

Error, Exactitud y Precisión

Texto original: Error, Accuracy, and Precision - Kenneth E. Foote and Donald J. Huebner, Dpto. of Geography of Texas at Austin, 1995. The Geographer Graft Project, Dpto. of Geography, The University of Colorado at Boulder . (Contactar con k.foote@colorado.edu)

Traducción: José Antonio Rodríguez Esteban, Prof. de Geografía, Dpto. Geografía, Lab. Geografía Aplicada, Proyecto GEOteca, Universidad Autónoma de Madrid (2002).

1. La importancia del Error, la Exactitud y la Precisión
2. Algunas definiciones básicas
3. Tipos de error
 - 3.1. Exactitud y precisión posicional
 - 3.2. Exactitud y precisión de los atributos
 - 3.3. Exactitud y precisión conceptual
 - 3.4. La lógica de la exactitud y la precisión
4. Fuentes de inexactitud e imprecisión
 - 4.1. Fuentes obvias de error
 - 4.1.1. Antigüedad de los datos
 - 4.1.2. Área de cobertura
 - 4.1.3. Escala del mapa
 - 4.1.4. Densidad de las observaciones
 - 4.1.5. Relevancia
 - 4.1.6. Formato
 - 4.1.7. Accesibilidad
 - 4.1.8. Coste
 - 4.2. Errores resultantes de la variación natural de los datos originales
 - 4.2.1. Exactitud posicional
 - 4.2.2. Precisión en el contenido
 - 4.2.3. Fuentes de variación de los datos
 - 4.3. Errores originados durante los procesos
 - 4.3.1. Errores numéricos
 - 4.3.2. Errores en los análisis topológicos
 - 4.3.3. Problemas de la clasificación y la generalización
 - 4.3.4. Digitalización y los errores geocodificados
5. Los problemas de la propagación y de la conexión en cascada
 - 5.1. Propagación
 - 5.2. Conexión en cascada
6. Cuidado con la falsa precisión y la falsa exactitud
7. Los peligros de los datos indocumentados
 - 7.1. Investigue cuando pida prestado o compre datos
 - 7.2. Prepare un informe de la calidad de los datos que vaya a utilizar
 - 7.3. Haga preguntas sobre la indocumentación de los datos
8. Principios del manejo del error (en inglés)
9. Bibliografía utilizada
10. Referencias y bibliografía adicional

1. La importancia del Error, la Exactitud, y la Precisión

Sólo recientemente, los usuarios y desarrolladores de los Sistemas de Información Geográfica (SIG) han prestado atención a los problemas causados por el error, la exactitud y la imprecisión en el conjunto de datos espaciales. Ciertamente, existía la conciencia de que todos los datos contenían cierta inexactitud e imprecisión, pero su efecto en los problemas y soluciones de los SIG no ha sido considerada con gran detalle. Las principales introducciones a los SIG, tales como la de C. Dana Tomlin *Geographic Information Systems and Cartographic Modeling* (1990), la de Jeffrey Star y John Estes's *Geographic Information Systems: And Introduction* (1990), o la de Keith Clarke's *Analytical and Computer Cartography* (1990), apenas tratan esta cuestión.

Esta situación ha cambiado sustancialmente en los últimos años. Ahora existe un reconocimiento general de que el error, la inexactitud y la imprecisión pueden "quebrar" algunos tipos de proyectos SIG. Esto es, los errores no detectados, pueden dejar sin valor algunos de los análisis GIS.

La ironía está en que el problema del error es inherente a uno de las grandes potencialidades de los SIG. Gran parte de las soluciones aportadas por los SIG son posibles gracias a que cotejan y cruzan diversos tipos de datos con localización. Esto es particularmente útil al posibilitar integrar diversos conjuntos de datos discretos bajo un único sistema. Desafortunadamente, cada vez que se importa un nuevo conjunto de datos, el SIG arrastrará el error inherente a los mismos. La mezcla y combinación de errores puede llevar al conjunto de datos por caminos impredecibles.

Una de las primeras discusiones en profundidad sobre el problema y las fuentes de error aparece en P. A. Borrough's *Principles of Geographical Information Systems form Land Resources Assessment* (1986). Ahora la cuestión aparece tratada en varias introducciones al los SIS, como en *Geographical Information System: A Guide to the Technology* (1991) de John Antenucci, Kay Broen, Peter Crosswell, Michael Kevany and Hugh Archer.

El punto central está en que a través del error se pueden alterar los análisis del SIG, que hay diversos caminos para reducir el error al mínimo mediante una cuidadosa planificación y métodos para estimar sus efectos en las soluciones SIG. El conocimiento del problema del error ha tenido como consecuencia beneficiosa el hacer sensible a los usuarios de los SIG de las potenciales limitaciones inherente al proceso para alcanzar exactitud y precisión en las soluciones.

2. Algunas definiciones básicas

Es importante distinguir desde el principio la diferencia entre exactitud y precisión:

Exactitud es el grado en el cual la información de un mapa o en una base de datos digital se muestra verdadera o con valores aceptables. La exactitud es un asunto perteneciente a la cualidad de los datos y al número de errores contenidos en un conjunto de datos o mapa. Analizando una base de datos de

un SIG, es posible considerar la exactitud horizontal y vertical con respecto a la posición geográfica, tanto atributiva y conceptual, como en la agudeza lógica. El nivel de exactitud requerido puede variar enormemente de unos casos a otros.

Producir y compilar una gran exactitud en los datos puede ser muy difícil y costoso.

Precisión hace referencia a la medida y exactitud de las descripciones en las base de datos de un SIG. Los atributos de información precisos pueden especificar las características de los elementos con gran detalle. Es importante observar, no obstante, que los datos precisos - no importando el cuidado en su medida - pueden ser inexactos. Los topógrafos pueden cometer errores, o bien los datos pueden ser introducidos en las bases de datos incorrectamente.

El nivel de precisión requerido puede variar enormemente de unos casos a otros. Los proyectos de ingeniería como el de una carretera, y las herramientas de construcción, requieren una muy precisa medida, de milímetros a decenas de centímetros. Análisis demográficos de las tendencias del electorado pueden prescindir de esta precisión mediante un código postal o de circunscripción.

Obtener datos altamente precisos puede ser verdaderamente difícil y costoso. Topografías cuidadosamente las localizaciones requieren de compañías específicas para la recogida de la información.

Gran precisión no es indicativa de gran exactitud y tener gran exactitud no implica gran precisión. Pero gran exactitud y gran precisión son bastante expresivas.

Los usuarios de los SIG no son siempre conscientes en el uso de los términos. En ocasiones ambos términos son intercambiables lo que resulta contraproducente.

Dos términos adicionales son igualmente usados:

Calidad de los datos hace referencia a la relativa exactitud y precisión de una base de datos particular en un GIS. Estos hechos están a menudo documentados en los informes de calidad.

Error acompaña tanto a la imprecisión de los datos como a su inexactitud.

3. Tipos de error

El error posicional es el que más a menudo concierne a los SIG, pudiendo afectar a diferentes características de la información almacenada en una base de datos.

3.1. Exactitud y precisión posicional

Es aplicable tanto a la posición horizontal como a la vertical.

Exactitud y precisión están en función de la escala en la que ha sido creado el mapa (impreso o digital). Los mapas estándar empleados por el Servicio Geológico de los Estados Unidos (USGS) especifican que:

"se requiere una exactitud horizontal del 90 % en todos los puntos tomados que deben de estar entre 1 y 30 pulgadas (2,54 y 76.2 cm) para mapas de escala superior a 1:20.000 y entre 1 y 50 pulgadas (2,54 y 127 cm) para mapas de

escala inferior a 1:20.000"

Precisiones estándar para algunas escalas de mapas

1:1.200 ± 3,33 pies (± 1,015 m)

1:2.400 ± 6,67 pies (± 2,033 m)

1:4.800 ± 13,33 pies (± 4,063 m)

1:10.000 ± 27,78 pies (± 8,467 m)

1:12.000 ± 33,33 pies (± 10,159 m)

1:24.000 ± 40,00 pies (± 12,192 m)

1:63.360 ± 105,60 pies (± 32,187 m)

1:100.000 ± 166,67 pies (± 50,80 m)

(Nota: 1 pie = 30,48 cm = 0,3048 m)

Esto significa que cuando nosotros vemos un punto en un mapa, tendremos esta probabilidad de que se encuentre dentro de cierta área. Esto se hace extensivo a las líneas.

Por otra parte, están los peligros de la falsa exactitud y de la falsa precisión, que son leídos en la información locacional desde los mapas con niveles de exactitud y precisión bajo los cuales han sido creados. Esto es un verdadero peligro en los ordenadores, puesto que permite a los usuarios aumentar y reducir las vistas en un número infinito de escalas. Exactitud y precisión están unidos a la escala original del mapa y no cambia aunque se use el zoom para aumentar o reducir la vista. Estas operaciones pueden incluso hacer creer - falsamente- que la exactitud y la precisión son mejores.

3.2. Exactitud y precisión de los atributos

Los datos no espaciales unidos a la localización pueden ser inexactos o imprecisos. La inexactitud puede ser consecuencia de errores de distinto tipo. Los datos no espaciales pueden variar mucho también en precisión. La información precisa de los atributos describen fenómenos con gran detalle. Por ejemplo, la descripción precisa de una persona que vive en una dirección particular puede incluir género, edad, ingresos, ocupación, nivel de educación y muchas otras características. Una descripción imprecisa puede incluir sólo los ingresos o sólo el género.

3.3. Exactitud y precisión conceptual

Los SIG dependen sobretodo de la abstracción y la clasificación de los fenómenos del mundo real. Los usuarios determinan que cantidad de información debe usarse y como ésta debe ser clasificada en categorías apropiadas. En ocasiones, los usuarios pueden usar inapropiadas categorías o una clasificación errónea de la información. Por ejemplo, la clasificación de ciudades por el comportamiento del voto electoral es una vía inadecuada para estudiar la fertilidad de las parejas; fallos en la clasificación de las líneas de alto voltaje puede limitar la efectividad en el diseño de un SIG en la construcción de las infraestructuras eléctricas. Aún empleando correctas categorías los datos

pueden estar mal clasificados. Un estudio de los sistemas de drenaje puede necesitar de una clasificación de las corrientes y ríos por su "orden", atendiendo su jerarquía al lugar donde una corriente particular desagua en el sistema tributario de la red. Los canales individuales pueden estar mal clasificados si los tributarios están mal localizados. Por ello, algunos estudios pueden no requerir un tipo preciso de categorización del orden de las corrientes. Todo lo más que pueden necesitar es la localización y el nombre de las corrientes fluviales, sin tener en cuenta el orden.

3.4 La lógica de la exactitud y precisión

La información almacenada en una base de datos puede estar ilógicamente introducida. Por ejemplo, los permisos necesarios para construir una subdivisión residencial en un plano de inundación pueden necesitar comparar la proposición con el mapa del plano de inundación. Por lo tanto, la construcción puede ser posible en algunas zonas del plano de inundación pero su uso no será conocido hasta que las variaciones de la inundación potencial hayan sido registradas y puedan ser usadas en la comparación. La cuestión es, pues, que la información almacenada en la base de datos de un SIG puede ser usada y cuidadosamente comparada, si produce resultados útiles. Los SIG están normalmente incapacitados para avisar a los usuarios cuando se produce una inapropiada comparación o si los datos han sido utilizados incorrectamente. Algunas reglas de uso pueden ser incorporadas en el diseño de un SIG, como sucede con los "sistemas expertos", pero los desarrolladores necesitarían estar seguros que las reglas empleadas corresponden al mundo real de los fenómenos que ellos modelan.

Finalmente señalar, cometeremos una equivocación si creemos que una gran exactitud y una gran precisión de la información es necesario para todas las aplicaciones de los SIG. La necesidad de exactitud y precisión puede variar radicalmente dependiendo del tipo de información codificada y del nivel de medida necesario para una particular aplicación. Son los usuarios los que deben determinar el alcance de su trabajo. Excesiva exactitud y precisión no sólo es costoso, sino también puede resultar un gran engorro.

4. Fuentes de inexactitud e imprecisión

Son muchas las fuentes de error que pueden afectar la calidad del conjunto de datos de un SIG. Esto, que resulta muy obvio, puede no ser tan difícil de discernir. Algunas de ellas serán automáticamente identificadas por el mismo SIG, pero es responsabilidad del usuario su prevención. Algunos casos particulares puede necesitar de comprobaciones específicas de error, porque los propios SIG son capaces de inducir al usuario una falsa sensación de exactitud y precisión sin garantizar la validez de los datos. Por ejemplo, suavizar cambios en las líneas fronterizas, en las curvas de nivel y en las zonas de cambio de los mapas de coropletas es una "elegancia que falsea" la realidad. En realidad, estas cuestiones son a menudo "vagas, graduales o azarosas" (Burrough 1986). Hay una imprecisión inherente en la cartografía como resultado de los procesos de proyección y la necesaria distorsión producida en algunos de sus datos (Koeln et all, 1994); una imprecisión que

puede continuar a través de los procesos aplicados con los SIG. Los usuarios de los SIG deben ser capaces, no sólo de reconocer el error, sino el grado de error tolerable y asumible del sistema.

Burrough (1986) divide las fuentes de error en tres grandes categorías:

Fuentes de error obvias.

Errores resultantes de la variación natural de las mediadas originales.

Errores surgidos en los procesamientos.

Generalmente los dos primeros errores son más fáciles de detectar que aquellos errores del tercer tipos, surgidos al procesar los datos, por permanecer un tanto escondidos y ser difíciles de identificar. Burrough divide estos grupos principales en distintas categorías, tratadas a continuación.

4.1 Fuentes obvias de error

4.1.1. Antigüedad de los datos.

Las fuentes de datos pueden ser simplemente antiguas para ser usadas en un proyecto SIG. Las colecciones estándar del pasado pueden ser desconocidas, inexistentes o desfasadas. Por ejemplo, los datos topográficos del Gran Cañón obtenidos por el decimonónico John Wesley Powell, contienen falta de precisión para ser utilizados hoy en día. Además, una parte de la información base puede, además, haber cambiado como consecuencia de la erosión, la deposición o cualquier otro proceso geomorfológico. Pese al poder de los SIG, la dependencia de datos antiguos puede tergiversar, sesgar o convertir en negativos los resultados.

4.1.2. Área de cobertura.

Los datos de una zona determinada pueden haber desaparecido completamente, o únicamente una parte de los niveles de información pueden ser utilizables en un proyecto SIG. Por ejemplo, los mapas de vegetación o de suelo pueden estar incompletos en las zonas de transición o faltarles exactitud en la representación. Otro ejemplo, es la falta de datos proporcionados por los sensores remotos en ciertas partes del mundo al estar permanentemente nublado. La exactitud de una cobertura uniforme puede no estar disponible y el usuario debe decidir que nivel de generalización debe ser necesaria si una nueva colección de datos es requerida.

4.1.3. Escala del mapa.

La posibilidad de los mapas para mostrar detalles está determinada por la escala. Un mapa con una escala 1:1.000 puede ilustrar detalles más precisos que otro a una escala 1:25.000. La escala determina el tipo, cualidad y cantidad de los datos (Star and Estes 1990). Se debe elegir siempre la escala adecuada al nivel de detalles requerido en el proyecto. Transformar la pequeña escala de un mapa en otra más grande no aumentará el número de detalles o el nivel de precisión de dicho mapa.

4.1.4. Densidad de las observaciones.

El número de observaciones realizadas en un área es una guía de la verosimilitud del mapa y debe ser conocido por los usuarios del mismo. Un insuficiente número de observaciones puede no proporcionar el adecuado nivel de resolución requerido para efectuar análisis espaciales y resolver los objetivos marcados en el proyecto SIG. En el caso de un punto, si las curvas de nivel poseen un intervalo de 120 cm, no es posible bajar el nivel de precisión. Las líneas de un mapa son una generalización basada en el intervalo de datos grabados, de este modo el más cercano al intervalo muestral, alcanza la mayor precisión de datos dibujados.

4.1.5. Relevancia.

Bastante a menudo, no es posible satisfacer los deseos de obtener datos de un lugar o de un área, y en su lugar debe haber sustitución estos datos por otros. Ha de existir una relación de validez entre los datos sustituidos y el fenómeno a estudiar, pero aún así, pueden producirse errores en tanto en cuanto los fenómenos no han sido medidos directamente. Un ejemplo local en el uso de datos sustituidos puede tomarse de los estudios del hábitat de la curruca en Hill Country. Es muy costoso (y molesto para los pájaros) inventariar este hábitat mediante observación directa. No obstante, las currucas prefieren vivir en viejos cedros *Juniperus ashei*. El hábitat puede ser localizados por fotografía aérea. La densidad de los *Juniperus ashei* puede utilizarse para deducir la densidad del hábitat de las currucas. Por supuesto, algunas áreas de cedro estarán inhabitadas o, por contra, tendrán una gran densidad. Estas áreas pueden no ser visibles cuando se utiliza fotografía aérea para tabular el hábitat.

Otro ejemplo de deducción de datos se produce con la señal electrónica utilizada para estimar mediante imágenes de satélite la cobertura vegetal, los tipos de suelo, la erosionabilidad y otras tantas características. Los datos son obtenidos por métodos indirectos. Los sensores de los satélites no pueden "ver" los árboles, si no únicamente ciertas firmas digitales típicas de los árboles y la vegetación. En ocasiones, estas firmas son almacenadas por los satélites aunque no estén presentes los árboles y la vegetación (falso positivo) o no ser recogidas cuando los árboles y la vegetación si están presentes (falso negativo). Dado el alto coste de obtención de datos in situ, sustituir los datos por deducción es con frecuencia utilizado y el usuario debe

entender estas variaciones y asumir, o no, su validez en función de la exactitud requerida en el proyecto.

4.1.6. Formato.

Los métodos para transmitir, almacenar y procesar la información de forma digital, pueden introducir error en los datos. Las conversiones de escala y proyección, los cambios desde raster a vector y la resolución del tamaño y profundidad del píxel, son ejemplos de los posibles errores inherentes al formato de los datos. En ocasiones, los datos han de ser transmitidos y utilizadas en múltiples SIG por lo que deben reformarse bajo un mínimo denominador común. Múltiples conversiones desde un formato a otro pueden crear un efecto similar a realizar copia tras copia en un máquina fotocopidora. Además, hay que tener en cuenta que los estándares internacionales para la transmisión, almacenamiento y recuperación de datos cartográficos no estén totalmente realizados.

4.1.7. Accesibilidad.

La accesibilidad de los datos es otra cosa. Lo que está disponible en un país puede no estarlo en otros. Antes de la desaparición de la Unión Soviética, no pocos mapas eran considerados como documentación clasificada y por lo tanto, inobtenibles por la mayor parte de la gente. Las restricciones militares, la rivalidad entre agencias, las leyes de privacidad y los factores económicos pueden restringir la validez de los datos o su nivel de exactitud.

4.1.8. Coste.

Extensos y veraces datos son, a menudo, demasiado caros de obtener o convertir. Iniciar una nueva colección de datos puede ser demasiado caro para los beneficios generados en un determinado proyecto GIS y sus diseñadores deben moverse entre su deseo de exactitud y el coste de la información. La verdadera exactitud es cara y puede ser inasequible.

4.2. Errores resultantes de la variación natural de los datos originales.

En ocasiones estas fuentes de error pueden no ser tan obvias, una revisión cuidadosa puede mostrar su trascendencia en el proyecto.

4.2.1. Exactitud posicional.

La exactitud en la posición es una medida del desajuste entre los elementos del mapa y la verdadera posición de los atributos (Antenucci and others, 1991, 102). Depende del tipo de datos usados u observados. Los cartógrafos pueden situar con exactitud objetos bien definidos, como carreteras, edificios, líneas divisorias y unidades topográficas discretas en mapas y en sistemas digitales, mientras que separaciones menos discretas como las existentes entre la vegetación o los tipos de suelo pueden ser el resultado de las estimaciones del cartógrafo. El clima, los biomas, el relieve, los tipos de suelo, el drenaje y otros

elementos faltos de una clara delimitación en la naturaleza, son susceptibles de ser interpretados.

Defectos o trabajos parciales, errores de digitalización de mapas y de conversión en los mapas o en los escáner, pueden todos ellos producir mapas inexactos en un proyecto SIG.

4.2.2. Precisión en el contenido.

Los mapas deben de ser correctos y estar libre de presuposiciones. La precisión cualitativa hace referencia a la corrección en la clasificación y a la presencia de elementos específicos. Por ejemplo, un bosque de pinos puede estar incorrectamente clasificado como un bosque de abetos, introduciendo de esta forma errores que no pueden ser conocidos o sospechados por el usuario de los datos o del mapa. Ciertos elementos pueden ser omitidos tanto desde el mapa como desde la base de datos espacial por descuido o intencionadamente.

Otros errores en exactitud cuantitativa pueden ocurrir por los defectos de los instrumentos de calibración usados para medir aspectos concretos como la altitud; el pH del suelo o del agua, o los gases atmosféricos. Los errores cometidos en el campo o en el laboratorio, pueden ser indetectables en un proyecto SIG, salvo que el usuario confirmara o corroborara la validez de la información.

4.2.3. Fuentes de variación de datos.

Las variaciones en los datos pueden realizarse por la introducción de errores de medida durante la observación, por la predisposición del observador o por falta de una adecuada calibración del equipamiento. Por ejemplo, se puede no esperar precisiones submétricas con un GPS de mano sin corrector diferencial. Por otra parte, una incorrecta calibración en las formas de disolver el oxígeno puede producir valores incorrectos de concentración del mismo en una corriente.

Pudiendo ser una variación natural durante la toma de datos. Así, por ejemplo, la salinidad en la bahía y en el estuario de Texas varía durante el año dependiendo del influjo de las corrientes frías en profundidad y de la evaporación. Si alguien no fuera consciente de esta variación natural, ideas y decisiones erróneas pudieran ser tomadas, introduciendo un significativo error en el proyecto SIG. En algunos casos, si el error no da lugar a inesperados resultados, su detección sería extremadamente difícil.

4.3. Errores originados durante los procesos

Los errores originados durante los procesos son los más difíciles de detectar por los usuarios de los SIG. Pueden ser específicamente buscados para lo cual se requiere conocimiento de la información y de los sistemas usados en su procesamiento. Hay suberrores que ocurren de diferentes modos, habiendo otros potencialmente más insidiosos, por que pueden ocurrir en múltiples conjuntos de datos durante su manipulación en un proyecto SIG.

4.3.1. Errores numéricos.

Diferentes ordenadores pueden no tener la misma capacidad para construir complejas operaciones matemáticas, pudiendo producir resultados significativamente diferentes desde un mismo problema. Borrough (1990) cita un ejemplo en la elevación al cuadrado de un número, lo que produce una diferencia del 1.200 %. Los errores en los procesos de cálculo ocurren en las operaciones de redondeo y son inherentes al número de dígitos manipulados por el procesador. Otra fuente de error puede deberse a defectos del propio procesador, como ha sucedido con un problema matemático identificado en los chips del Pentium de Intel (tm). En ciertos cálculos, el chip ofrecía respuestas equivocadas.

Un mayor reto es el de la exactitud en la conversión de mapas existentes en formato digital (Muehcke 1986). Como los ordenadores manipulan los datos en formato digital, los errores numéricos pueden producir resultados inexactos. En cualquier caso, los errores en los procesos numéricos son extremadamente difíciles de detectar, y quizá requieran de una sofisticación no presente en la mayoría de los usuarios de SIG o promotores de proyectos.

4.3.2. Errores en los análisis topológicos.

Los errores lógicos pueden causar una incorrecta manipulación de los datos y de los análisis topológicos. Se pueden reconocer qué datos no son uniformes y están sujetos a variaciones. La superposición de múltiples capas de mapas puede resultar ocasionar problemas del tipo "Slivers", "Overshoots" y "Dangles". Variaciones en la exactitud entre diferentes capas de mapas pueden oscurecer durante el proceso en la creación de "datos virtuales los cuales pueden dificultar el reconocimiento de los datos reales" (Sample 1994).

4.3.3. Problemas de clasificación y generalización.

Para el entendimiento humano, la comprensión de una vasta cantidad de datos reside en su clasificación, y en algunos casos en su generalización. Siguiendo a Borrough (1986, 137) clasificar los datos en torno a siete divisiones es el ideal, ya que se pueden retener fácilmente en la memoria. Definir como se harán los intervalos es otro problema. Por ejemplo, definir las causas de muerte en hombres de entre 18-25 años será significativamente diferente que entre 18-40 años. Los datos son más exactos y manipulables entre múltiples pequeños. Definir un múltiplo razonable y preguntases "por qué hay que comparar", es esencial (Tufte 1990, 67-79). La clasificación y la generalización de atributos usada en un GIS está sujeta a errores de interpolación y puede introducir irregularidades en los datos difíciles de detectar.

4.3.4. Digitalización y errores geocodificados.

Los errores ocurridos durante el transcurso de las fases de manipulación de datos tales como la digitalización y la geocodificación, el recubrimiento y las intersecciones de los límites, y los errores de rasterización de un mapa

vectorial. Los errores fisiológicos del operador por contracciones involuntarias del músculo pueden dar lugar a "spikes" (puntos), a switchbacks (zig-zags), a "polygonal knots" (nudos poligonales), y a "loops" (lazos). Los errores asociados con los mapas fuente dañados, el error del operador mientras lo convertía a digital, y los prejuicios puede ser comprobados comparando los mapas originales con versiones convertidas a digital. Otros errores resultan más evasivos.

5. Los problemas de la propagación y de la conexión en cascada

Esta discusión se ha enfocado en relación a los errores que pueden estar presentes en sistemas de datos individuales. Los SIG dependen generalmente de comparaciones de muchos sistemas de datos. Este diagrama esquemático demuestra cómo una variedad de conjunto de datos discretos puede tener que ser combinados y ser comparados para solucionar un problema de análisis del recurso. Es inverosímil que la información contenida en cada capa sea exactamente igual y precisa. Los errores pudieron también haber surgido compilando la información. Si éste es el caso, la solución al problema del SIG en si mismo puede ser inexacta, imprecisa o errónea.

El problema es esa inexactitud, imprecisión y el error se puede formar en los SIG que empleen muchas fuentes de datos. Hay dos maneras posibles para que esto ocurra.

5.1. Propagación.

La propagación ocurre cuando un error conduce a otro. Por ejemplo, si un punto de registro del mapa se ha convertido a digital en una cobertura y después se utiliza para colocar una segunda cobertura, la segunda cobertura propagará el primer error. De esta manera, un solo error puede conducir a otro y separarse hasta que corrompe los datos a través del proyecto entero del SIG. Para evitar este problema utilice el mapa de la escala más grande para colocar sus puntos.

La propagación ocurre a menudo en una manera acumulativa, como cuando los mapas de diversa exactitud se compaginan.

5.2. Conexión en cascada.

La conexión en cascada significa que la información errónea, imprecisa e inexacta sesgará la resolución de SIG cuando la información se combina selectivamente en nuevas capas y coberturas. En este sentido, la conexión en cascada ocurre cuando los errores se pueden propagar de manera incontrolada de capa en capa.

Los efectos de la conexión en cascada pueden ser, igualmente muy difíciles de predecir. Pueden ser aditivos o multiplicativos y pueden variar dependiendo de cómo se combina la información, variando de situación en situación. Como la conexión en cascada puede tener tales efectos imprevisibles, es importante probar su influencia en una solución dada de un SIG. Esto se hace calibrando

una base de datos de un SIG mediante técnicas tales como la del análisis de la sensibilidad. El análisis de la sensibilidad permite que los usuarios calibren cómo y cuántos errores tendrán solución. El análisis de la calibración y de la sensibilidad se discute en el error de manejo.

También es importante señalar que la propagación y la conexión en cascada pueden afectar a la horizontal, la vertical, la cualidad, la conceptualidad y a la exactitud lógica y la precisión

6. ¡Cuidado con la falsa precisión y la falsa exactitud !

Los usuarios de los SIG no están siempre enterados de los difíciles problemas causados por el error, la inexactitud y la imprecisión. Caen a menudo presa de la falsa precisión y de la falsa exactitud, así que divulgan sus resultados a un nivel de precisión o de exactitud que son imposibles alcanzar con sus materiales de fuente. Si las localizaciones en una cobertura de un SIG se miden solamente en cientos de pies de su posición verdadera, no tiene ningún sentido divulgar localizaciones predichas en una solución a una décima del pie. Es decir, sólo porque los ordenadores pueden almacenar imágenes numéricas con muchos espacios decimales no significa que todos esos espacios decimales sean "significativos". Es importante que las soluciones de un SIG sean divulgadas honestamente y sólo bajo la exactitud y la precisión en la que se puedan apoyar.

Esto significa en la práctica que las soluciones de un SIG son a menudo las mejores divulgadas como gamas o graduación, o presentadas dentro de intervalos estadísticos de confianza. Estas ediciones se tratan en el módulo, tratamiento del error.

7. Los peligros de los datos indocumentados

Después de lo dicho, es fácil imaginarse el peligro de usar datos indocumentados en un proyecto de SIG. A menos que el usuario tenga una idea clara de la exactitud y de la precisión del conjunto de datos, mezclar éstos en un GIS puede ser muy aventurado. Los datos que se han elaborado cuidadosamente se pueden interrumpir por errores que algún otro ha cometido. Esto trae a la luz tres situaciones importantes.

7.1. Investigue cuando pida prestado o compre datos.

Muchos de los datos de productos gubernamentales y comerciales importantes, trabajan dentro de los establecidos estándares de exactitud y precisión que están disponibles al público en forma impresa o digital. Estos documentos explican exactamente cómo fueron compilados los mapas y el conjunto de los datos, y tales informes se deben estudiar cuidadosamente. Los informes de calidad de los datos se proporcionan generalmente en las agencias locales y del estado o en los de agentes privados.

7.2. Prepare un informe de la calidad de los datos que vaya a utilizar.

Sus datos no tendrán valor a menos que se prepare también un informe de la calidad de los datos. Incluso si usted no planea compartir sus datos con otros, debe preparar un informe -para el caso de que se utilice el conjunto de datos otra vez en el futuro-. Si no se documentan el conjunto de datos cuando se crean, puede terminar por perder el tiempo más adelante comprobarlos una segunda vez. Utilice los informes de la calidad de los datos encontrados como modelos para documentar su conjunto de datos.

7.3. En ausencia de un informe de la calidad de los datos, haga preguntas acerca de los datos indocumentados antes de utilizarlos.

¿Cuál es la antigüedad de los datos?

¿De dónde proceden?

¿Por qué medio se crearon originalmente?

¿Cuál es la cobertura regional de los datos?

¿A qué escala del mapa fueron convertidos a digital los datos?

¿Qué proyección, sistema de coordenadas y 'datum' fueron utilizados en los mapas?

¿Cuál era la densidad de las observaciones usadas para su compilación?

¿Cómo de exactas son las características posicionales y de cualidad?

¿Parecen lógicos y consistentes los datos?

¿Parecen limpias las representaciones cartográficas?"

¿Son relevantes los datos para el proyecto actual?

¿Qué formato se mantienen los datos?

¿Cuándo fueron comprobados los datos?

¿Por qué fueron compilados los datos?

¿Cuál es realmente la fiabilidad del proveedor?

8. Bibliografía utilizada

Antenucci, J.C., Brown, K., Croswell, P.L., Kevany, M. and Archer, H. 1991. Geographic Information Systems: a guide to the technology. Chapman and Hall. New York.

Burrough, P.A. 1990. Principles of Geographical Information Systems for Land Resource Assessment. Clarendon Press. Oxford.

9. Referencias y bibliografía adicional

Antenucci, J.C., Brown, K., Croswell, P.L., Kevany, M. and Archer, H. 1991. Geographic Information Systems: a guide to the technology. Chapman and Hall. New York.

Burrough, P.A. 1990. Principles of Geographical Information Systems for Land Resource Assessment. Clarendon Press. Oxford.

Koeln, G.T., Cowardin, L.M., and Strong, L.L. 1994. "Geographic Information Systems". P. 540 in T.A. Bookhout ed. Research and Management Techniques for Wildlife and Habitat. The Wildlife Society. Bethesda.

Muehrcke, P.C. 1986. Map Use: Reading, Analysis, and Interpretation . 2d Ed. JP Publications, Madison.

Sample, V.A. (Ed). 1994. Remote Sensing and GIS in Ecosystem Management . Island Press. Washington, D.C.

Star, J. and Estes, J. 1990. Geographic Information Systems: an Introduction . Prentice Hall. Englewood Cliffs.

Tufte, E.R. 1990. Envisioning Information. Graphics Press, Cheshire, Conn.

Creado el 31 de mayo de 2002. Artículo de GEOteca de la Universidad Autónoma de Madrid, España.