Accuracy and error GIS", Green, D. R. y D. Rix (eds.), *The AGÍ Sourcebook for Geographic Information Systems 1995,* AGÍ, London, pp. 87-91.

Card, H. D. (1982), "Using known map category marginal frequencies to improve estimates of thematic map accuracy", *Photogrammetric Engineering and Remote Sensing,* 48 (3):431-439

Carmel, Y., D. J. Dean y H. F. Curtís (2001), "Combining location ancl classification error sources for estimating multi-temporal datábase accuracy", *Photogrammetric Engineering and Remote Sensing* 67(7):865-872.

Chrisman, N R., (1989), "Modeling error in overlaid categorical maps" trie accuracy of spatial databases, Goodchild, M. y Gopal, S. (eds), Chapter 2, Taylor & Francis, London, pp. 21-34.

Chuvieco, E. (1996), *Fundamentos de teledetección espacial,* 3ra. ed., Ediciones RIALP, Madrid, España.

Cochran, W. G. (1980), *Técnicas de muestreo,* CECSA, México.

Conese, C. y F. Maselli (1992), "Use error matrices to improve área estimates with máximum likelihood classification procedures", *Remote Sensing ofthe Environment,* 40:113-124.

Congalton, R. G. (1988a), "Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data", *Photogrammetric Engineering and Remote Sensing,* 54 (5):587-592.

Congalton, R. G. (1988b), "A comparison of sampling scheme use in generating error matrices for assessing the accuracy of maps generated from remotlely sensed data", *Photogrammetric Engineering and Remote Sensing,* 54(5):593-600.

Congalton, R. G. (1991), "A review of assessing the accuracy of classifications of remotely sensed data", *Remote Sensing of the Environment,* 37:35-46.

Congalton, R. G.. y K. Green (1993), "A practical look at the sources of confusión in error matrix generation", *Photogrammetric Engineering and Remote Sensing,* 59(5):641-644.

Czaplewski, R. y G. Catts (1992), "Calibration of remotely sensed proportion or área estimates for misclassification error", *Remote Sensing ofthe Environment,* 39:29.43.

Dicks, S. E. y T. H. C. Lo (1990), "Evaluation of thematic map accuracy in a land-use and land-cover mapping program", *Photogrammetric Engineering and Remote Sensing,* 56(9): 1247-1252.

Dymond, J. R. (1992), "How accurately do image classifiers estímate área?", *International Journal Remote Sensing* 13(9):1735-1742.

Fitzpatrick-Lins, K., (1981), "Comparison of sampling procedures and data analysis for a land-use and land-cover map", *Photogrammetric Engineering and Remote Sensing,* 47:343-351.

Foody, G. M. (1992), "On the compensation for chance agreement in image classification accuracy assesmenf, *Photogrammetric Engineering and Remote Sensing,* 58{10): 1459-1460.

Foody, G. (2002), "Status of land cover classification accuracy assessment", *Remote Sensing of Environment,* 80:185-201.

Friedl, M. A., C. Woodcock, S. Gopal, D. Muchoney, A. H. Strahler y C, Barker-Schaaf (2000), "A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data", *International Journal Remote Sensing,* 21(5):1073-1077.

Goodchild, M. F., S. Gouquing y Y. Shiren (1992), "Development and test of an error model for categórica! data", *International Journal of Geographical Information Systems,* 6:87-104.

Gopal, S., C. E. Woodcock y A. Strahler (1999), "Fuzzy neural network classification of global land cover from a 1^{o} AVHRR Data Set", *Remote Sensing Environment,* Elsevier Science Inc., New York, 67:203-243.

Gopal, S., y C. E. Woodcock (1994), "Accuracy of thematic maps using fuzzy sets I: theory and methods", *Photogrammetric Engineering and RemoteSensing,* 58:35-46.

Green, D. R. y S. Hartley (2000), "Integrating photointerpretation and GIS for vegetation mapping: some issues of error", Alexander, R. y A. C. Millington (eds.), *Vegetation Mapping: From Patch to Planet,* John Wiley & Sons Ltd., pp. 103-134.

Hammond, T. O. y D. L Verbyla (1996), "Optimistic bias in classification accuracy assessment", *International Journal Remote Sensing,* 17 (6):1261-1266.

Hay, A. M. (1988), "The derivation of global estimates from a confusión matrix", *International Journal Remote Sensing,* 9(8):1395-1398.

tU Hord, R. M. y W. Brooner (1976), "Land-use map accuracy criteria", *Photogrammetric Engineering and Remote Sensing,* 42(5):671-677.

Janssen, L. F. y F. J. van der Wel (1994), "Accuracy assessment of satellite derived land-cover data, a review", *Photogrammetric Engineering and Remote Sensing,* 60(4):419-426.

Jupp, D. (1989), "The stabtüty of global estimates from confusión matrices", *International Journal Remote Sensing,* 10(9):1563-1569.

Khorram, S. (1999), "Accuracy assessment of remote sensing-derived change detection", *Photogrammetric Engineering and Remote Sensing.* Monograph Series.

Khorram, S , J Knight, H. Cakir, H. Yan, Z Mao y X. Dai (2000), "Improving estimates of the accuracy of thematic maps when using aerial photos as the ground reference source", *Proceedings of the ASPRS Symposium,* Washington, USA.

Luneta, R., R. G. Congalton, L K. Fnstermaker, J. R. Jensen, K. C. McGw y L. R. Tinney (1991), "Remote sensing and geographic ¡nformation systems data integration: error sources and research issues", *Photogrammetric Engineering and Remote Sensing,* 57(6): 677-687.

Ma, Z. y R. L Redmond (1995), "Tau coefficients for accuracy assessment of classification of remote sensing data", *Photogrammetric Engineering and Remote Sensing,* 61:435-439.

Mas, J. F., (2000), "Une revue des techniques et méthodes de télédétection du changement", *Canadian Journal of Remote Sensing I Journal Canadien de télédétection,* vol. 26(4):349-369.

Mas, J. F., J. L. Palacio, A. Velázquez y G. Boceo (2001), "Evaluación de ¡a confiabilidad temática de bases de datos cartográficas", *Memoria Digital CD interactivo,* ler. Congreso Nacional de Geomática, Guanajuato, 26-28 de septiembre de 2001, http://indy2.igeograf.unam.mx/dote/publicaciones/evalconf_congreso.htm

Mas, J. F., A. Velázquez, J. L. Palacio-Prieto, G. Boceo, A. Peralta y J. Prado (2002), "Assessing forest resources in México: Wall-to-wall land use/cover mapping", *Photogrammetric Engineering and Remote Sensing* 68(10):966-968. http://asprs.org/asprs/publications/pe&rs/2002journal/october/high!ighthtml.

Millington, A.C. y R. W. Alexander (2000), "Vegetation mapping in the last three decades of the twentieth century", Millington, A. C y R W Alexander (eds.), *Vegetation Mapping,* John Wiley & Sons, Chochester, England, pp. 321-331.

Naeseet, E. (1996), "Use of the weighted Kappa coefficient in classification error assessment of thematic maps", *International Journal Geographical Information Systems* 10(5):591-604.

Peralta-Higuera, A., J. L Palacio, G. Boceo, J, F. Mas, A. Velázquez, A. Victoria, R. Bermúdez, U. Martínez y J. Prado (2001), "Nationwide sampling of México with airborne digital cameras: an image datábase to validate the interpretaron of satellite data", *American Society for Photogrammetry and Remote Sensing,* 18th Biennal

Workshop on Color Photography & Videography in Resource Assessment. Amherst, Mass., USA, mayo 16-18, pp. 1-9.

Pontius R.G. (2000), "Quantification error *versus* location error in comparison of categorical maps", *Photogrammethc Engineerίng and Remote Sensing,* 66(8)1011-1016.

Pontius R. G. (2002), "Statistical methods to partition of quantity and location during comparison of categorical maps at múltiple resolutions", *Photogrammetric Engineering and Remote Sensing,* 68(10):1041-1049.

Rosenfield, G. H. y K. Fitzpatrick-Lins (1986), "A coefficient of agreement as a measure of thematic classification accuracy", *Photogrammethc Engineering and Remote Sensing,* 52(2):223-227.

Singh, A. (1989), "Digital change detection techniques using remotely-sensed data", *International Journal of Remote Sensing,* 10(6):989-1003.

Stehman, S. V. y R. L. Czaplewski (1998), "Design and analysis for thematic map accuracy assessment: fundamental principies", *Remote Sensing Environment,* 64:331-344.

Stehman, S. V. (1992), "Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data", *Photogrammetric Engineering and Remote Sensing,* 58(9):1342-1350.

Stehman, S. V. (1996), "Estimating the kappa coefficient and its variance under stratified random sampling", *International Journal Remote Sensing,* 62:401-407.

Stehman, S. V. (1997), "Selecting and interpreting measures of thematic classification accuracy", *Remote Sensing, Environment,* 62:77-89.

Stehman, S. V. (1999), "Basic probability sampling designs for thematic map accuracy", *International Journal Remote Sensing* 20 (12):2423-2441.

Stehman, S. V. (2000), "Practical implications of design-based sampling inference for thematic map accuracy assessment", *Remote Sensing of Environment,* 72:35-45.

Stehman, S. V. (2001), "Statistical rigor and practical utility in thematic map accuracy assessment", *Photogrammetric Engineering and Remote Sensing,* 67(6):727-734.

Turk, G. T. (1979), "GT Índex: a measure of the success of prediction", *Remote Sensing Environment,* 8:65-75.

Vogelmann, J. E., S. M. Howard, L Yang, C. R. Larson, B. K. Wylie y N. Van Driel (2001), "Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources", *Photogrammetric Engineering and Remote Sensing,* 67(6): 650-662.

Walsh, J. E., D. R. Lightfoot y D. R. Buttler (1987), "Recognitíon and assessment of error in geographic informatíon systems", *Photogrammetric Engineering and Remote Sensing,* 53 (10):1423-1430.

Wonnacott, T. H. y R. J. Wonnacott (1991), *Statistiques,* Ed. Económica, 4a edición.

Woodcock, C. y S. Gopal (2000), "Accuracy assessment and área estimates using fuzzy sets". *International Journal of Geographical Information Science.* 14(2): 153-172.