

Control de errores en Datos Geográficos

Carlos López Vázquez
carlos.lopez@ieee.org

Enfoque

- Existen cursos de Control de Calidad
 - Típicamente para Ingeniería Industrial o Química
- Métodos concebidos para productos tangibles
- Apuntan a:
 - Cadenas de producción en serie
 - Medir atributos de la Calidad
 - Detectar problemas rápidamente...
 - Actuar eliminando la causa
- Con datos es un poco más difícil...

Con datos...

- Muchas veces no salen de una *cadena*
 - Ejemplos:
 - Mapa de una ciudad
 - Catastro urbano
 - Contraejemplos:
 - Datos meteorológicos
 - Series o colecciones cartográficas
 - Imágenes satelitales
- Lo normal es que ya *estén* recogidos o generados
 - No incidiremos en su creación ☺

Problemas...

- Problemas principales:
 - ¿Cómo saber si un dato/registro está bien?
 - ¿Nos afecta?
 - Si está mal: ¿cómo asignarle un sustituto?
- Problemas secundarios:
 - ¿Cómo medir la Exactitud?
 - ¿Cómo comunicar la Calidad?
 - Una vez lograda, ¿cómo mantener la Calidad en una producción en serie?
- Por cierto... ¿Qué es *exactamente* Calidad?

Anecdotario...

- Ilustra algunos ejemplos de la vida real
- Barbas del vecino arder...
- Son ejemplos geográficos; hay muchos más en otras áreas

El BMW último modelo...

Fuente: P. Fisher (1999)

- En Alemania, 1998 un BMW se precipitó en el río Havel en el embarcadero de un ferry
- El dueño seguía fielmente las instrucciones de un sistema GPS con mapas instalado a bordo
- El mapa decía "puente" y debió decir "ferry"



Embajada china en Belgrado

- Bombardeada por la OTAN en 199X por error
- Los mapas decían que era una instalación serbia
- Había sido... ¡antes!
- Tres muertos, roces, etc.

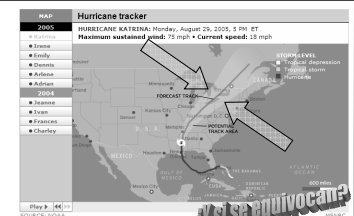


Derrame de petróleo

- Punta del Este, Uruguay en febrero de 1997
- Escollo de 24 mt. no figuraba en las cartas náuticas oficiales
- Un petrolero se lo llevó por delante



Huracán Katrina



Más casos marítimos

- 1980, Canadá. El *Sea Fever* pierde un marinero en una tormenta
- El pronóstico decía que la tormenta no llegaría donde llegó
- El error se debió a una boya meteorológica fuera de servicio
- El litigio por USS 3.2 millones fue ganado

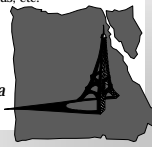


Fuente: Richard O. Mason (1986)

Hay casos más antiguos...

- El canal de Suez existió en la época faraónica
- Napoleón no lo reconstruyó porque se creía que $\Delta h \approx 6$ mts.
 - Esa diferencia requería esclusas, etc.
- En realidad son 80 cm

La India podría haber caído en otras manos... y la historia hubiese sido diferente



¿Ud. quisiera estar en esta lista?



Temario

- Definiendo la Calidad → 4 h
- Comunicando la Calidad → 4 hs
- Midiendo la Exactitud → 6 hs
- Mejorando la Exactitud → 12 hs
- Manteniendo la Calidad → 4 hs

Total previsto: 30 horas

¿Por qué tanto?

Definiendo la Calidad (4 hs.)

- Hay que comenzar por el principio
- Terminología, jerga, etc.
- Quizá sea familiar al público (en parte...)
- Dar contexto de datos geográficos
 - Estándares recientes
 - Informatización *pervasiva*
 - Revolución en varias profesiones
- Lo que es de *alta calidad* para uno ...

Comunicando la Calidad (4 hs.)

- ¿Alguien ha escuchado de la INDE?
- ¿Alguien ha escuchado de *metadatos*?
- ¿Alguien ha visto algún metadato?
- Veremos qué es, cómo interpretarlo y qué buscar en relación a Calidad
- Revisaremos los estándares vigentes



Midiendo la Exactitud (6 hs.)

- ¿Qué métodos o estándares se aplican?
- Caso posicional:
 - Los que saben son uds.
 - Revisaremos algún estándar vigente
- Caso temático:
 - Hay métodos pero ...
 - Veremos criterios académicos
- Servirá para saber
 - Si tenemos que preocuparnos (*problem dependent!*)
 - Si efectivamente *mejoramos...*



Mejorando la Exactitud (12 hs.)

- Tema *muy* nuevo
- Ámbito académico (1996, 1998,...)
- Grupo pequeño
- Tema *demasiado técnico*
 - Journals estadísticos no lo publican mucho
 - Journals temáticos (p.ej. Meteorológicos) lo ignoran
- ¿A quién le importa?
 - Productor del dato
 - Usuario del dato
 - (sufrido usuario debi decir...)
- Constituye el núcleo *duro* del curso



Mejorando la Exactitud (cont.)

- Las preguntas básicas son:
 1. ¿Cómo detectar un dato malo?
 2. ¿Cómo corregirlo?
- La segunda es más trillada: interpolación, mínimos cuadrados, etc.
- La primera... tiene cola
 - ¿Existe un (único) *valor correcto*?
 - ¿Es accesible? ¿A qué precio?
- ¿Cómo hago?
- ¿Lo que dijo Juan, no será para peor?



Manteniendo la Calidad (4 hs.)

- ¿Cómo asegurar niveles estables de error?
- Importante para:
 - Compradores de datos a proveedores externos
 - Productores de datos
 - Académicos (*of course...*)
- Tema nuevo en Agrimensura: típico y tradicional de Industriales o Químicos
- Veremos estándares, métodos, etc.

Referencias y materiales

- Ariza, 2002
- López *et al.*, 1999
- CD con material vario
 - PPT con "transparencias"
 - PDF consultados
 - Informes de proyecto
 - Tesis
 - Páginas y sitios web integrales
 - Quizá algo de software
 - Etc....

Módulo 1: Definiendo la Calidad

Carlos López Vázquez

carlos.lopez@ieee.org

Plan

- Definiendo la Calidad
- Exactitud vs. Precisión
- Componentes de la Calidad



Algunos problemas técnicos...

- Por ejemplo:
 - dar coordenadas a objetos
 - cómo crear representaciones digitales
 - cómo procesar esas representaciones
- Hay otros...
 - error, incertidumbre, escala, resolución...
- Complejivamente: *Calidad de datos*

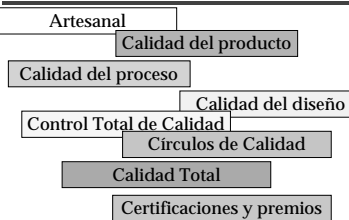
¿Qué es Calidad?

- RAE:

Propiedad o conjunto de propiedades inherentes a una cosa que permiten apreciarla como igual, mejor o peor que las restantes de su especie.
- Otra definición: ISO 9000

Grado en que un producto o servicio cumple con los requisitos especificados. No se limita a atributos, generalmente implica una relación de orden único aspecto.

Evolución histórica



Exactitud vs. Precisión

- Exactitud (Geodetic Glossary):

Es una cercanía de los resultados, cálculos o estimaciones a los valores verdaderos o valores que se acepta son verdaderos
- El idioma español...

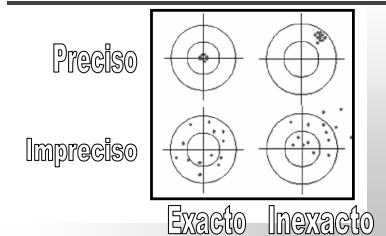
Precisión ≠ Exactitud
- Hay varias "Precisión"
 - Precisión (NCGIA Core):

Se define como el número de cifras decimales o de dígitos significativos de una medida
 - Precisión (Geodetic Glossary):

En estadística, una medida de la tendencia de un conjunto de números aleatorios a agruparse alrededor de un número determinado por el conjunto



Ilustrando Exactitud vs. Precisión



Más sobre Exactitud...

- A veces no hay "valor verdadero..."
- $E(x)$ no es lo mismo que $E(f(x))$...
 - Ej.: MDE vs. Pendientes
- Hay Exactitud *alta* y *baja*
- No son sólo medidas. Comparar:
 - Un valor exacto ($\alpha+\beta+\gamma=180$)
 - Una unidad convencional (metro patrón)
 - Un valor que se asume más exacto (monolito)

Precisión (NCGIA)

- Suele ser >>Exactitud
- Herencia informática...
 - Cifras significativas ≠ Cifras previstas
 - Es fácil confundirse
 - PC/ARC/INFO → Single Precision
 - Algunos GIS → Sólo Integer
- Es un error relativo
- Más relacionada con Resolución
- Usada por Burroughs, 1986 y otros
- No es la definición más moderna, ni la más apropiada

Precisión (Geodetic Glossary)

- Tiene en cuenta la repetibilidad
 - +Precisión ↔ -Varianza
- Un aspecto clave:
 - Precisión depende sólo del juego de datos
 - En cambio la Exactitud no
- Para confundir más...
 - "Nivelación de alta precisión"
- Es la definición que usaremos

Los problemas son...

- Cómo medir la Exactitud
 - ¡Repetir las medidas no ayuda!
- Cómo se propaga la Exactitud
- Cómo no atribuir más Exactitud de la correcta
 - Es un problema del lado del *usuario*

Otro concepto: Resolución

- Mezclable con Precisión (NCGIA)
- Se distinguen tres casos:
 - Espacial
 - Temática
 - Temporal
- No abundaremos...

Plan

- ✓ Definiendo la Calidad
- ✓ Exactitud vs. Precisión
- Componentes de la Calidad



Componentes de la Calidad

- Recogida en varios estándares
 - CEN/TC 287
 - ISO/TC 211
 - USGS 1994
 - ISO 19113
- Hay muchos elementos comunes
 - Exactitud Posicional
 - Exactitud Temática
 - Consistencia lógica
 - Completitud
 - Linaje



Exactitud Posicional

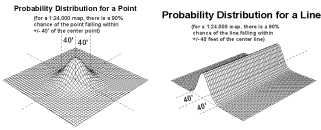
- Definida como la cercanía a la posición verdadera
- Usualmente involucra coordenadas
- Convención: 0.5 mm a la escala del plano
 - Ej.: $0.5 / (1/10000) = 0.5 \cdot 10^4 \text{ mm} = 50 \text{ m}$
 - En general: 0.5 / escala
 - ¿Escala? ¡Regla para el mundo en papel!
- Cifras significativas...
Ej.: 1:24.000; (x,y)=(123456.789,987654.321) mts.
Error=0.5 / (1/24000)=12.0 mts → sólo valen algunos dígitos (1.2345x10⁵; 9.8765x10³)

¿Y si no hay/hubo papel?

- Se debe declarar la "Exactitud"
- ¿Cómo?
 - La posición de los puntos difiere de la verdadera en menos de X mts.
 - Pregunta: ¿Cuales puntos? ¿Todos?
 - USGS: En el 90% de los casos, la posición difiere de la verdadera en menos de X mts.
 - Aparece el percentil 90%
 - No asume una distribución a priori
 - BSI: La desviación estándar del error es menos de X mts.
- ¿Cómo interpretar la declaración?
- ¿Cómo verificarla?

Cómo interpretar...

■ Según USGS



Cómo verificarla...

■ Tres alternativas:

- Usar una fuente de mayor Exactitud
 - GPS
 - Mapa de mayor escala
 - Volver a los datos originales
 - Usar evidencia interna
 - Tamaño de los defectos
 - Calcular la Exactitud propagando...
 - 0.5 mm por aquí, 1 mm por allá...
- Será desarrollado en detalle luego

Exactitud de Atributos

- También definida como *cercanía al valor verdadero*
- Atención: el tiempo ahora pesa más
 - Aparece la Vigencia
- ¿Cómo se expresa la Exactitud? Depende del tipo de dato:
 - Atributos continuos
 - Ej.: MDE, pluviometría, etc.
 - Parecido al caso posicional
 - Ej.: *La elevación tiene un error menor a 1 mt*
 - Podrían estar categorizados (imágenes RGB, etc.)
- Atributos categóricos

Para Atributos Categóricos

■ Definir "cercanía" es diferente...

- Está definido el "#", pero no el "-"
- Podría estar definido el ">"
- Hay otras diferencias
 - ¿Existe un "verdadero valor"?
 - ¿Cómo son las categorías?
 - ¿Apropiadas?
 - ¿Suficientemente detalladas?
 - ¿Bien definidas? ¿Nitidas?
- También será analizado en detalle luego

Cómo verificarla...

■ Hay varias alternativas

- Construyendo una *matriz de confusión*
 - Presume una inspección binaria
 - Existen índices para caracterizarla
 - Usando Métodos Difusos
 - La inspección ahora no es binaria
 - Aún bajo investigación
 - Otros índices
 - Los hay, pero son algo simples...
 - En cualquier caso, se hace Prueba de Hipótesis
- Aún no está recogido en los estándares

Consistencia Lógica

■ ¿Es la BD consistente con sus definiciones?

- Geométricas:
 - Una etiqueta y sólo una por polígono
 - Todo polígono con etiqueta
 - Todo punto es de un polígono
 - Ningún arco que se cruza con otro sin cortarse
 - Tiene topología
 - Temáticas:
 - Los atributos están dentro del rango
- La Consistencia Lógica se mide en % de cumplimiento

Más sobre Consistencia...

- ¿Fueron *todas* las capas obtenidas...?
 - Misma escala, mismas hipótesis, mismos procesos...
- ¿Fueron *todas* las capas editadas...?
 - Mismos criterios (overshoots, tolerancias,...)
- ¿Fueron *todas* las capas obtenidas simultáneamente?
 - Hora, estación, año, etc. (lo que corresponda)

Completitud

- Difícilmente medible
- No hay una definición formal
 - "...Dado el Modelo Cartográfico, la BD y la Realidad, los objetos que *deberían estar efectivamente están...*"
- Calidad depende del Modelo Cartográfico
 - Ej.: Incluir líneas de 30kV o más
 - La BD está completa, aunque falten las de 6.3kV
- Podría ser por Atributos
 - El objeto está, pero incompleto
 - Ej.: Falta la profundidad de un pozo
- No hay métricas consensuadas

Linaje

- Es un registro de historia de la BD
- Responde a preguntas como:
 - ¿Cómo fue digitalizada? ¿A partir de qué documentos?
 - ¿Cuándo fueron recogidos los datos base?
 - ¿Qué agencia u organización recolectó los datos?
 - ¿Qué pasos u etapas se utilizaron para procesar los datos originales?
 - ¿Con qué precisión fueron efectuados los cálculos (o qué error tenían los resultados numéricos)?
- El Linaje es usualmente un indicador útil de Exactitud
- No hay métricas consensuadas

Hay otros también relevantes...

- Vigencia
- Accesibilidad
- Pertinencia o Relevancia

- Son considerados en algunos estándares de metadatos
- Veremos de qué se trata

Vigencia

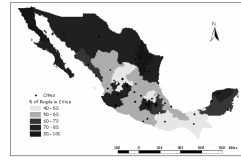
- ¿Están los datos actualizados?
 - Ej: uso de tierra con fines de agricultura



Formato OK, Resolución OK, Completitud OK etc....

Otro ejemplo...

- Población urbana, México 1990



¿1965 vs. 1990? Puede ser mejor 1965...

Accesibilidad...

- Que los datos existan...



- Burocracia, formatos exóticos, etc.

Pertinencia o Relevancia

- Ciertos datos pueden no aportar nada...
- A veces se descubre; a veces no
 - Modelos importados
 - Condiciones locales
- Estadística tramposa; un ejemplo:
 - Idea: Búsqueda de individuos con mayor Respuesta a promociones
 - ≠ Ídem con mayor propensión a pagar!!!

Módulo 1: Definiendo la Calidad

Carlos López Vázquez

carlos.lopez@ieee.org