

Facultad de Ingeniería

ISO 19157 - Información geográfica - Calidad de datos

Hebenor Bermúdez - Miguel Gavirondo

Octubre 2024



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Control posicional

Como vimos para la evaluación de la calidad posicional de un conjunto de datos geográficos es muy frecuente realizarla a partir de METODOLOGÍAS DE CONTROL POSICIONAL POR PUNTOS.

Entre estas se encuentran las medidas D.29 a D.56 de la norma ISO 19157 y los estándares NMAS, ASLSM o NSSDA.

Para realizar las evaluaciones es necesario recurrir a fuentes de mayor exactitud. No siempre es necesario relevar datos en campo.

NSSDA

El **National Standard for Spatial Data Accuracy** de la **Federal Geographic Data Committee** es de uso obligatorio en las agencias federales de Estados Unidos y uno de los más reconocidos en el mundo.

El estándar no acepta o rechaza un producto sino que simplemente declara la calidad posicional absoluta de un conjunto de datos.

NSSDA

El estándar se basa en el **Error Medio Cuadrático** que se calcula considerando las diferencias entre las coordenadas del punto a evaluar y el punto asumido como verdadero.

La componente Z se maneja de forma independiente a las coordenadas X e Y.

El estándar diferencia en el caso de las Z entre territorios sin vegetación y con vegetación. En el caso de territorios con vegetación la normalidad en los errores es muy difícil de cumplir por lo que se recurre a los percentiles.

NSSDA

Requiere algunas consideraciones especiales:

- Distribución adecuada de los puntos de control.
- Control en los errores calculados
 - Detección y eliminación de valores atípicos.
 - No presencia de sistematismos.
 - Normalidad de los errores en cada componente.
 - Independencia entre los errores de la componente X e Y.
 - Aleatoriedad de los errores en cada componente.

Tamaño muestral

Se pueden utilizar los procedimientos vistos para determinar el tamaño muestral para hacer el control posicional.

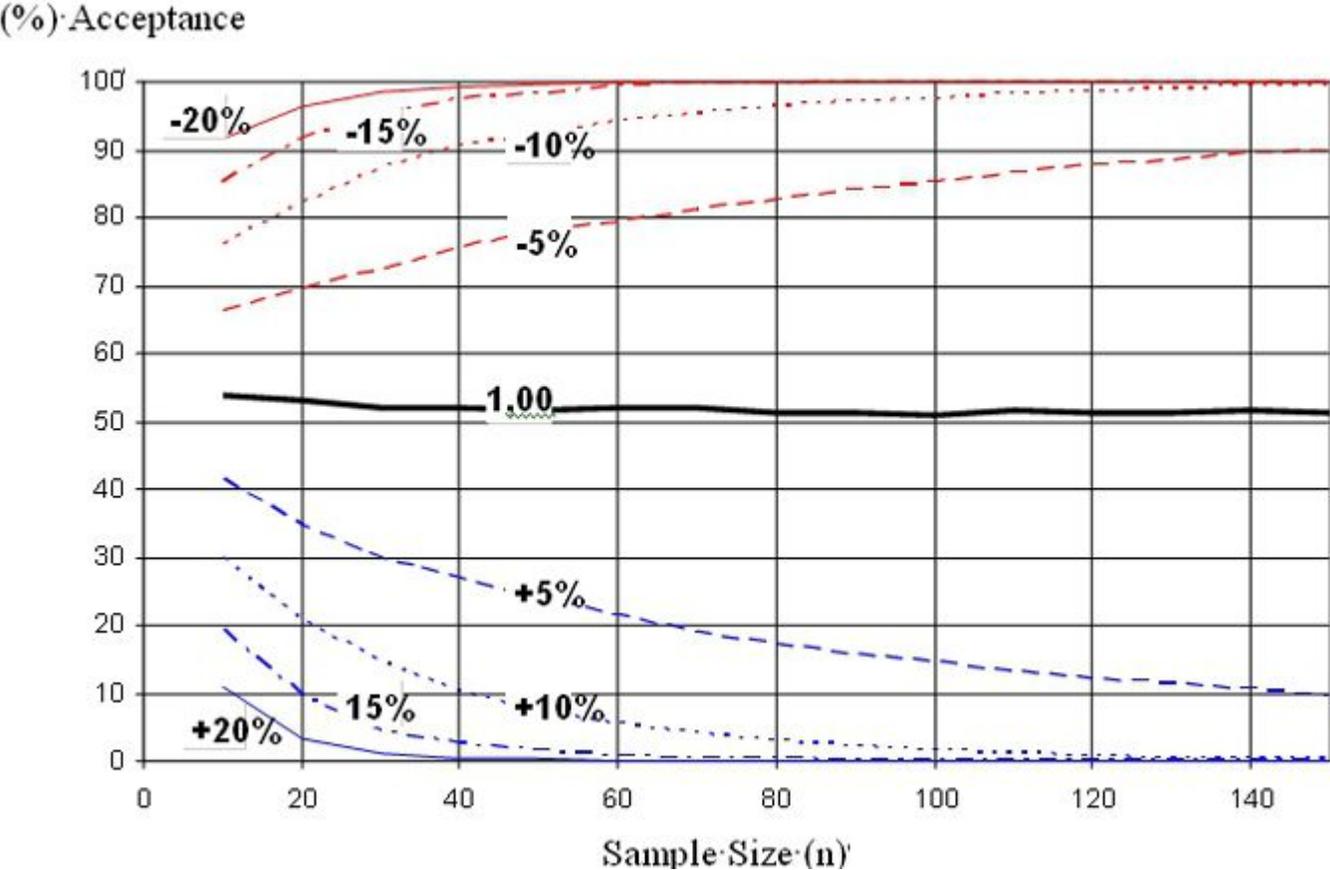
TABLE C.1 RECOMMENDED NUMBER OF CHECKPOINTS BASED ON AREA

Project Area (Square Kilometers)	Horizontal Accuracy Testing of Orthoimagery and Planimetrics	Vertical and Horizontal Accuracy Testing of Elevation Data sets		
	Total Number of Static 2D/3D Checkpoints (clearly-defined points)	Number of Static 3D Checkpoints in NVA ⁹	Number of Static 3D Checkpoints in VVA	Total Number of Static 3D Checkpoints
≤500	20	20	5	25
501-750	25	20	10	30
751-1000	30	25	15	40
1001-1250	35	30	20	50
1251-1500	40	35	25	60
1501-1750	45	40	30	70
1751-2000	50	45	35	80
2001-2250	55	50	40	90
2251-2500	60	55	45	100

⁹Although vertical check points are normally not well defined, where feasible, the horizontal accuracy of lidar data sets should be tested by surveying approximately half of all NVA check points at the ends of paint stripes or other point features that are visible and can be measured on lidar intensity returns.

Ariza et al. Estudiaron el tema de tamaño muestral con la perspectiva del usuario y del productor.

Tamaño muestral



Control posicional de modelos digitales de elevación

El problema que se presenta al controlar un MDE es que los puntos no están claramente identificados principalmente si se trata de un raster o son puntos equiespaciados.

Esto implica que no se puede hacer coincidir exactamente el punto relevado con el que corresponde en el modelo o requiere de procedimientos de interpolación.

Siempre hay que elegir el método de interpolación más elemental y que mejor se adapte a la realidad del terreno.

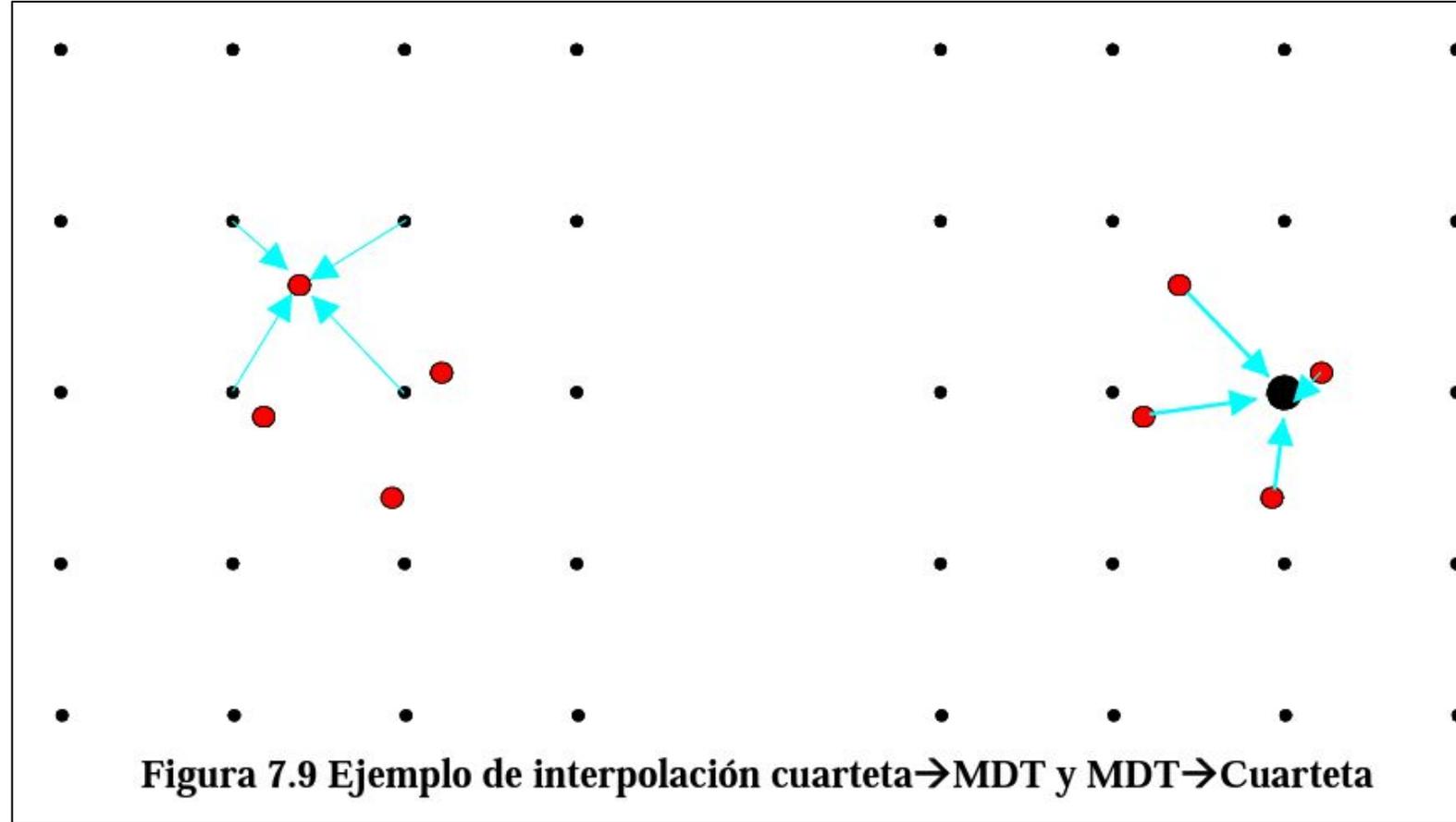
Control posicional de modelos digitales de elevación

La principal condición es que los puntos de control se tomen sobre superficies de terreno consideradas planas (pendiente constante) y con un tamaño de al menos 2 veces el tamaño de la celda del MDE.

En campo se debe:

- buscar el plano de pendiente constante. La proyección horizontal de esta superficie debe ser equivalente a un cuadrilátero de lado el doble del paso de malla.
- Levantar cuatro puntos con una metodología adecuada formando un cuadrilátero de aproximadamente el lado del paso de malla.

Control posicional de modelos digitales de elevación



Lectura recomendada

MEDIDAS

- **EMC absoluto en altitud con interpolación sobre el MDE.** Se obtiene de las diferencias entre las altitudes medidas y las interpoladas del MDE.
- **EMC absoluto en altitud con interpolación sobre puntos del terreno.** Se obtiene de las diferencias entre la altitud del MDE y las interpoladas a partir de los puntos de campo.
- **EMC relativo en altitud con interpolación sobre el MDE.** Se calcula comparando desniveles entre las altitudes derivadas de campo y los desniveles obtenidos a partir de la interpolación del MDE.
- **EMC relativo en altitud con interpolación sobre los puntos del terreno.** Se calculará comparando desniveles obtenidos de las altitudes del MDE con desniveles obtenidos a partir de las interpolaciones de campo.

Componente temática

El primer punto y muy importante que hay que definir en estos elementos es el denominado **UNIVERSO DE DISCURSO**.

Esto es así porque en el proceso de evaluación no se compara la cartografía con la realidad sino que debo hacerlo frente al **terreno nominal** y que se puede definir como el **subconjunto de elementos del mundo real que cumplen con las especificaciones del usuario** respecto del producto.

Componente temática

Tabla 7.15 Proceso de abstracción: desde el mundo real al terreno nominal	
Mundo real ⇓	Es la realidad en toda su complejidad. Es el conjunto de elementos, características, relaciones y hechos que pueden ser o no conocidos por el ser humano.
Mundo conceptual o universo abstracto ⇓	El discurso humano no tiene lugar al nivel del mundo real, puesto que mediante nuestro lenguaje natural, dando nombres a las cosas, abstraemos la realidad. De esta manera se conforma un mundo conceptual, o universo abstracto, como el conjunto de entidades percibidas de la realidad que sirve como marco de referencia para la obtención de un conjunto de datos.
Mundo Geoespacial ⇓	El universo abstracto no es suficientemente adecuado para su plasmación en BDG. El modelo geoespacial, con sus restricciones y peculiaridades, establece una nueva simplificación o abstracción que facilita el tratamiento de la información, conformándose así en mundo geoespacial
Universo de discurso ⇓	Visión del mundo real, o hipotético, que incluye todo lo que es de interés para una aplicación geoespacial concreta. Esta visión está condicionada por el proceso de abstracción que genera el mundo conceptual y por las restricciones que impone el modelo geoespacial y por los requisitos de la aplicación concreta (especificaciones de producto).
Terreno nominal	Concepción del mundo real derivada de las especificaciones de un conjunto de datos geográficos; por tanto es la plasmación del universo de discurso como selección de elementos del mundo real. El terreno nominal constituye el conjunto ideal de datos geográficos, con el que se compara el conjunto real, la BDG, para evaluar su calidad. En ocasiones puede venir definido por otro producto cartográfico, siendo recomendable que su calidad esté contrastada.

Componente temática

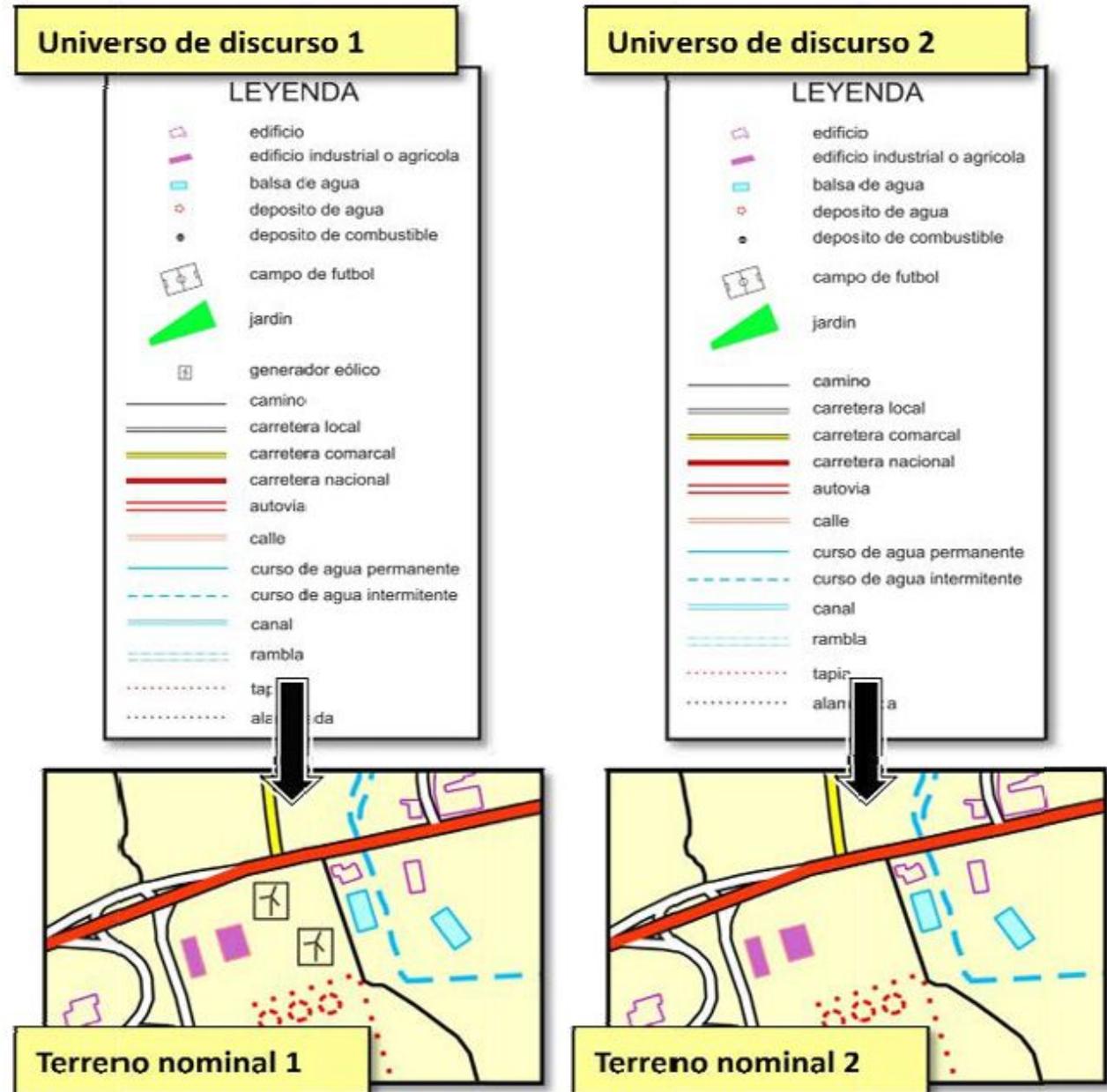


Figura 7.21 Ejemplo de dos universos de discurso que dan lugar a terrenos nominales diferentes

Componente temática



Figura 7.22 Visión del mundo real del universo de discurso 1 (izquierda) y del universo de discurso 2 (derecha)

Componente temática

La **Matriz de error en la clasificación** (medida D.65 de la norma ISO 19157) y **sus índices** (ver tabla D.67) asociados son los principales elementos para evaluar la corrección en la clasificación.

Los elementos de la matriz corresponden a los elementos de la clase X clasificados como Y para cada una de las posibles combinaciones definidas por el número de clases. La verdad terreno esta en las columnas y el conjunto de datos en las filas.

- Las clases deben ser independientes, mutuamente excluyentes, exhaustivas y en número suficiente.
- Deben usarse métodos de muestreo que no introduzcan correlación.
- Conviene usar estratificación.

Componente temática

Tabla D.65 – Matriz de error de la clasificación

Línea	Componente	Descripción																														
1	Nombre	matriz de error de la clasificación																														
2	Alias	matriz de confusión																														
3	Nombre del elemento	corrección de la clasificación																														
4	Medida básica																															
5	Definición	matriz que indica el número de ítems de la clase (i) clasificados como clase (j)																														
6	Descripción	<p>la matriz de error de la clasificación (MCM, misclassification matrix) es una matriz cuadrada de n columnas y n filas. n indica el número de clases consideradas</p> <p>$MCM(i,j) = [n^\circ \text{ de ítems de la clase } (i) \text{ clasificados como clase } (j)]$</p> <p>los elementos de la diagonal de la matriz de error de clasificación contiene los ítems clasificados correctamente, y los elementos fuera de la diagonal contienen el número de errores de la clasificación</p>																														
7	Parámetro	<p>nombre: n</p> <p>definición: número de clases consideradas</p> <p>tipo de valor: entero</p>																														
8	Tipo de valor	entero																														
9	Estructura del valor	matriz ($n \times n$)																														
10	Fuente de referencia	–																														
11	Ejemplo	<table border="1"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="3">Clase del conjunto de datos</th> <th rowspan="2">Recuento</th> </tr> <tr> <th>A</th> <th>B</th> <th>C</th> </tr> </thead> <tbody> <tr> <th rowspan="4">Clase verdadera</th> <th>A</th> <td>7</td> <td>2</td> <td>1</td> <td>10</td> </tr> <tr> <th>B</th> <td>1</td> <td>2</td> <td>2</td> <td>5</td> </tr> <tr> <th>C</th> <td>1</td> <td>1</td> <td>3</td> <td>5</td> </tr> <tr> <th>Recuento</th> <td>9</td> <td>5</td> <td>6</td> <td>20</td> </tr> </tbody> </table>			Clase del conjunto de datos			Recuento	A	B	C	Clase verdadera	A	7	2	1	10	B	1	2	2	5	C	1	1	3	5	Recuento	9	5	6	20
		Clase del conjunto de datos			Recuento																											
		A	B	C																												
Clase verdadera	A	7	2	1	10																											
	B	1	2	2	5																											
	C	1	1	3	5																											
	Recuento	9	5	6	20																											
12	Identificador	62																														

Componente temática

Tabla 7.20 Ejemplo de matriz de confusión

Unidad Cartográfica	Clase (Verdad) Terreno					Total
	1	2	...	v-1	v	
1	x_{11}	x_{12}	...	$x_{1,v-1}$	x_{1v}	n_{1+}
2	x_{21}	x_{22}	...	$x_{2,v-1}$	x_{2v}	n_{2+}
.
.
.
c-1	$x_{c-1,1}$	$x_{c-1,2}$...	$x_{c-1,v-1}$	$x_{c-1,v}$	$n_{c-1,+}$
c	$x_{c,1}$	$x_{c,2}$...	$x_{c,v-1}$	$x_{c,v}$	$n_{c,+}$
Total	n_{+1}	n_{+2}	...	$n_{+,v-1}$	$n_{+,v}$	N

Tabla 7.21 Notación seguida

N	número de casos (celdas-píxeles)
M	número de clases
V	número de clases "verdad terreno" ($V = M$)
C	número de clases "unidades cartográficas" ($C = M$)
i	filas, unidades del mapa o producto
j	columnas, verdad-terreno
V_i	Clase verdad-terreno
C_i	Clase unidad cartográfica
x_{ij}	número de casos en $C_i V_j$
n_{i+}	total casos C_i
n_{+j}	total casos V_j
P_{ij}	proporción en $C_i V_j$ ($= x_{ij}/N$)
P_{i+}	proporción C_i ($= n_{i+}/N$)
P_{+j}	proporción V_j ($= n_{+j}/N$)

