

Determinación de la calidad de OpenStreetMap (OSM) para la Comunidad de Madrid

Antonio Villena Martín¹, Antonio F. Rodríguez Pascual²

¹ Universidad Politécnica de Madrid
a_villenam@hotmail.com

² Jefe de Área de Infraestructura de IG
IGN
afrodriguez@fomento.es

Resumen

El presente estudio se ocupa de la determinación de la calidad de la cartografía de uso libre Open Street Map (OSM) en España, concretamente en la Comunidad de Madrid, conforme a las normas ISO 19100 aplicables (ISO 19113 Principios de Calidad, ISO 19114 Métodos de Evaluación de la Calidad e ISO 19138 Medidas de calidad de datos) y teniendo en cuenta tanto las Especificaciones de la Base Topográfica Armonizada BTA v1.0 aprobada por el Consejo Superior Geográfico, como los estudios de calidad realizados en la comunidad OSM (Muki y otros).

Palabras clave: Openstreetmap, calidad, ISO 19100.

1 Introducción

OSM es un proyecto colaborativo de Neocartografía en el que se genera un conjunto de datos de callejero y red viaria que cubre todo el mundo para su publicación en la web mediante servicios de visualización y descarga. Los datos son libres bajo licencia CC: by-sa 2.0¹. Se han considerado únicamente los viales,

¹ <http://creativecommons.org/licenses/by-sa/2.0/deed.es>

es decir las carreteras y calles ya que son los objetos geográficos más relevantes para OSM. OSM dispone de objetos de otro tipo, como ríos, ferrocarriles, lugares, edificios, etc. que cumplen el papel de elementos de referencia y solo aparecen en los núcleos de población más importantes.

2 Ámbito

El estudio que hemos realizado se halla encuadrado en la Comunidad de Madrid, considerándose tres zonas de tipología diferenciada, tales como:

- Zona de alta densidad de población. Hoja MTN50: 0559 - Madrid
- Zona de edificación disgregada. Hoja MTN50: 0581 – Móstoles
- Zona de baja densidad de población. Hoja MTN50: 0533 – San Lorenzo del Escorial

Hemos seleccionado esta zona debido al conocimiento que tenemos de la Comunidad de Madrid y a que es una de las áreas más desarrolladas por OSM y en la que hemos creído conveniente determinar la calidad de los datos por su interés temático y riqueza de información.

3 Tareas realizadas

Las tareas que se han realizado, atendiendo a la norma ISO 19113 Principios de calidad, son:

- Identificar que parámetros, medidas y métodos de calidad de datos se van a utilizar.
- Seleccionar una muestra representativa.
- Elegir los datos de referencia.
- Evaluar la calidad de un conjunto de datos, aplicando lo anterior.
- Obtener conclusiones.

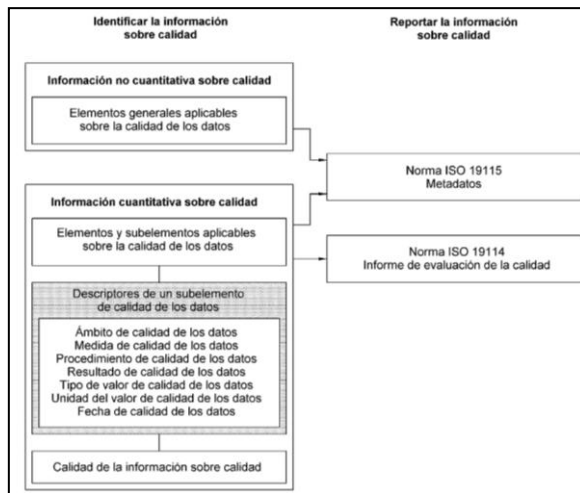


Figura 1. Visión general de la información sobre la calidad de los datos

3.1 Elección de parámetros de calidad

De la norma ISO 19113 se han extraído una serie de componentes para describir y valorar cuantitativamente el grado de adecuación con que un conjunto de datos se ajusta a su universo de discurso². Los elementos y subelementos, teniendo en cuenta aspectos espaciales y temáticos elegidos han sido:

- Compleción
 - Comisión
 - Omisión
- Exactitud posicional
 - Exactitud absoluta horizontal (externa)
- Exactitud temática
 - Corrección de la clasificación
 - Corrección de atributos no cuantitativos (nombre geográfico)

La metodología, a grandes líneas ha sido la siguiente:

- Preparación del conjunto de datos
- Análisis completación

² Visión del mundo real, o hipotético, que incluye todo aquello que es de interés.

- Análisis de exactitud posicional
- Análisis de la exactitud temática

3.2 Evaluación de la calidad OSM-CartoCiudad

Los datos geográficos usados para hacer la determinación de la calidad han sido los de CartoCiudad³, para lo que se han descargado las capas correspondientes en formato shape disponibles en el Centro de Descargas del CNIG⁴. Además de CartoCiudad se ha usado el PNOA⁵ como apoyo en determinadas zonas que presentaban situaciones poco claras.

Según la norma ISO19113 se recomienda presentar la información sobre la calidad en un formato breve, fácilmente comprensible y recuperable, ya que, en ocasiones, la cantidad de información puede exceder la del conjunto de datos. Por lo que presentamos las siguientes tablas resumen de los estudios que hemos realizado.

Los elementos según la norma ISO 19113 aplicables que hemos seleccionado son los que aparecen en la siguiente tabla 1.

Tabla 1. Resumen de información cuantitativa pertinente sobre calidad.

Elementos de la calidad de datos	Subelementos de la calidad de datos	¿Pertinente?
compleción	comisión	si
	omisión	si
consistencia lógica	consistencia conceptual	no
	consistencia del dominio	no
	consistencia del formato	no
	consistencia topológica	no
exactitud posicional	exactitud absoluta o externa	si

³ <http://www.cartociudad.es/portal/1024/index.htm>

⁴ <http://centrodedescargas.cnig.es/CentroDescargas/>

⁵ <http://www.ign.es/PNOA/>

	exactitud relativa o interna	no
	exactitud posicional de datos en malla	no
exactitud temporal	exactitud de una medida del tiempo	no
	consistencia temporal	no
	validez temporal	no
exactitud temática	corrección de la clasificación	si
	corrección de atributos no cuantitativos	si
	exactitud de atributos cuantitativos	no

Para cada elemento de los anteriormente vistos, clasificado como «pertinente», se ha realizado la metodología resumida en la tabla 2.

Tabla 2. Valoración de la información cuantitativa sobre la calidad.

Párrafos relevantes de la especificación del producto para documentar la aplicabilidad	Elemento/subelemento aplicable	Ámbito(s) de la calidad identificado(s)
Porcentaje de viales de OSM que no están representados en CartoCiudad ni en PNOA y que corresponden a carreteras interurbanas y calles.	compleción comisión/omisión	8 áreas de muestreo por cada hoja del MTN50 considerados en el análisis
Porcentaje de vías interurbanas de OSM que omiten el atributo «ref» comparando con en el «mapa oficial de carreteras 2011».	compleción omisión	el conjunto de datos
Porcentaje de entidades lineales de OSM del total del ámbito de las hojas MTN50 tomadas que no disponen de atributo «nombre».	compleción omisión	el conjunto de datos
EMC de las diferencias entre los viales de OSM con CartoCiudad medidos en las intersecciones de las entidades con las retículas de 8x8 para cada hoja del MTN50.	exactitud posicional exactitud absoluta horizontal (externa)	30 puntos en horizontal y 30 puntos en vertical homogéneamente distribuidos de las intersecciones de las entidades con una retícula de 8x8 para cada hoja del MTN50

Porcentaje de las tipologías «type» y nomenclaturas «ref» de los viales de OSM erróneamente clasificados comparando con el «mapa oficial de carreteras 2011».	exactitud temática corrección de la clasificación	el conjunto de datos
Porcentaje de los nombres de vías urbanas de OSM erróneamente asignados comparando con CartoCiudad que intersectan con una diagonal trazada en cada área de muestreo de las hojas del MTN50	exactitud temática corrección de atributos no cuantitativos (nombre geográfico)	8 áreas de muestreo por cada hoja del MTN50 considerados en el análisis.
Porcentaje de entidades de población de OSM erróneamente definidos comparando con la base de datos del INE.	exactitud temática corrección de atributos no cuantitativos (nombre geográfico)	el conjunto de datos

3.3 Resultados de la determinación de la calidad

Los resultados de la determinación de la calidad se han reportado según se indica en la norma ISO 19114 en informes.

A continuación mostramos un cuadro resumen con los resultados generales del análisis en la tabla 3.

Tabla 3. Valoración de la información cuantitativa sobre calidad.

Elemento/subelemento aplicable	Ámbito(s) de la calidad identificado(s)	Resultado	Resultado esperado	Conformidad
compleción omisión	8 áreas de muestreo por cada hoja del MTN50 considerados en el análisis	26%	$\leq 2\%$	No conforme
compleción omisión	8 áreas de muestreo por cada hoja del MTN50 considerados en el análisis	15%	$\leq 4\%$	No conforme
compleción omisión atributo «ref»	el conjunto de datos	73%	$\leq 4\%$	No conforme
Compleción omisión atributo nombre vial	el conjunto de datos	73%	$\leq 4\%$	No conforme
exactitud posicional exactitud absoluta horizontal (externa)	30 puntos en horizontal y 30 puntos en vertical homogéneamente distribuidos de las intersecciones de las entidades con una retícula de 8x8 para cada hoja del	EMC 4,85 m	EMC ≤ 12 m (1/100.000)	Conforme

	MTN50			
exactitud temática corrección de la clasificación atributo <i>ref</i>	el conjunto de datos	99%	$\geq 96\%$	Conforme
exactitud temática corrección de la clasificación atributo <i>type</i>	el conjunto de datos	39%	$\geq 96\%$	No conforme
exactitud temática corrección de atributos no cuantitativos (nombre geográfico)	8 áreas de muestreo por cada hoja del MTN50 considerados en el análisis.	64%	$\geq 95\%$	No conforme
exactitud temática corrección de atributos no cuantitativos (nombre geográfico)	el conjunto de datos	89%	$\geq 95\%$	No conforme

Dado que en OSM se utilizan trazadores GPS de 10 m de precisión para tomar datos de posición, se toma como hipótesis y aplicando el Teorema del Muestreo que el detalle menor registrado es del orden de 20. Si esa dimensión se supone que es equivalente al límite de percepción visual (0,2 mm) podemos suponer que la escala aproximada de OSM es de 1:100.000. Si la BTA toma como límite el 60% de 0,2 mm a escala (0,60 m a 1:5.000), el límite para la exactitud posicional admisible a 1:100.000 será de 12 metros. Incluso si se toma como escala nominal de OSM 1:50.000, lo que en algunas zonas puede ser más realista, la exactitud medida estaría por debajo del límite de 6 metros.

Hemos consultado otros estudios como el de *Muki*⁶ sobre Inglaterra, en el que han comparando OSM con la cartografía *Meridian 2*, y han obtenido un 20% de omisión. También tenemos otro estudio de *ITO World*⁷ de ámbito Reino Unido tomando como referencia la cartografía de *OS Locator* en el que han obtenido un 15% para la omisión.

3.4 Conclusiones

Como conclusión podríamos decir que, atendiendo a los resultados del análisis que hemos efectuado, OSM se puede analizar mediante las especificaciones de la BTA v1.0 con ciertos matices o adaptaciones, ya que están pensadas para cartografías escala 1: 5.000 y nosotros hemos considerado una escala para OSM de 1:100.000.

Los 4,85 m de EMC nos dan idea de la excelente exactitud de elementos lineales analizados para cartografías de escalas 1:50.000 o incluso 1:25.000, además de la inmensa cantidad de información de que disponemos. El análisis que hemos

⁶ http://www.ucl.ac.uk/~ucfamha/OSM%20data%20analysis%20070808_web.pdf

⁷ http://www.itoworld.com/product/data/osm_analysis/main

realizado (BTA v1.0) es muy estricto y está pensado para escalas mucho mayores (1:5.000). Los datos de OSM representan la realidad probablemente a una escala de 1:100.000 (como hemos visto más arriba) por lo que no está dentro de la tolerancia para casi ningún elemento de las normas ISO 19113 para la BTA. Pero si extrapolamos la BTA a escala 1:100.000 o 1:50.000 la exactitud posicional debería estar entre 6 y 12 m, con lo que OSM entraría en tolerancia.

Sabemos que CartoCiudad tiene una mayor exactitud en zonas urbanas que en interurbanas, y en OSM es todo lo contrario, ya que en campo abierto los dispositivos de captura de datos de las trazas GPS disponen de mayor precisión al tener más cobertura de satélites en zonas interurbanas, mientras que en zonas urbanas, debido al apantallamiento producido por los edificios la cobertura es menor y en conclusión menos precisa. Todo esto es aplicable si atendemos a los datos cuya forma de captura ha sido mediante dispositivos GPS. Esto nos lleva a la conclusión de que teniendo un método de captura de información tan sencillo y barato, se obtienen unas exactitudes en posición para una escala entre 1:100.000 y 1:50.000 excelentes, con un EMC de 4,85 m.

Un aspecto destacable de OSM es que en lo relativo a caminos rurales supera en compleción a CartoCiudad y esto en el análisis de la comisión nos ha dado un resultado poco fiable (26%), ya que hemos comprobado que OSM dispone de muchas carreteras y calles que aún CartoCiudad no contempla. Es decir que la comisión de OSM en vez de ser un resultado negativo, en este caso es algo positivo ya que OSM presenta una información actualizada casi al 100% y CartoCiudad no nos serviría para este elemento de nuestro análisis.

En cuanto a la omisión obtenida del 15% (similar a los estudios de *Muki* e *ITO World*) es uno de los puntos débiles de OSM, ya que hay zonas bastante escasas, sobre todo las de baja densidad de población como la zona de San Lorenzo del Escorial (21%) o Móstoles (21%) en contraposición de zonas de mayor densidad de población como Madrid capital (5%).

Otro de los puntos débiles de OSM es la exactitud temática o semántica, siendo bastante escasa en cuanto a los nombres de viales (64%) y en nombres de población (89%).

Tenemos que señalar que el conjunto de datos de CartoCiudad hereda los datos del Dirección General del Catastro, Instituto Nacional de Estadística, Sociedad Estatal de Correos y Telégrafos e Instituto Geográfico Nacional. CartoCiudad no contempla las tildes ni las minúsculas en nombres, ya que el INE (su principal fuente en los nombres de calles) no las tiene en cuenta, poniendo todos los nombres

en mayúsculas, por lo que en el análisis hemos obviado estos aspectos y no los hemos considerado en los cálculos.

Hay que apuntar que la *web* de OSM no es muy usable en lo relativo a descarga de datos. Por ejemplo para la descarga del *planet* con los *shapes* de España se han realizado 6 pasos hasta llegar al enlace de descarga y no han sido fáciles de encontrar, en cambio, en CartoCiudad, para llegar al enlace de descarga de los *shapes* de la Comunidad de Madrid han sido necesarios 9 pasos, pero encontrarlos ha sido muy intuitivo y sencillo.

Finalmente concluimos que OpenStreetMap puede ser usada para aplicaciones que no requieran de una exactitud temática elevada de una manera libre, gratuita y con unas exactitudes aceptables teniendo en cuenta la modalidad de captura de la información geográfica.

A la vista de los resultados obtenidos se considera aconsejable el uso de OSM, según escala y exactitud, para:

- Cartografía para países subdesarrollados (proyecto Mapping Bolivia⁸).
- Estudios previos en proyectos de obra civil.
- Base de datos geográfica para SIG, cálculos de rutas de flotas, etc.
- Estudios previos de actualización de cartografía oficial.

OSM es un proyecto amateur que por ello, no sigue los estándares más extendidos de adquisición de datos, verificación y uso, pero porque es una de sus filosofías y no entra dentro de sus objetivos.

4 Capturas de pantalla de ejemplo

En estas capturas de pantalla las líneas representadas en rojo se corresponden con los viales de OSM, las líneas en azul a CartoCiudad, y las que están en amarillo son las que se encuentran seleccionadas.

Compleción

⁸ <http://mappingbolivia.blogspot.com/>

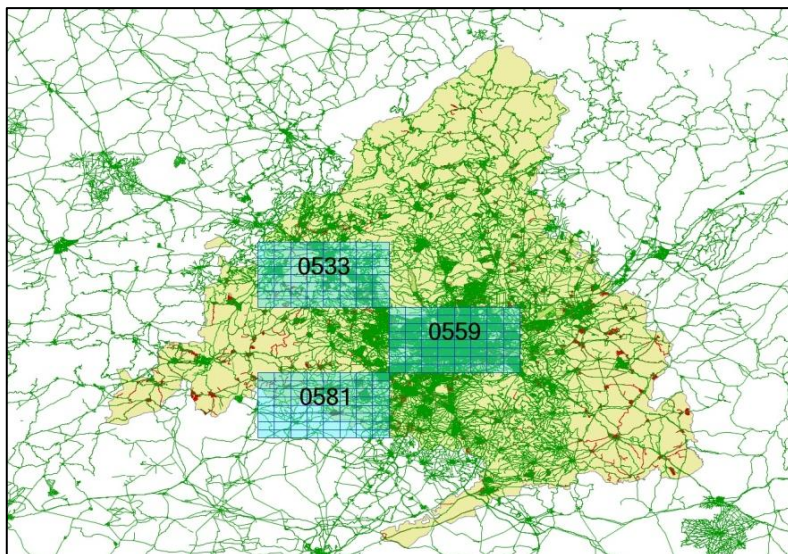


Figura 2. Esquema general de la muestra tomada

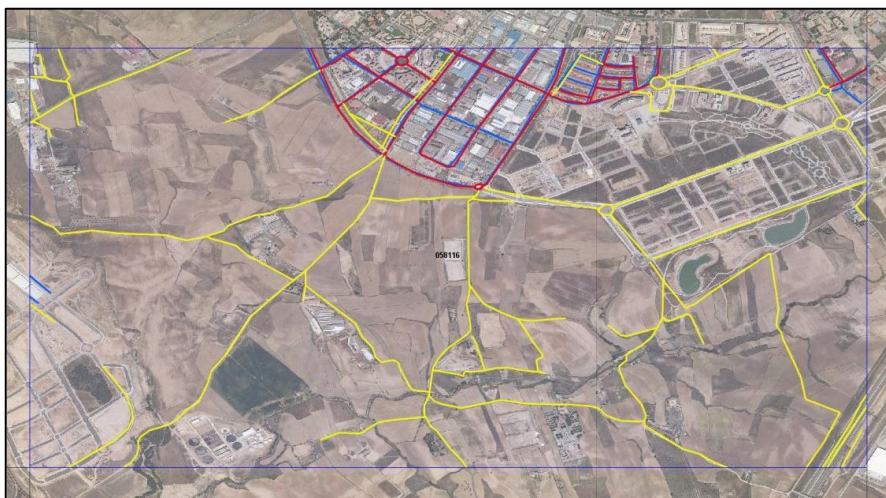


Figura 3. Compleción – Comisión hoja MTN50- 0581 área 16

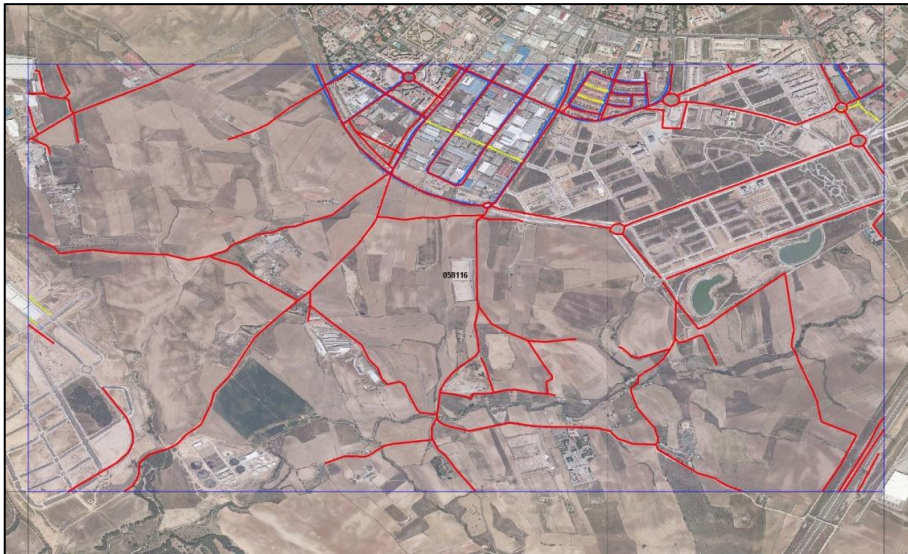


Figura 4. Compleción – Omisión hoja MTN50- 0581 área 16

Exactitud posicional

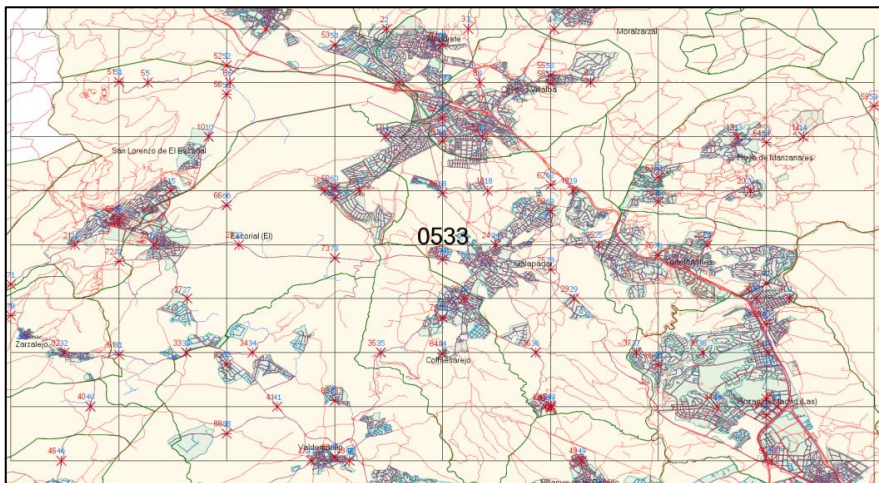


Figura 5. Exactitud posicional. Esquema distribución puntos en hoja MTN50-0533

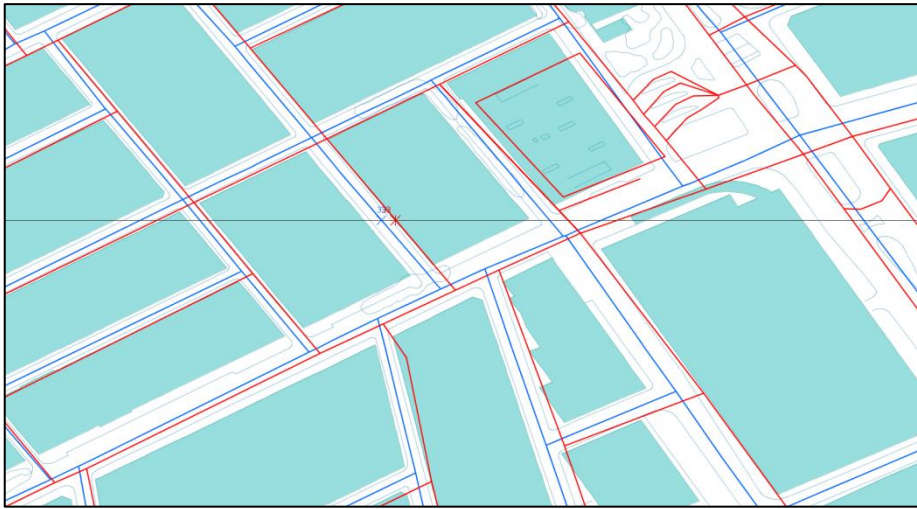


Figura 6. Exactitud posicional. Detalle de punto medido en malla horizontal

Exactitud temática

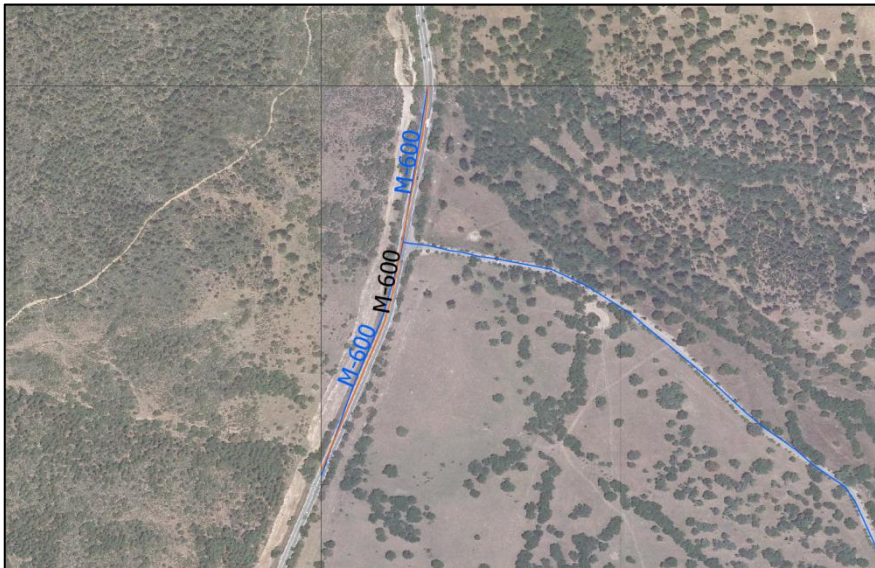


Figura 7. Exactitud temática. Detalle de comparación de clasificación del atributo «ref»

Agradecimientos. Antonio F. Rodríguez Pascual (Jefe Área de Infraestructura de IG, IGN), Iván Sánchez Ortega (OpenStreetMap España), CartoCiudad.

Referencias

- [1] Ariza López, Francisco Javier, Rodríguez Pascual, Antonio Federico: Introducción a la normalización en Información Geográfica: la familia ISO 19100 v2. Grupo de Investigación en Ingeniería Cartográfica, Universidad de Jaén. (2008)
- [2] AENOR. UNE-EN ISO 19113. Información geográfica. Principios de Calidad. (2005)
- [3] AENOR. UNE-EN ISO 19114. Información geográfica. Procedimientos de evaluación de la calidad. (2005)
- [4] AENOR. UNE-EN ISO 19138. Información geográfica. Medidas de Calidad de los Datos. (2004)
- [5] Rodríguez Pascual, Antonio Federico. Servicios OGC y Neocartografía en la IDEE. (2008)
- [6] <http://www.openstreetmap.org/>
- [7] <http://www.cartociudad.es/portal/1024/index.htm>
- [8] <http://www.ign.es/PNOA/>