

Sistema de Información de Ocupación del Suelo en España

Manual de control de calidad SIOSE

Versión 3.1

Editor	Equipo Técnico Nacional SIOSE
Fecha	5 de marzo de 2015
Unidad	D.G. Instituto Geográfico Nacional. Observación del Territorio. S.G de Geodesia y Cartografía Servicio de Ocupación del Suelo
Descripción	Documento que describe los procesos de control de calidad de la base de datos SIOSE
Documentos Relacionados	“Documento técnico SIOSE” “Descripción del modelo de datos SIOSE” “Manual de fotointerpretación SIOSE” “Estructura y consulta de la base de datos SIOSE” “Metodología de actualización SIOSE”
Diseminación	Pública
Web	www.siose.es
Idioma	ES

CONTROL DE VERSIONES

Nº	Fecha	Comentarios
1	06/01/2012	Primera Versión
2	03/12/2012	Segunda versión, actualización de contenidos
3	05/03/2015	Revisión
3.1		Actualización nomenclatura documentación

ACRÓNIMOS

Acrónimo	Definición
IGN	Instituto Geográfico Nacional
ISO	International Organization for Standardization / Organización Internacional para la Normalización
SIOSE	Sistema de Información de Ocupación del Suelo en España
UTM	Universal Transversa de Mercator / Universal Transverse Mercator

ÍNDICE

1.	INTRODUCCIÓN	1
2.	ASPECTOS A REVISAR.....	2
2.1	TIPO I: ASPECTOS GENERALES.....	2
2.1.1	Convención de nombres.....	2
2.1.2	Formato de entrega	2
2.1.3	Unidades de producción	2
2.1.4	Sistema geodésico de referencia.....	2
2.1.5	Proyección cartográfica	2
2.1.6	Contenido de las bases de datos.....	2
2.1.7	Unidad mínima a representar.....	3
2.1.8	Características geométricas.....	3
2.1.9	Límites ficticios	3
2.1.10	Control de los polígonos	4
2.1.11	Metadatos.....	4
2.2	TIPO II: CONTROL TOPOLÓGICO.....	4
2.2.1	Entidades con geometría distinta a la de tipo polígono	4
2.2.2	Arcos colgantes – extremos libres	4
2.2.3	Resolución de anclajes.....	4
2.2.4	Huecos y solapes entre polígonos	4
2.2.5	Puntos superpuestos y repetidos.....	5
2.2.6	Bucles en elementos perimetrales	5
2.3	TIPO III: CASADO ENTRE BLOQUES.....	5
2.3.1	Coincidencia geométrica y semántica.....	5
2.3.2	Tratamiento de elementos perimetrales.....	5
2.4	TIPO IV: CONTROL GEOMÉTRICO Y SEMÁNTICO DE POLÍGONOS DE CAMBIO.....	5
2.4.1	Control geométrico de la fotointerpretación.....	6
2.4.2	Control semántico de la fotointerpretación	7
3.	RESULTADO DEL CONTROL DE CALIDAD E INFORMES	7
	ANEXO A: METODOLOGÍA DEL CONTROL DE CALIDAD GEOMÉTRICO	8

1. INTRODUCCIÓN

Los requisitos técnicos del SIOSE y su estructura organizativa hacen necesario considerar desde un principio un Sistema de Gestión y Control de Calidad de la Producción SIOSE que deberá:

- Asegurar el cumplimiento de los estándares técnicos de calidad del SIOSE
- Asegurar el cumplimiento de plazos previstos durante la fase de producción
- Establecer mecanismos ágiles de comunicación entre los equipos de trabajo de producción y control de las comunidades autónomas, y la Dirección nacional del proyecto.
- Poner a disposición de los organismos participantes información actualizada sobre el estado del proyecto, los controles de calidad realizados y datos disponibles, provisionales o definitivos, resultado del mismo.

Por lo tanto, será prioritario definir con claridad los procesos productivos del SIOSE, y sus controles de calidad sobre los productos resultantes; es decir, el flujo de trabajo en la producción y control de calidad del SIOSE. Estos procesos quedarán establecidos mediante un conjunto de procedimientos comunes para todos los equipos productivos de las comunidades autónomas, y luego serán consensuados y especificados en detalle para cada comunidad considerando las particularidades de la producción del SIOSE en la misma.

El Manual de control de Calidad, redactado de acuerdo a la metodología final del SIOSE, se considera como una de las actividades del procedimiento a seguir dentro del Sistema de Gestión y Control de Calidad de la Producción para asegurar el cumplimiento de los requisitos técnicos establecidos para el proyecto SIOSE y que, por tanto, deberán cumplir todos los productos SIOSE generados por los equipos de producción autonómicos.

En el Manual se describen los procesos de control que serán realizados a las bases de datos SIOSE entregadas por los equipos de producción y que serán evaluados por la Dirección Técnica del proyecto SIOSE, a lo largo del proceso de producción del proyecto y al final del mismo para conseguir un producto final homogéneo y de calidad.

2. ASPECTOS A REVISAR

Para la totalidad de las bases de datos entregadas, correspondientes a las unidades de producción (o conjunto de unidades de producción), se revisarán los siguientes apartados en el proceso de control de calidad, según se define en la “Guía Técnica SIOSE2009”

2.1 TIPO I: ASPECTOS GENERALES

2.1.1 Convención de nombres

Las bases de datos entregadas deberán seguir las normas de nomenclatura de ficheros definida para el proyecto SIOSE.

2.1.2 Formato de entrega

Las bases de datos entregadas cumplirán los formatos de entrega definidos para el proyecto SIOSE, siguiendo su modelo de datos, tanto en estructura como en contenido.

2.1.3 Unidades de producción

La entrega final constituirá la información completa SIOSE por comunidad autónoma, con información continua y sin costuras (sin huecos ni solapes), en un único fichero o en varios (bloques) y sin cambios con los límites externos de la base de datos SIOSE de partida.

2.1.4 Sistema geodésico de referencia

Para la creación de la base de datos SIOSE se empleará el sistema geodésico ETRS89, según las recomendaciones dictadas por el Consejo Superior Geográfico, atendiendo a los requisitos determinados por la Directiva Europea INSPIRE.

2.1.5 Proyección cartográfica

El sistema de representación cartográfica será el oficial: la proyección conforme Universal Transversa de Mercator (UTM) establecida como reglamentaria por el Decreto 2303/1970, en los husos 28, 29, 30 y 31, dependiendo de la zona y en el huso al que corresponda mayoritariamente el territorio en cada comunidad autónoma, en el caso de estar entre dos husos, se entregará en huso 30.

2.1.6 Contenido de las bases de datos

Las bases de datos entregadas deberán contener todas y cada una de las tablas y entidades definidas por el modelo de datos del proyecto SIOSE MF2 y acorde al Modelo de datos conceptual SIOSE.

2.1.7 Unidad mínima a representar

El tamaño mínimo del polígono a representar estará en función del tipo de superficie de la que se trate:

- Coberturas de agua, coberturas húmedas, cultivos forzados, playas, dunas y arenales, acantilados marinos y vegetación de ribera: 0,5 ha
- Superficies artificiales: 1 ha
- Zonas agrícolas, forestales y resto de clases de vegetación natural: 2 ha

Todos los polígonos representados deberán cumplir con el tamaño mínimo según lo anteriormente descrito.

Para poder determinar si un polígono con una cobertura compuesta es correcto en cuanto a su superficie, según las restricciones de unidad mínima, se considerará la cobertura simple más restrictiva que lo compone. Es decir, si alguna de las coberturas simples existentes pertenece a alguna de las categorías anteriores, se considerará todo el polígono correspondiente a esa categoría.

2.1.8 Características geométricas

Como norma general El ancho mínimo a representar como polígono será igual o superior de 15 m.

Asimismo, se comprobará que los polígonos no formen estrangulamientos o pasillos de ancho menor o igual a 15m y que tengan longitud igual o mayor de 60m.

Nota Estas normas generales no serán aplicables en los casos de coberturas de red viaria o ferroviaria o coberturas que puedan tener un tamaño mínimo de polígono de 0.5 ha), dónde no se aplicarán las limitaciones generales de ancho mínimo de elementos lineales a 15m.

2.1.9 Límites ficticios

Las bases de datos finales entregadas no tendrán bordes de hoja ni límites ficticios (límites de término municipal, de provincia, etc) dentro de cada comunidad autónoma.

2.1.10 Control de los polígonos

Se comprobará que:

- Todos los polígonos deberán estar asociados a una cobertura y todas las coberturas deben tener un polígono asociado
- El porcentaje (%) de superficie cubierta de cada polígono debe ser del 100%.
- No deben existir polígonos contiguos con un identificador o cobertura asociada que presente la misma cobertura y atributos totalmente coincidentes y en los mismos porcentajes.

2.1.11 Metadatos

Se comprobará que para los metadatos se ha seguido el manual de Metadatos SIOSE, basado en las recomendaciones y normativa CSG: NEM v1.1 e ISO TC211 (19115-2:2009) e INSPIRE.

El formato de los metadatos generados deberá poderse importar a la aplicación CatMDEdit. (Formato XML).

2.2 TIPO II: CONTROL TOPOLÓGICO

2.2.1 Entidades con geometría distinta a la de tipo polígono

En las bases de datos SIOSE sólo se permitirá la existencia de entidades de tipo polígono.

2.2.2 Arcos colgantes – extremos libres

No habrá arcos colgantes o extremos libres en la geometría final del proyecto SIOSE. Será necesario que estén resueltas las intersecciones entre elementos, estando estos conectados mediante nodos.

2.2.3 Resolución de anclajes

Entre polígonos adyacentes los nodos deberán ser coincidentes

2.2.4 Huecos y solapes entre polígonos

Los polígonos no presentarán solapes ni huecos entre ellos, debiendo ser las líneas comunes entre polígonos adyacentes exactamente las mismas

2.2.5 Puntos superpuestos y repetidos

No existirá el mismo nodo superpuesto y repetido en los elementos perimetrales de los polígonos, consecuencia de la digitalización del mismo nodo varias veces.

2.2.6 Bucles en elementos perimetrales

Los elementos perimetrales de los polígonos no presentaran bucles superponiéndose en algún punto a sí mismo.

2.3 TIPO III: CASADO ENTRE BLOQUES

2.3.1 Coincidencia geométrica y semántica

Deberá existir una coincidencia exacta tanto geométrica como semántica de la información entre los diferentes ficheros de entrega por comunidad autónoma.

2.3.2 Tratamiento de elementos perimetrales

Todos los elementos perimetrales deberán estar cerrados sobre el límite de la unidad de producción.

2.4 TIPO IV: CONTROL GEOMÉTRICO Y SEMÁNTICO DE POLÍGONOS DE CAMBIO

El control geométrico y semántico realizará sobre una muestra representativa de los polígonos de cambio o revisión de las bases de datos SIOSE generadas.

La muestra de polígonos de cambio se realizará de acuerdo a los siguientes criterios:

- Se realizará estratificación dividiendo la población en 5 estratos según el nivel 1 de CLC00.
- Se seleccionarán una serie de puntos de forma aleatoria que corresponderán a una serie de polígonos de la base de datos, de tal forma que, la superficie de los polígonos seleccionados sea el un porcentaje significativo de la superficie de cada estrato.
- La muestra de polígonos obtenida se utilizará tanto para el control geométrico como para el semántico.

2.4.1 Control geométrico de la fotointerpretación

Sobre los polígonos de la muestra, se seleccionan un número significativo de puntos de control geométrico, pudiendo aplicar para la selección de puntos representativos de la muestra la metodología descrita en el Anexo A.

Sobre estos puntos, se controlará la precisión geométrica de la base de datos del proyecto SIOSE mediante ajuste correcto de las líneas a la imagen de referencia, teniendo en cuenta los errores admitidos en el mismo.

El **control geométrico** se realizará comparando las coordenadas de la imagen SPOT5 con las coordenadas de la base de datos del punto más cercano al obtenido en el muestreo que sea reconocible tanto en la imagen como en la base de datos (bordes de parcelas, cruces, etc.).

De esta forma se obtendrá para cada punto controlado los residuos en x e y, así como el error longitudinal en cada punto y el error medio cuadrático para la muestra de la base de datos considerada. Estos parámetros se obtendrán de la siguiente forma:

$$E_L = \sqrt{XR^2 + YR^2}$$

$$RMSE = \sqrt{\frac{\sum (XR_i^2 + YR_i^2)}{n-1}}$$

siendo:

RMSE = Raíz cuadrada del error medio cuadrático. Se expresará en unidades de longitud.

n = nº total de puntos de control de la muestra.

XR_i = Residuo en X del punto de control i. Se expresará en unidades de longitud.

YR_i = Residuo en Y del punto de control i. Se expresará en unidades de longitud.

Con esta información se generará la Tabla I de Control geométrico

Tabla I. Control geométrico

Nº Punto	X SPOT5	Y SPOT5	X SIOSE	Y SIOSE	XR	YR	EL
1							
2							
3							
...							
n							

Se considera que el error longitudinal (EL) de cada punto que forman la muestra de la base es ≤ 5m, aunque se aceptarán bases con puntos de EL > 5m, siempre que representen menos de un 2% de los puntos de la muestra.

Lo **errores groseros** se definen como tres veces el valor del error longitudinal manejado como límite anteriormente, por lo tanto 15m. Se aceptarán bases con presencia de errores groseros (≥ 15m), siempre y cuando estos puntos representen menos de un 0,2% de la muestra.

2.4.2 Control semántico de la fotointerpretación

El control semántico de la fotointerpretación se realizará sobre los polígonos de la muestra, con el fin de detectar:

- Errores en la asignación total o parcial del tipo de cobertura
- Errores en la asignación del tipo de cobertura
- Errores por omisión o exceso de coberturas representativas (mayores del 5%) en un cierto polígono
- Errores en la asignación del porcentaje de coberturas (%)
- Errores en la asignación de atributos
- Cambios no detectados

3. RESULTADO DEL CONTROL DE CALIDAD E INFORMES

El control de calidad realizado sobre los datos SIOSE quedará reflejado en su correspondiente informe, que recogerá tanto la metodología seguida como los resultados de los aspectos revisados previamente descritos.

Cada base de datos revisada superará el control de calidad si cumple que:

- No presenta errores de Aspectos generales
- No presenta errores topológicos
- No presenta errores de casado entre bloques en el caso de entrega en distintos bloques colindantes.
- En la muestra de polígonos de cambio, el control geométrico, la raíz cuadrada del error medio cuadrático (RMSE) de todos los puntos que forman la muestra de la toda la base es $\leq 3\text{m}$.
- En la muestra de polígonos de cambio, el control semántico es aceptable.

ANEXO A: METODOLOGÍA DEL CONTROL DE CALIDAD GEOMÉTRICO

Como complemento al Manual de Calidad aquí expuesto, se ha visto interesante hacer referencias a la metodología llevada a cabo en el proceso del control de calidad geométrico. Debido a las pocas experiencias que se han realizado en el campo de la comprobación geométrica superficial, este trabajo puede ilustrar para futuros desarrollos de este aspecto.

La exactitud geométrica de un producto cartográfico se puede determinar mediante el análisis de lo ocurrido en una muestra significativa de puntos del producto. El chequeo de estos puntos, se hace mediante comparación de sus coordenadas planimétricas en el producto, junto con las coordenadas de una fuente de mayor exactitud.

En este caso concreto del Sistema de Información sobre Ocupación del Suelo en España, la asignación de las coberturas homogéneas se fotointerpretan desde una imagen SPOT5 de 2,5 m de resolución espacial. El control de calidad geométrico del proyecto SIOSE se realiza comparando las coordenadas de los puntos de la base de datos de polígonos digitalizados, con las coordenadas del punto homólogo en la imagen SPOT5, es decir, los datos sobre los que se realiza el control son de tipo vectorial, y la fuente de mayor exactitud es una imagen.

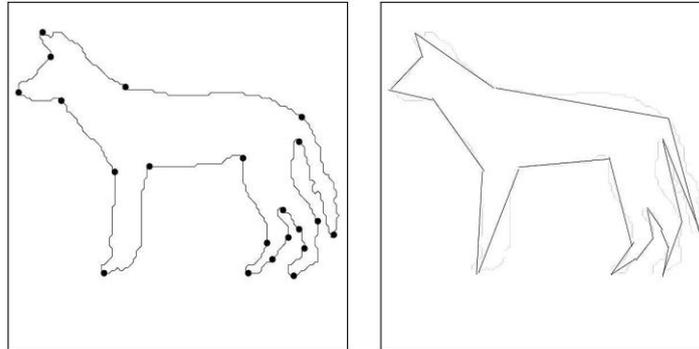
Pero la diferencia entre formatos de información no resulta la mayor problemática. El aspecto más complejo es manejar elementos superficiales, como son los polígonos SIOSE, junto con las demás armas de evaluación estadística y de la calidad, que se centran mayoritariamente en cálculos unidimensionales, o cartográficamente hablando, basados en puntos. Por ello, la tarea más revolucionaria de este estudio es evaluar una base de información geográfica de elementos superficiales. Aunque en última instancia se recurran a cálculos unidimensionales, el proceso de obtención de los puntos sobre los que calcular supone un pequeño reto, el cual es resuelto reduciendo las dimensiones de estudio según avanza el proceso de control de calidad.

Las superficies son estudiadas mediante sus contornos, las líneas. Este enfoque basado en geometrías lineales es más complejo en ejecución, cálculo y gestión que el proceso clásico basado en puntos de control. Existen estudios en los que se evalúa la calidad cartográfica de elementos lineales basándose en el concepto de “banda o halo de indeterminación”, pero todos ellos requieren disponer de herramientas de emparejamiento de elementos lineales homólogos, que no siempre existen. Esta circunstancia ocurre en este caso, se disponen de elementos vectoriales a evaluar, pero no elementos vectoriales de mayor confianza con los que comparar.

El problema planteado inicialmente consiste en determinar qué puntos, de las líneas que conforman los polígonos SIOSE, deben ser empleados en el estudio para la evaluación de la calidad geométrica, de forma que la muestra sea representativa de la geometría de cada polígono de la base de datos.

El método de selección de puntos de control geométrico elegido, se basa en investigaciones psicológicas sobre el sistema visual humano, que han demostrado que los puntos de máxima curvatura en el perímetro de una figura, juegan un papel importante en el proceso de reconocimiento de su contorno. El sistema visual humano es capaz de extraer y procesar las características visuales no redundantes, de tal forma que se pierde poca información.

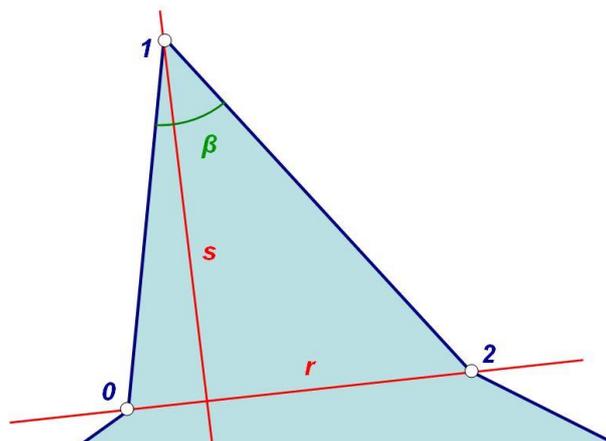
En 1954 el psicólogo F. Attneave propuso que “*el sistema visual humano percibe los puntos de máxima curvatura como la información importante de las figuras*”, es decir, estos puntos son discontinuidades en los bordes como curvas, ángulos, esquinas y en general cualquier punto en el que una línea recta cambia de dirección. Dichos puntos de máxima curvatura son menos redundantes que los propios bordes, y contienen más información que líneas con curvatura cero, o mínima curvatura, lo que lleva a pensar que las esquinas pronunciadas, generan respuestas fisiológicas más intensas que las esquinas poco pronunciadas.



En la primera silueta de un perro se han marcado los puntos de máxima curvatura, que unidos mediante líneas rectas sirven para generar la segunda silueta, fácil de reconocer e identificar con la figura original (Ghosh y Petrov 2006).

A pesar de que este principio es fácilmente entendible, su implementación dentro de una herramienta numérica que evalúe cada punto del perímetro de un polígono no resulta tan trivial. Existen expresiones numéricas que emparejan de forma directa el valor de la curvatura de un elemento lineal con esta representabilidad (por ejem.: Feldman and Sing, 2005), pero para fácil y rápido manejo dentro de una cadena productiva, como es la del SIOSE, se optó por un método más sencillo (Tragsatec, 2008).

A grandes rasgos, este método se basa en el análisis de tres puntos secuenciales dentro del contorno del polígono ($0, 1, 2$). Sobre 0 y 2 se traza la recta que los une r , y sobre el punto 1 se traza la recta s que siendo perpendicular a r pasa por él. A partir de las coordenadas de todos los puntos es posible calcular los valores de los lados y ángulos del triángulo. El ángulo β sustentado en 1 juega un papel importante, ya que es el que encierra, en su valor y su signo, la información de la curvatura en el punto 1 . Calculándolo sobre todos los puntos del polígono, se hace posible seleccionar aquellos puntos del perímetro que posean valores de curvatura interesante.



Una vez que se poseen los puntos representativos de los polígonos, se hace una muestra estadística con todos ellos, tal y como se menciona en páginas anteriores de este manual. Consecuentemente se registraran sus coordenadas en la capa vectorial, así como se hace medida interactiva de las coordenadas en la imagen SPOT5 mediante fotointerpretación de cada punto. Esta identificación en la imagen no resulta muy complicada ya que la mayoría de puntos de alta curvatura pertenecen a límites visibles entre coberturas.

Comentando el análisis estadístico, su primer paso es dotar a los diferentes puntos de un valor de error, utilizando la contrastación entre la digitalización y la imagen SPOT5. Se calculan los errores en ambas direcciones coordenadas, X e Y , a estos valores se les denomina dentro de este Manual de Control de Calidad, "*Residuos en X e Y de los puntos de la muestra*". Pero para sintetizar estos valores y manejarlos más fácilmente dentro de los análisis estadísticos, se opta por generar la componente cuadrática de ambos en cada punto. De este concepto se obtiene el valor del "*Error Longitudinal de cada punto de la muestra*".

Para evaluar la dispersión de los valores del error longitudinal, se aplica la desviación típica muestral de los errores longitudinales, debido a su gran difusión y sencilla aplicación en un proceso de control de calidad. La explicación por la cual se fija un valor que represente a toda la muestra es la de evaluar en conjunto todas las bases de datos cartográficas, y restringir aquellas bases que a pesar de no poseer errores groseros, tengan errores que en conjunto no alcancen el nivel de calidad esperado para el proyecto. Se incluye en el control de calidad el parámetro llamado "*Raíz cuadrada del error medio cuadrático, RMSE*".

Para definir el valor limitante de la RMSE se tienen en cuenta dos aspectos fundamentales dentro de la elaboración de las bases, como son la imagen de referencia y la escala del producto. Partiendo de un tamaño de píxel de 2.5m y una escala final de 1:25.000 cuyo error máximo aceptable en la representación es de 5m, se acordó establecer un límite a la RMSE de 3m.

Por otro lado, analizando de forma separada cada error longitudinal de los puntos, se ve necesario incluir restricciones en sus valores para no contener dentro del producto final errores muy elevados. Se establecen dos niveles de análisis, uno de ellos a los 5m, por ser el error máximo aceptable en la representación cartográfica a esta escala, y otro definido como "*errores groseros*" para delimitar aquellos muy discrepantes. Este segundo nivel se establece a los 15m, por ser tres veces el valor del anterior valor. Para estos nuevos valores, también se debe establecer el porcentaje máximo de errores dentro de la muestra. Existen diversas técnicas estadísticas encaminadas a la detección y limitación de porcentajes dentro de una población de valores. Pero ante todo, y debido a la naturaleza práctica que debe tener el control de calidad, prima la sencillez y homogeneidad del método de análisis. Existen alternativas a implementar en el proceso, entre ellas las basadas en análisis individual y conjunto de los errores longitudinales, así como las más estrictas y metódicas que necesitan la utilización de distribuciones estadísticas. Pero éstas no garantizaban un valor limitante independiente del número de puntos a manejar en cada base, y cada base analizada tendría resultados diferentes. Por ello se optó por un estudio empírico sobre las primeras experiencias y a partir de él se establecieron los límites a los errores longitudinales de 5m y 15m, fijados en 2% y 0.2% respectivamente.