

Evaluación de la calidad posicional de conjuntos de datos geográficos

LÍNEA DE BASE PARA LA CUANTIFICACIÓN DE LA MEJORA

Ing. Agrim. HEBENOR BERMÚDEZ

**Departamento de Geomática, Instituto de
Agrimensura, Facultad de Ingeniería, UdelaR**

Montevideo, Agosto de 2019

ÍNDICE

MOTIVOS	4
METODOLOGÍA	5
ZONAS CENSALES RURALES	8
Resultados para el total de los puntos	8
Control posicional relativo	11
Control posicional absoluto por puntos	14
Detección de outliers	14
Control de normalidad	15
Control de aleatoriedad	16
Control de independencia	16
Control de sistematismo	17
National Standard for Spatial Data Accuracy (NSSDA)	18
Resultados	19
CAMINERÍA	20
Resultado para el total de los puntos	20
Control posicional relativo	23
Control posicional absoluto por puntos	26
Detección de outliers	26
Control de normalidad	27
Control de aleatoriedad	28
Control de independencia	28
Control de sistematismo	29
National Standard for Spatial Data Accuracy (NSSDA)	30
Resultados	31
CATASTRO RURAL	32
Resultados para el total de los puntos	32
Control posicional relativo	36
Control posicional absoluto por puntos	38
Detección de outliers	38
Control de normalidad	39
Control de aleatoriedad	40
Control de independencia	40
Control de sistematismo	41
Resultados	43

PLAN CARTOGRÁFICO 1/50000	44
Resultados para el total de los puntos	46
Control posicional relativo	49
Control posicional absoluto por puntos	51
Detección de outliers	51
Control de normalidad	52
Control de aleatoriedad	53
Control de independencia	54
Control de sistematismo	54
National Standard for Spatial Data Accuracy (NSSDA)	55
Resultados	56
ANÁLISIS GENERAL	57
BIBLIOGRAFÍA	60

MOTIVOS

Existen en Uruguay muchos conjuntos de datos geográficos contruidos principalmente por organizaciones estatales con anterioridad al Vuelo Fotogramétrico Nacional (VFN) con el objetivo de servir a sus fines. A pesar de ser conjuntos de datos geográficos muy utilizados, pocos de ellos cuentan con evaluación de la calidad en algunas de sus dimensiones. Según Méndez y López (2010) en nuestro país no ha habido un proceso sistemático de evaluación de la exactitud, si bien existen antecedentes aislados. Estos antecedentes se han centrado exclusivamente en el aspecto posicional y se han aplicado a pequeñas extensiones territoriales sin abarcar completamente el universo del conjunto de datos (por ejemplo 1 hoja del PCN 1:50000).

El VFN es un punto de inicio para la actualización de los conjuntos de datos geográficos existentes. Esta actualización permitirá generar un salto de calidad ya que se estará usando una base única, por lo que resulta necesario para productores y usuarios tener información que permita cuantificar dicha mejora. Este trabajo busca fijar la línea de base para permitir cuantificar el salto cuantitativo en cuanto a la calidad de la información en su aspecto posicional.

Según Ariza (2009), “..las medidas son la base de los procesos de mejora..”, Por lo cual toda evaluación debería verse como un insumo para los productores de forma que se implementen procesos de mejora continua en su producción.

Se espera con esta evaluación comparar un grupo de datos espaciales considerados básicos con uno de los principales productos del VFN como son las ortoimágenes para el ámbito Nacional. Con este trabajo se espera analizar en términos cuantitativos dar una perspectiva de su calidad posicional.

En tercer lugar, los productos de datos preexistentes van a coexistir -hasta tanto no se hagan ajustes- con los productos del VFN los cuales han sido ampliamente controlados. Este trabajo busca servir de base para medir la interoperabilidad geográfica que los productos existentes presentan con las ortoimágenes.

METODOLOGÍA

El control que se plantea es únicamente posicional planimétrico para los siguientes conjuntos de datos geográficos.

1. Zonas censales rurales del Instituto Nacional de Estadística.
2. Parcelario rural de la Dirección Nacional de Catastro.
3. Caminería Nacional del Ministerio de Transporte y Obras Públicas.
4. Plan Cartográfico 1:50000 del Servicio Geográfico Militar.

El control se realizará considerando como verdad terreno las ortoimágenes de 32 centímetros de tamaño de pixel obtenidas del VFN. Para esto se elegirán puntos para el control posicional planimétrico absoluto y relativo en los diferentes conjuntos de datos geográficos y sus homólogos correspondientes en las ortoimágenes obtenidas del VFN proporcionadas por la IDEUy a través del visualizador Topovision.

El control se realizará en la cuenca del Río Santa Lucía por ser este el ámbito geográfico para el que se entregó (por parte de Topocart y se aceptó por parte de la IDEUy), en primera instancia, los productos principales del Vuelo Fotogramétrico Nacional. Esta cuenca cubre un área de aproximadamente 13500 km² (7.6% del territorio terrestre de Uruguay) distribuida en 6 departamentos: Montevideo, Canelones, San José, Flores, Florida y Lavalleja.

El control para la aceptación de los productos del VFN se hicieron de forma estadística utilizando la norma española UNE 148002:2016 y las ISO 2859. Se entiende entonces que los datos están aceptados por lo que cumplen con los niveles de calidad establecidos de 1 metro al 95% de confianza.

Para realizar el control se seleccionarán 100 pares de puntos homólogos para cada conjunto de datos a evaluar. Este número de puntos de control surge de las afirmaciones de Ariza y Atkinson (2008) que para el caso del estándar NSSDA se llegaría al 95% de nivel de confianza con una muestra de aproximadamente 100 puntos.

Para identificar los puntos se utilizará una herramienta SIG de distribución aleatoria de puntos. Con esto se obtendrán 100 zonas donde buscar los pares de puntos homólogos para la evaluación. Para lograr una mejor distribución espacial de estas zonas se ingresará como parámetro a la herramienta de selección de puntos la distancia mínima entre zonas. Esta distancia se fijó en 10 kilómetros.

La selección de los puntos en Topovisión (puntos considerados como verdad terreno) se harán utilizando las herramientas de digitalización con que cuenta este visualizador. Esta permite dibujar elementos geográficos y su posterior descarga en formato KML para luego ser procesados en un software SIG de escritorio. El emparejamiento de puntos se hará a través de un identificador único consistente en un número correlativo del 0 al 99.

El sistema de coordenadas utilizado en este proceso de evaluación será SIRGAS-ROU 98 proyección UTM 21 Sur. En caso de que el sistema de coordenadas del conjunto de datos a evaluar sea distinto del antes mencionado se utilizará QGIS para hacer la transformación de coordenadas.

Una vez identificados los pares de puntos y obtenidas sus coordenadas se procederá a realizar los siguientes cálculos tendientes a la evaluación:

- Medidas básicas de los errores: suma, máximo, mínimo, media, mediana, raíz cuadrada del error cuadrático medio (REMC).

- Medidas ISO 19157: Valor medio de las incertidumbres posicionales (Tabla D.29) y Sesgo de las posiciones (Tabla D.30)
- Para el Control posicional relativo
 - Se calcularán todas las combinaciones de distancias entre los puntos que se seleccionen tanto en el conjunto de datos a evaluar como en el conjunto de datos considerados como verdad terreno.
 - Se obtendrán las diferencias de distancias homólogas entre ambos conjuntos de datos.
 - Detección de outliers sobre las diferencias usando el método de Chauvenet.
 - Verificación de normalidad de las diferencias usando Test de Kolmogorov-Smirnov.
 - Verificación de la aleatoriedad de los errores para cada componente usando el Test de Rachas.
 - Se calcularán las medidas de tendencia central y dispersión.
- Para Control Posicional Absoluto por Puntos.
 - Detección de outliers para ambas componente usando el método de Chauvenet.
 - Verificación de normalidad de los errores para cada componente usando Test de Kolmogorov-Smirnov.
 - Verificación de la aleatoriedad de los errores para cada componente usando el Test de Rachas.
 - Se estudiará la presencia de sistematismos en los datos utilizando un Test sobre la media con el supuesto de varianza poblacional desconocida.
 - Se calculará el coeficiente de Pearson para estudiar la independencia en los errores de las componentes posicionales.
 - Se calculará el *National Standard for Spatial Data Accuracy* (NSSDA).

Tabla 1 - Elementos de la calidad, medidas y métodos de evaluación		
Elemento de la calidad	Medida	Método de evaluación
Exactitud absoluta o externa	ISO 19157 - D.29: Valor medio de las incertidumbres posicionales	Directa externa - Muestreo
Exactitud absoluta o externa	ISO 19157 - D.30: Sesgo de las posiciones.	Directa externa - Muestreo
Exactitud absoluta o externa	National Standard for Spatial Data Accuracy (NSSDA).	Directa externa - Muestreo

Para la exactitud posicional absoluta planimétrica los errores (diferencias de coordenadas) surgen de restar la coordenada del punto en el conjunto de datos geográficos menos la coordenada considerada como verdadera. Los resultados se darán en metros.

En el caso de la exactitud posicional relativa los errores surgen de restar la distancia considerada como verdadera (calculada entre puntos obtenidos de las ortoimágenes) de la homóloga obtenida del conjunto de datos. Los resultados se darán en metros y en porcentaje. Es importante resaltar que esta forma de calcular la exactitud posicional relativa se utiliza para caracterizar los conjuntos de datos estudiados y los resultados se darán en porcentajes. Los resultados que se obtengan no son aplicables directamente a cualquier otra distancia.

En la parte de desarrollo se brindarán los resultados hasta la cifra significativa correspondiente al milímetro a los efectos de permitir la comprobación de los resultados obtenidos. En el apartado de resumen de resultados presentes en este documento, los valores se presentarán con las cifras significativas que corresponde según los datos y métodos usados.

En todos los resultados que se requiera el nivel de confianza seleccionado es de 95%, se presentarán mapas y tablas para mostrar e ilustrar los resultados obtenidos.

ZONAS CENSALES RURALES

Se realizó la evaluación de las zonas censales rurales 2011 de la cuenca del Río Santa Lucía que abarca parcialmente los departamentos de Montevideo, Canelones, San José, Flores, Florida y Lavalleja (635 zonas censales rurales). El dato fue producido por el Instituto Nacional de Estadística para el Censo Nacional 2011 y se puede encontrar en <http://www.ine.gub.uy/web/guest/338>

La zona censal es la menor unidad identificable de las unidades geoestadísticas del INE. Según el documento consultado en el sitio web del INE “En áreas no amanzanadas corresponde a porciones de territorio definidas por límites naturales o artificiales de fácil reconocimiento (cursos de agua, carreteras, caminos vecinales, vías férreas)”.

No se conoce información sobre la calidad teórica de los datos. Los metadatos de este conjunto de datos declara que el denominador de escala es de 50000.

Se seleccionaron 100 puntos en el área de la cuenca del Santa Lucía. Luego se ubicaron los puntos sobre vértices de las zonas censales. Los puntos a seleccionar se estratifican en dos grupos: cruces de caminos y cruce de arroyos (Figura 1). Se seleccionaron de esta forma ya que estos dos tipos de puntos, son los más representativos de los límites de las zonas censales además de tener características distintas por ser: naturales y artificiales.

En cuanto a las estadísticas generales los puntos se analizarán de forma conjunta y luego por separados, a los efectos de ver posibles diferencias en la exactitud posicional en relación al tipo de punto. Para los demás cálculos se considerarán de manera indistinta los tipos de puntos.

Resultados para el total de los puntos

En esta instancia no se eliminaron outliers ya que se revisó y se verificó que los puntos homólogos seleccionados estén bien identificados. La eliminación de outliers se entiende que pueden desvirtuar el resultado de la evaluación ya que son puntos con errores que son posibles de ser seleccionados en una muestra (Tabla 2). Los módulos y las direcciones de los errores se muestran en las figuras 2 y 3.

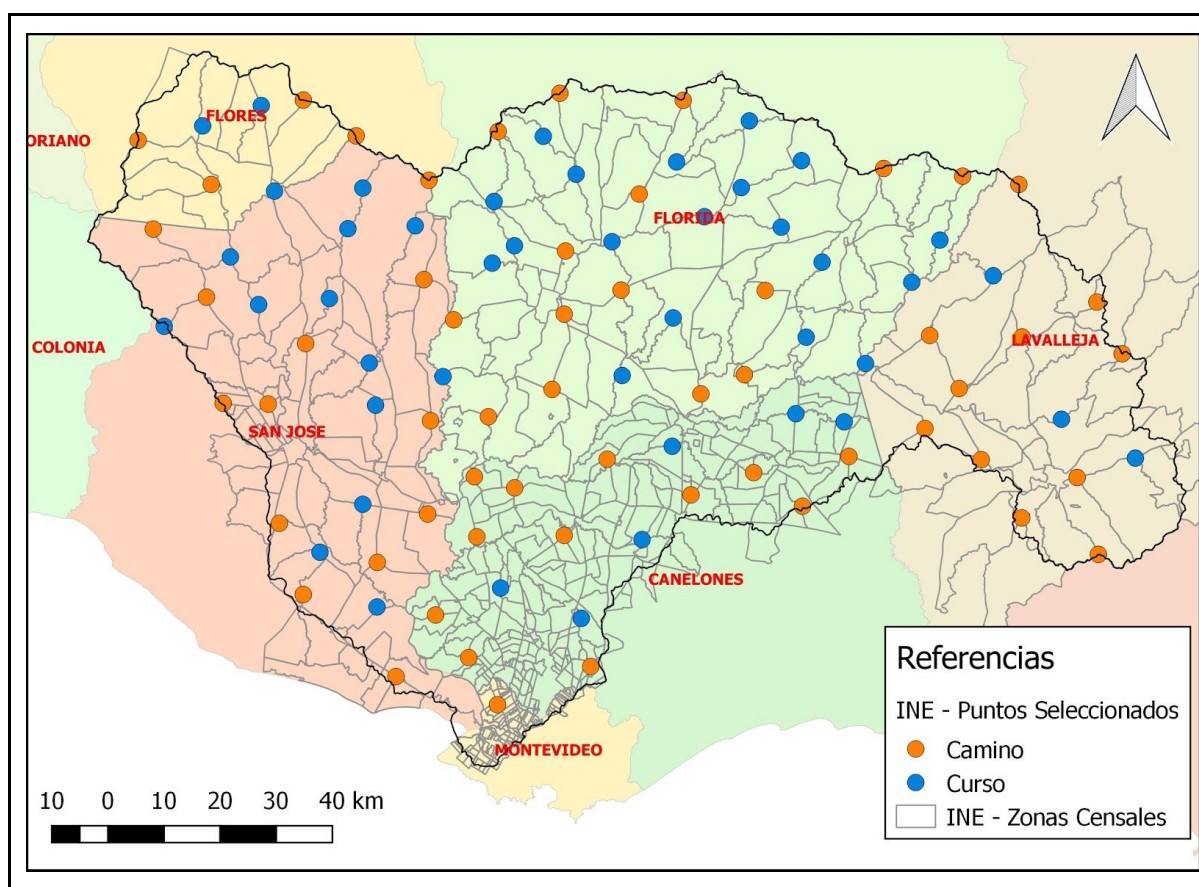


Figura 1 - Distribución de la muestra de puntos por tipo de punto para zonas censales.

Tabla 2 - Información básica para el total de puntos		
Medida	Diferencia X (metros)	Diferencia Y (metros)
Suma	423,475	362,241
Máximo	309,106	242,746
Mínimo	-168,322	-131,219
Media	4,235	3,622
Mediana	1,110	3,297
Error Medio Cuadrático	54,565	43,723

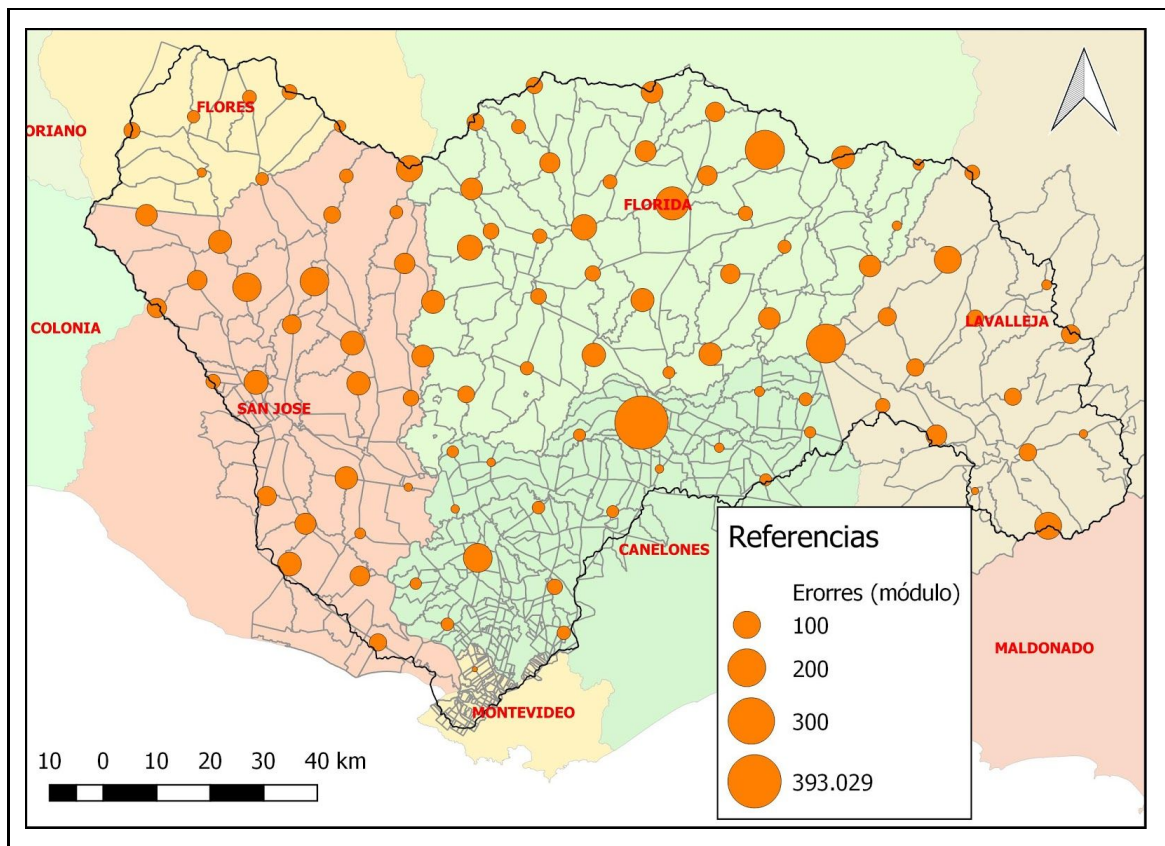


Figura 2 - Módulo de los errores para las zonas censales

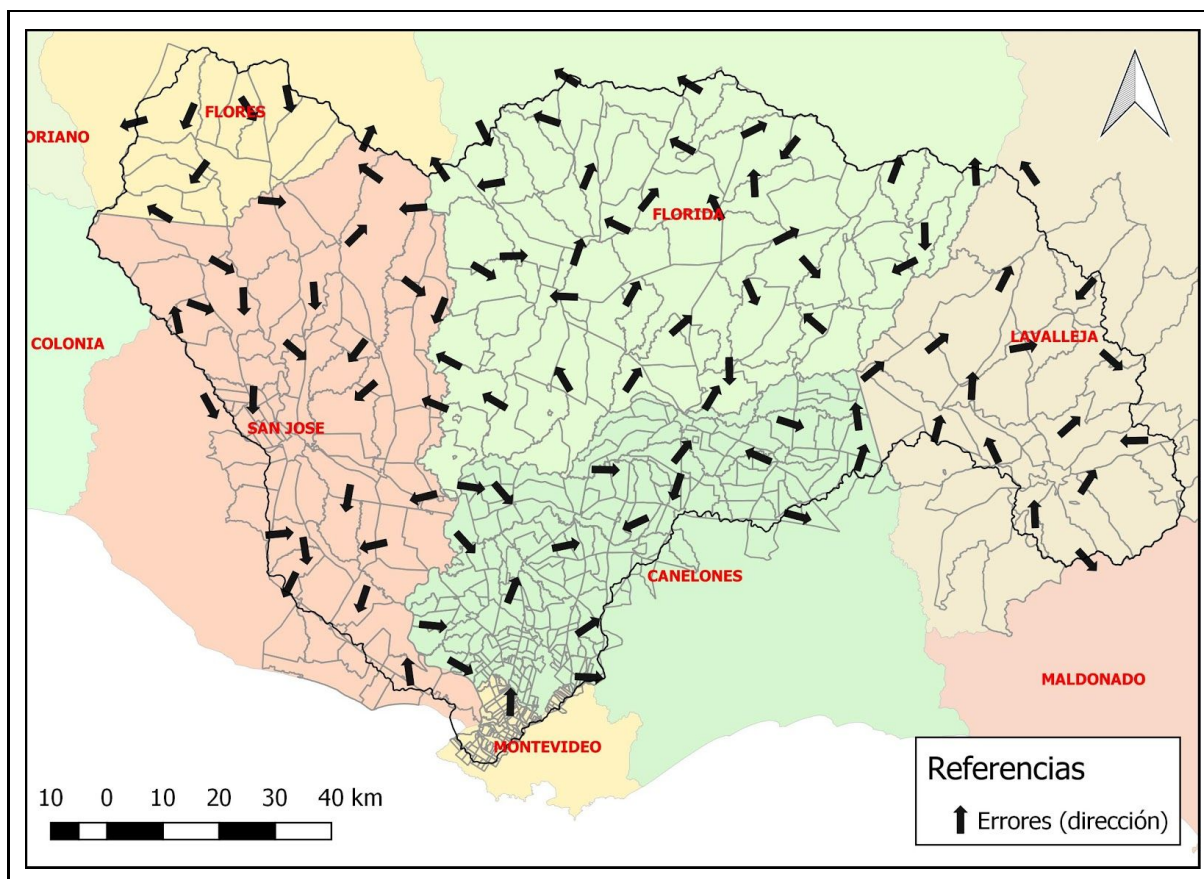


Figura 3 - Dirección de los errores para las zonas censales

Como se mencionó anteriormente se tomaron puntos homólogos de dos tipos: intersecciones de caminos y de cursos de agua. Para caracterizar estos puntos de forma diferenciada se muestran las estadísticas básicas por tipo de punto en la tabla 3.

Tabla 3 - Información básica para los puntos estratificados				
Caminos (56 puntos)			Cursos de agua (44 puntos)	
Medida	Diferencia X	Diferencia Y	Diferencia X	Diferencia Y
Suma	-17,858	148,680	441,333	213,561
Máximo	75,957	65,999	309,106	242,746
Mínimo	-77,098	-52,410	-168,322	-131,219
Media	-0,319	2,655	10,030	4,854
Mediana	1,110	3,297	3,817	2,303
Error Medio Cuadrático	30,934	24,380	74,490	59,903

Los resultados utilizando las medidas de la calidad establecidas en la norma ISO 19157 se muestran en las tablas 4 y 5.

Tabla 4 - Valor medio de las incertidumbres posicionales (Tabla D.29 Norma ISO 19157)		
1D X (metros)	1D Y (metros)	2D (metros)
33,611	27,636	46,968

Tabla 5 - Sesgo de las posiciones (Tabla D.30 Norma ISO 19157)		
ax (metros)	ay (metros)	ap (metros)
4,235	3,622	5,573

Control posicional relativo

Se calcularon las 4950 combinaciones de distancias entre los puntos seleccionados y se realizó la diferencia entre distancias homólogas. La diferencia se expresa como porcentaje de la distancia considerada como verdadera.

A los efectos de caracterizar los valores de distancias, el valor máximo del universo de distancias trabajadas es de 186456,426 metros y el mínimo es de 5014,467 metros. Sin quitar los outliers las estadísticas básicas se muestran en la tabla 6.

Tabla 6 - Información básica para los errores considerados en la exactitud relativa			
	Diferencia (m)	Diferencia (%)	Diferencia absoluta (%)
Suma	-36909,667	-44,104	509,343
Máximo	538,150	3,189	3,189
Mínimo	-451,205	-3,028	0,000
Media	-7,456	-0,009	0,103
Mediana	-7,118	-0,011	0,056
Error Medio Cuadrático	73,857	0,197	0,197

Para detectar outliers en las observaciones se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018).

P = 0,9999 → Error = $4,0 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 7 - Detección de outliers - Diferencia de distancias						
Iteración	Error	Tamaño muestra	Error máximo (+/- %)	Límite máximo (%)	Límite mínimo (%)	Cantidad de distancias eliminadas
0	$4,0 \times \sigma$	4950	0,786	0,777	-0,786	47
1	$4,0 \times \sigma$	4903	0,582	0,571	-0,594	53
2	$4,0 \times \sigma$	4850	0,513	0,501	-0,524	35
3	$4,0 \times \sigma$	4815	0,480	0,468	-0,492	16
4	$4,0 \times \sigma$	4799	0,467	0,455	-0,479	3
5	$4,0 \times \sigma$	4796	0,465	0,452	-0,477	3
6	$4,0 \times \sigma$	4793	0,462	0,450	-0,475	0

Luego de quitar los outliers la cantidad de observaciones a utilizar en el proceso de cálculo es de 4793 (tabla 7) diferencias de distancias, lo que implica que se eliminaron el 3,17% de las observaciones.

Para determinar la normalidad de los errores en las distancias se utilizará el Test de Kolmogorov-Smirnov para un nivel de confianza del 95% (Tabla 8).

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 8 - Parámetros del Test de Kolmogorov-Smirnov	
	Error X
N	4793
C_α	0,895
k(n)	70,358
D_n	0,087
D_α	0,013
Cumplimiento	No cumple

Para la aleatoriedad se utilizará el Test de Rachas el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96 (Tabla 9).

Tabla 9 - Parámetros del Test de Rachas	
	Error X
N	4793
n_0	2745
n_1	2048
R	2196
μ_R	2346,821
σ_R	33,880
Z_R	-4,452
Cumplimiento	No cumple

Luego de quitar los outliers si calculamos el error medio cuadrático 0,116%. Las diferencias entre las distancias no sigue una distribución normal por lo que en vez de calcular un intervalo de confianza se calcularon los percentiles 2,5% y el 97,5%. Los valores para estos parámetros son -0,268% y 0,252%.

Control posicional absoluto por puntos

Detección de outliers

Para aplicar el NSSDA primero se detectarán outliers en la componente X y luego en la componente Y. Para esto se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018). Los puntos eliminados se muestran en las tablas 10 y 11.

P = 0,995 → Error = $2,8 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 10 - Detección de outliers - Variable X						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Puntos eliminados
0	$2,8 \times \sigma$	100	153,088	157,323	-148,853	2 y 32
1	$2,8 \times \sigma$	98	118,261	121,146	-115,377	24 y 86
2	$2,8 \times \sigma$	96	106,320	106,436	-106,204	19 y 37
3	$2,8 \times \sigma$	94	97,069	99,548	-94,589	84
5	$2,8 \times \sigma$	93	92,810	94,184	-91,437	Ninguno

Tabla 11 - Detección de outliers - Variable Y						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Puntos eliminados
0	$2,8 \times \sigma$	93	83,400	84,439	-82,361	Ninguno

Es importante destacar que todos los puntos detectados como outliers corresponden a puntos obtenidos sobre cursos de agua (Figura 4).

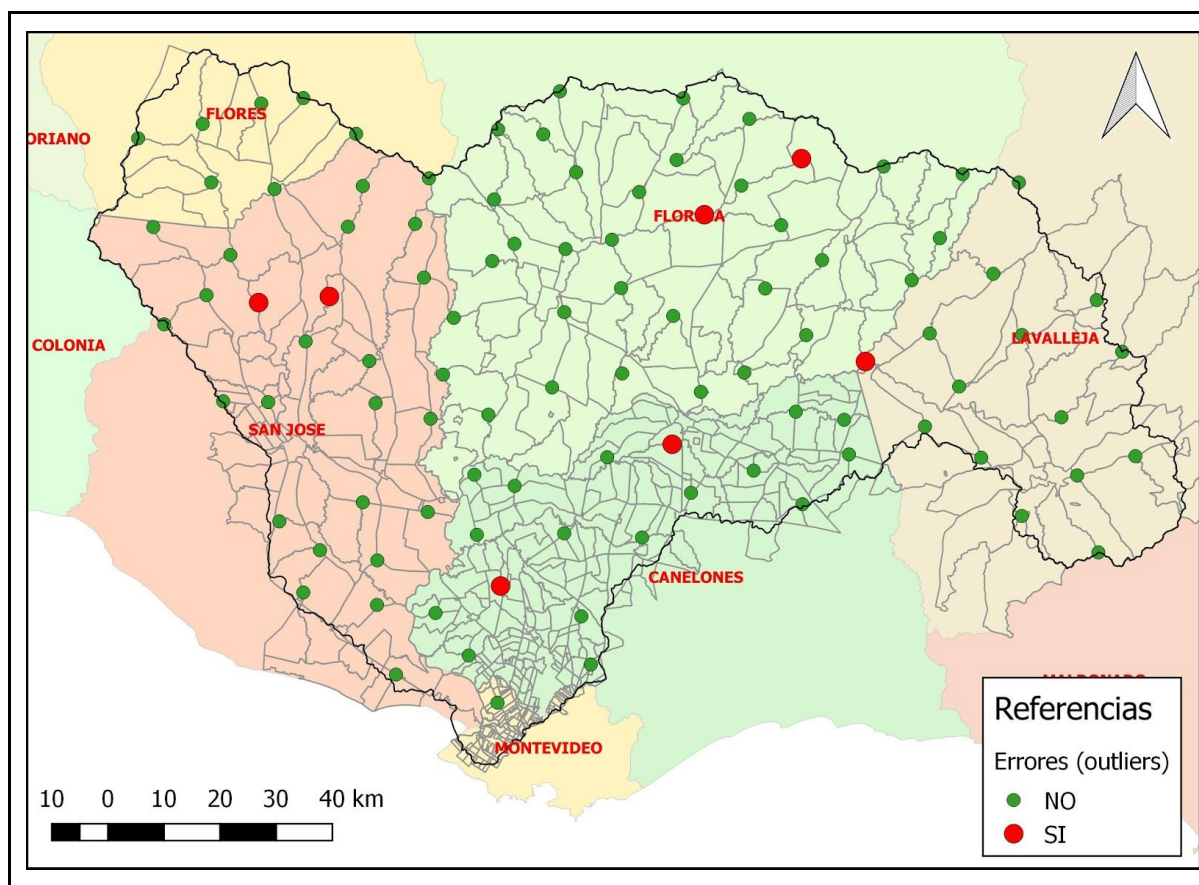


Figura 4 - Ubicación de los outliers detectados para las zonas censales

Control de normalidad

Para determinar la normalidad de los errores en las componentes estudiadas (X, Y) se utilizará el Test de Kolmogorov-Smirnov para un nivel de confianza del 95% (Tabla 12).

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 12 - Parámetros del Test de Kolmogorov-Smirnov		
	Error X	Error Y
N	93	93
C_{α}	0,895	0,895
k(n)	9,722	9,722
D_n	0,111	0,081
D_{α}	0,092	0,092
Cumplimiento	No cumple	Cumple

Control de aleatoriedad

Para la aleatoriedad se utilizará el Test de Rachas el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96 (Tabla 13).

Tabla 13 - Parámetros del Test de Rachas		
	Error X	Error Y
N	93	93
n_0	45	42
n_1	48	51
R	48	43
μ_R	47,452	47,065
σ_R	4,791	4,750
Z_R	0,114	0,856
Cumplimiento	Cumple	Cumple

Control de independencia

Para la independencia se calculará el coeficiente de correlación de Pearson. El resultado para este coeficiente es de -0.0597. La distribución de los errores se muestra en la figura 5 donde se puede ver la no correlación de los errores en la componente X e Y.

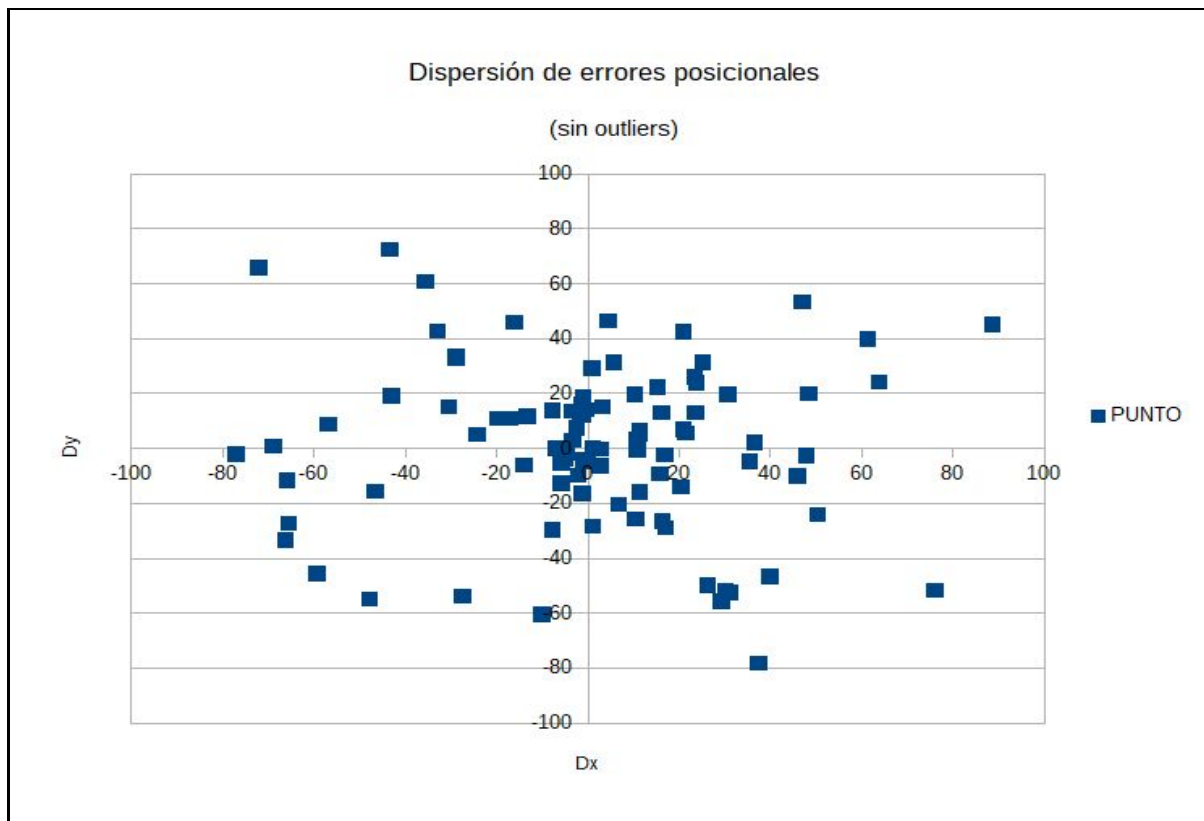


Figura 5 - Distribución de los errores posicionales para las zonas censales

Control de sistematismo

Para comprobar si el error sistemático es aceptable se realiza un test sobre la media con el supuesto de varianza poblacional desconocida para un 95% de nivel de confianza. Para esto las hipótesis son:

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$
$ t $ menor o igual a $t_{n-1,\alpha}$

Tabla 14 - Parámetros test de control de sistematismos		
	Error X	Error Y
N	93	93
Media	-1,374	-1,039
D. E. Muestral	33,147	29,786
Estadístico t	0,400	0,336
Estadístico $t_{0,05;92}$	1,986	
Cumplimiento	Cumple	Cumple

Se verifica que los errores sistemáticos son aceptables por lo que se puede considerar que los errores no presentan sistematismos o estos son muy pequeños (Tabla 14).

En resumen (tabla 15) se tiene la siguiente información para los errores calculados a partir de la muestra.

Tabla 15 - Resumen de resultados para el control de las zonas censales		
Para el 95% de confianza	Error X	Error Y
Detección de outliers	Si. Método de Chauvenet	
Normalidad	No cumple	Cumple
Aleatoriedad	Cumple	Cumple
Sistematismo	Cumple	Cumple
Independencia	El coeficiente de Pearson es de -0,0597	

National Standard for Spatial Data Accuracy (NSSDA)

El resultado de aplicar el NSSDA es:

Tabla 16 - Resultados para el NSSDA para zonas censales		
EMC	EMCmin / EMCmax	NSSDA
EMCx = 32,997 metros	0,898	76,662 metros
EMCy = 29,643 metros		

Resultados

Se puede ver en la Tabla 2 que las medias y medianas de los errores son similares entre sí y cercanos a 0. Si analizamos estas variables por tipo de puntos (tabla 3) se puede ver que estos parámetros se diferencian más cuando los puntos son sobre cursos de agua que sobre caminos. Esto es esperable ya que al estar tomados sobre elementos naturales la variación de los mismos y la indefinición en los criterios de digitalización hacen que los errores cometidos sean mayores.

Si se analiza el rango de valores en las diferencias de coordenadas los mismo son amplios comparados con los errores esperados o declarados. De forma general para la componente X el rango de valores es de 477,428 metros y en la componente Y de 373,965 metros. Si este rango se analiza por componente (tabla 3) resulta que los rangos en los valores de las diferencias de coordenadas error son del orden de 3 veces más grandes en los cursos de agua que en los caminos (tabla 17).

Tabla 17 - Valores máximo, mínimos y rangos para las diferencias de coordenadas por tipo de punto.				
	Camino		Curso de agua	
	DX	DY	DX	DY
Máximo	75,957	65,999	309,106	242,746
Mínimo	-77,098	-52,410	-168,322	-131,219
Rango	153,055	118.409	477,428	373,965

También de la tabla 3 se puede apreciar que los REMC para los caminos están en el orden de la mitad de los REMC para los cursos de agua. Esto refuerza la diferencia entre los elementos naturales y artificiales.

Como se mencionó anteriormente los 7 puntos detectados como outliers se tratan de puntos ubicados sobre cursos de agua lo que es consistente con los resultados antes mencionados al mostrar los errores más grandes

El coeficiente de correlación de Pearson tiene un valor muy cerca de 0 por lo que se puede concluir que no hay correlación lineal entre las variables. Esto se puede verificar por la distribución de los pares de errores DX, DY que se muestran en la figura 5.

Como se mencionó anteriormente se establece en los metadatos que la escala de los datos es de 1:50000. Considerando el criterio del cuarto de milímetro como error esperado en un conjunto de datos esto implicaría que de forma estándar el error esperado es de 12,5 metros (para el 67.5% de los puntos). Si este valor se extendiera para el 95% el error sería de 24,5 metros.

Considerando estos valores y el resultado del NSSDA se concluye que el error encontrado es del orden de 3 veces mayor al declarado.

CAMINERÍA

Se trabaja con los caminos de la base de datos del Ministerio de Transporte y Obras Públicas que se encuentran en la cuenca del Río Santa Lucía. En total se trabaja con 13528 elementos geográficos. Los datos se obtuvieron de los geoservicios del Geoportal del MTOP (geoportal.mtop.gub.uy) y los estos son producto de la información recabada por las intendencias departamentales, OPP y MTOP.

No se conocen evaluaciones de la calidad posicional para este conjunto de datos. Según los metadatos de este conjunto de datos disponible en el Geoportal del MTOP (geoportal.mtop.gub.uy) la escala declarada para este conjunto de datos es de 1/10000, lo que implica siguiendo el criterio del ¼ de milímetro que el error en metros es de 2.50 metros y de 4.90 metros para el 95% de nivel de confianza.

También de los metadatos se obtiene la metodología de captura del dato *“El relevamiento se efectuó recorriendo todo camino público transitable con vehículos propios, dotados de receptor GPS (Trimble Geoexplorer 3 y Leica GS 20 -en el caso de lo relevado por el SGM-) y antena sobre el mismo vehículo. Se tomaron posiciones cada 30 metros y 15 metros respectivamente; en aquellos casos donde se realizaron actualizaciones (octubre 2015 en adelante) se procedió a realizar una simplificación de los trazados para reducir el número de nodos de los arcos. Los datos recibidos y corregidos diferencialmente (con Pathfinder Office y GIS Data Pro 3.0 respectivamente) permiten obtener precisiones que rondan los 2 metros en planimetría. En el caso de rotondas e intercambiadores se apoyó el relevamiento con datos de imágenes (proveniente de las aplicaciones Google Earth, Bing Maps, HERE y Yandex) y digitalización en gabinete. En general en gabinete, además, se realizaron diferentes correcciones y controles, utilizándose para ello diversos programas y fuentes (imágenes satelitales, fotografías aéreas, información complementaria -otros relevamientos, cruzamiento de información, etc.)”*

Se puede ver que este conjunto de datos en su mayor parte fue generado a través de un relevamiento directo de los caminos aunque las actualizaciones son producto de la digitalización sobre imágenes no controladas. Esto puede servir de indicio que la calidad declarada puede estar degradada.

Resultado para el total de los puntos

En esta instancia no se eliminaron outliers ya que se revisó y se verificó que los puntos homólogos seleccionados estén bien identificados. La eliminación de outliers se entiende que pueden desvirtuar el resultados de la evaluación ya que son puntos con errores que son pasibles de ser seleccionados en una muestra. Las diferencias de coordenadas surge de restar la coordenada del punto en el conjunto de datos geográficos menos la coordenada considerada como verdadera. La distribución de los puntos seleccionados en la cuenca se muestra en la figura 6 y los principales resultados en la tabla 18.

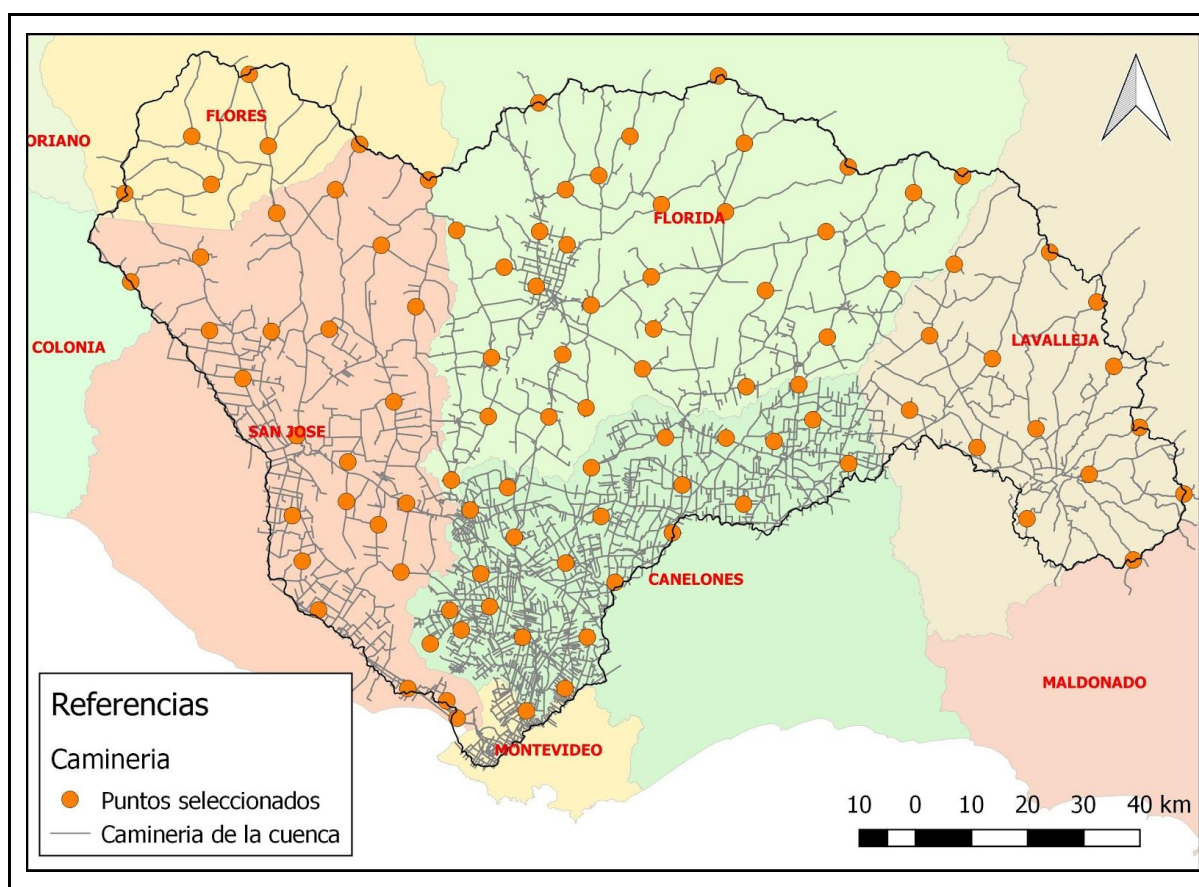


Figura 6 - Distribución de los puntos de control seleccionados para caminería

Tabla 18 - Información básica para el total de puntos		
Medida	Diferencia X (metros)	Diferencia Y (metros)
Suma	-334,544	742,300
Máximo	10,426	35,123
Mínimo	-21,182	-34,003
Media	-3,345	7,423
Mediana	-3,345	9,323
Error medio cuadrático	6,084	10,998

En las figuras 7 y 8 se pueden ver graficados los módulos y las direcciones de los errores.

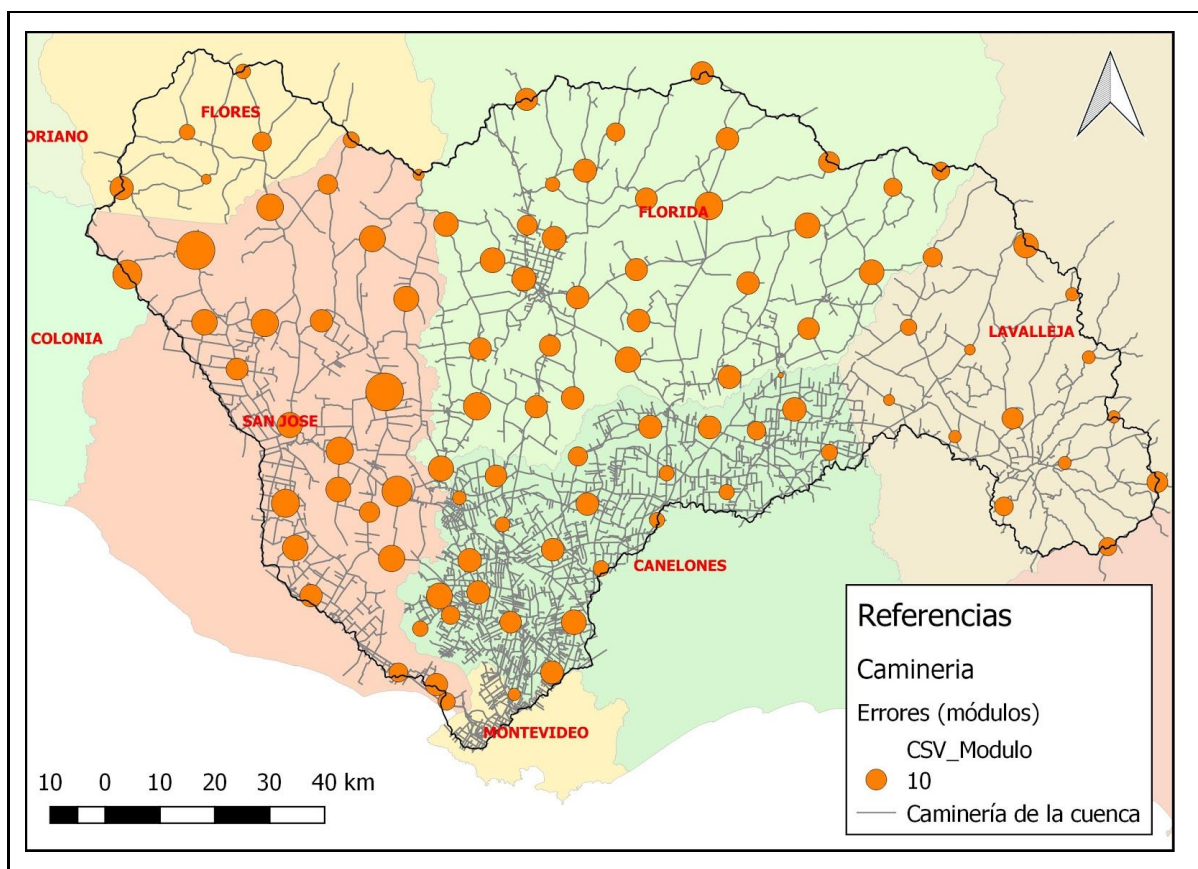


Figura 7 - Módulos de los errores en los puntos de control seleccionados para caminería

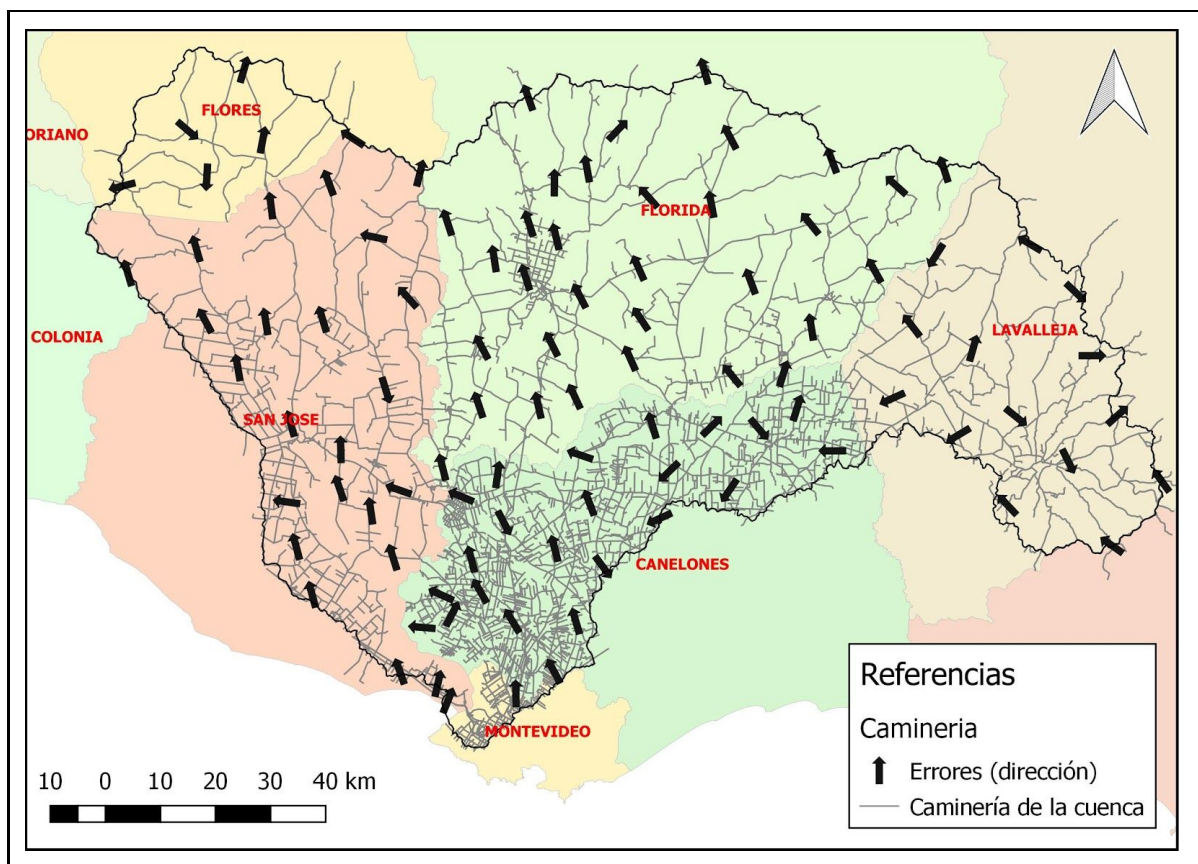


Figura 8 - Dirección de los errores en los puntos de control seleccionados para caminería

En las tablas 19 y 20 se muestran los resultados utilizando las medidas de la calidad establecidas en la norma ISO 19157.

Tabla 19 - Valor medio de las incertidumbres posicionales (Tabla D.29 Norma ISO 19157)		
1D X (metros)	1D Y (metros)	2D (metros)
4,818	9,153	11,013

Tabla 20 - Sesgo de las posiciones (Tabla D.30 Norma ISO 19157)		
ax (metros)	ay (metros)	ap (metros)
-3,345	7,423	8,142

Control posicional relativo

Se calcularon las 4950 combinaciones de distancias entre los puntos seleccionados y se realizó la diferencia entre distancias homólogas. La diferencia se expresa como porcentaje de la distancia considerada como verdadera.

A los efectos de caracterizar los valores de distancias, el valor máximo del universo de distancias trabajadas es de 196129,118 metros y el mínimo es de 3591,850 metros. Los resultados pueden verse en la tabla 21.

Tabla 21 - Información básica para los errores considerados en la exactitud relativa			
	Diferencia (m)	Diferencia (%)	Diferencia absoluta (%)
Suma	-7251,173	-12,983	71,349
Máximo	50,801	0,275	0,275
Mínimo	-57,965	-0,250	0,000
Media	-1,465	-0,003	0,014
Mediana	-0,755	-0,001	0,008
Error cuadrático medio	83,258	0,026	0,026

Para detectar outliers en las observaciones se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018).

P = 0,9999 → Error = $4,0 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 22 - Detección de outliers - Diferencia de distancias						
Iteración	Error	Tamaño muestra	Error máximo (+/- %)	Límite máximo (%)	Límite mínimo (%)	Cantidad de distancias eliminadas
0	$4,0 \times \sigma$	4950	0,105	0,102	-0,107	57
1	$4,0 \times \sigma$	4893	0,078	0,076	-0,080	56
2	$4,0 \times \sigma$	4837	0,069	0,066	-0,071	24
3	$4,0 \times \sigma$	4813	0,066	0,063	-0,068	5
4	$4,0 \times \sigma$	4808	0,065	0,063	-0,068	2
5	$4,0 \times \sigma$	4806	0,065	0,062	-0,068	0

Luego de quitar los outliers la cantidad de observaciones es de 4806 pares de distancias. Se eliminaron el 2,91% de las observaciones (tabla 22).

Para determinar la normalidad de los errores en las distancias se utilizará el Test de Kolmogorov-Smirnov para un nivel de confianza del 95% (tabla 23).

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 23 - Parámetros del Test de Kolmogorov-Smirnov	
	Error X
N	4806
C_α	0,895
k(n)	69,328
D_n	0,004
D_α	0,013
Cumplimiento	Cumple

Para la aleatoriedad se utilizará el Test de Rachas el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96 (tabla 24).

Tabla 24 - Parámetros del Test de Rachas	
	Error X
N	4806
n_0	2626
n_1	2180
R	2072
μ_R	2383,305
σ_R	34,361
Z_R	-9,060
Cumplimiento	No cumple

Luego de quitar los outliers el error cuadrático medio es 0,016%.

Las diferencias entre las distancias sigue una distribución normal con una media de -0,003% y una desviación estándar de 0,016%. El intervalo de confianza para las diferencias de distancias para el 95% de los casos es (-0,035% / 0,029%).

A los efectos de tener una línea de comparación con los otros conjuntos de datos se calculan los percentiles 2,5% y el 97,5%. Los valores para estos parámetros son -0,040% y 0,030%.

Control posicional absoluto por puntos

Detección de outliers

Para aplicar el NSSDA primero se detectarán outliers en la componente X y luego en la componente Y. Para esto se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018). Los puntos eliminados se muestran en las tablas 25 y 26 y en la figura 9.

P = 0,995 → Error = $2,8 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 25 - Detección de outliers - Variable X						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Punto eliminado (metros)
0	$2,8 \times \sigma$	100	14,300	10,955	-17,646	46 y 83
1	$2,8 \times \sigma$	98	12,725	9,723	-15,728	20 y 40
2	$2,8 \times \sigma$	96	11,673	8,667	-14,678	31 y 84
3	$2,8 \times \sigma$	94	10,741	7,735	-13,747	44
5	$2,8 \times \sigma$	93	10,290	7,164	-13,416	Ninguno

Tabla 26 - Detección de outliers - Variable Y						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Punto eliminado (metros)
0	$2,8 \times \sigma$	93	19,777	27,883	-11,671	22
1	$2,8 \times \sigma$	92	18,217	26,029	-10,404	Ninguno

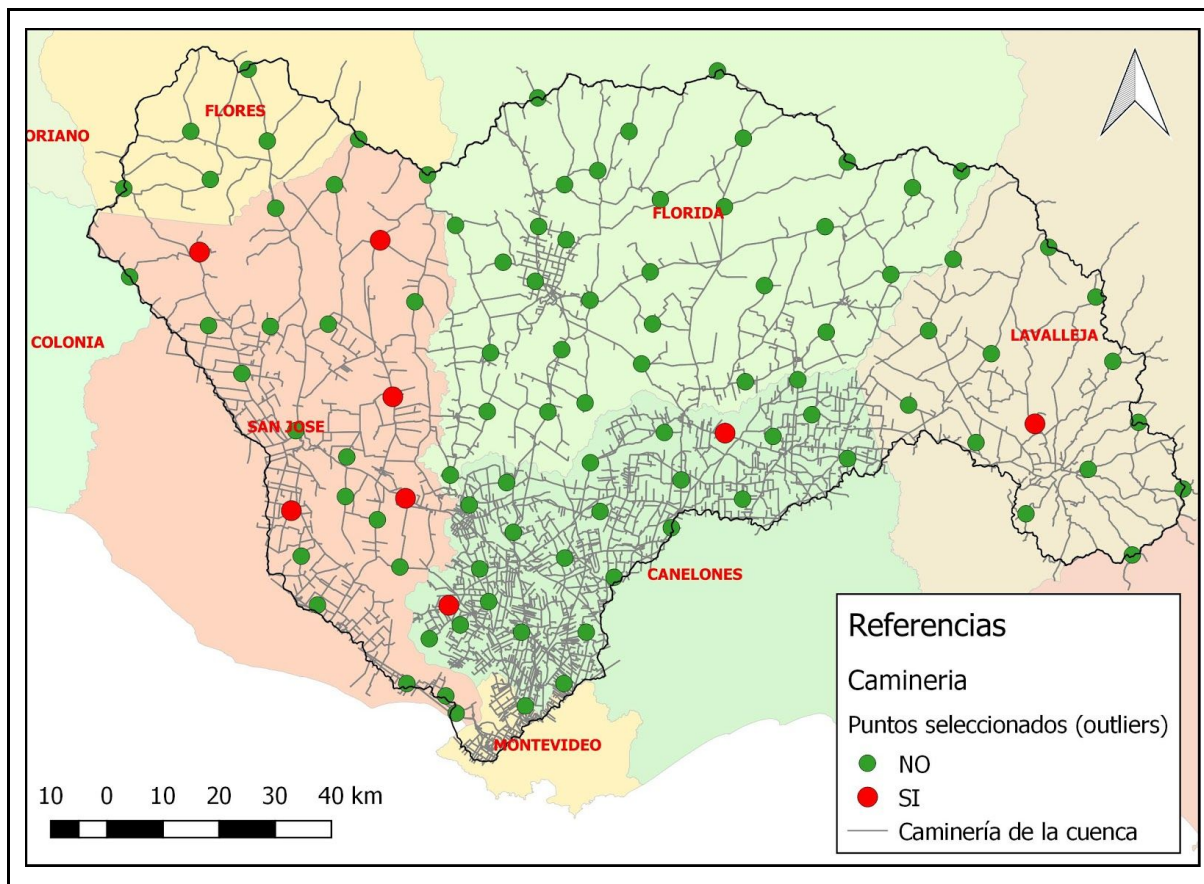


Figura 9 - Ubicación de outliers para los puntos de control de caminería

Control de normalidad

Para determinar la normalidad de los errores en las componentes estudiadas (X, Y) se utilizará el Test de Kolmogorov-Smirnov para un nivel de confianza del 95% (tabla 27).

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 27 - Parámetros del Test de Kolmogorov-Smirnov		
	Error X	Error Y
N	92	92
C_α	0,895	0,895
k(n)	9,670	9,670
D_n	0,109	0,126
D_α	0,093	0,093
Cumplimiento	No cumple	No cumple

Control de aleatoriedad

Para la aleatoriedad se utilizará el Test de Rachas el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96 (tabla 28).

Tabla 28 - Parámetros del Test de Rachas		
	Error X	Error Y
N	92	92
n₀	72	15
n₁	20	77
R	35	30
μ_R	32,304	26,109
σ_R	3,229	2,579
Z_R	0,835	1,509
Cumplimiento	Cumple	Cumple

Control de independencia

Para la independencia se calculará el coeficiente de correlación de Pearson. El resultado para este coeficiente es de -0.3182. El coeficiente tiene un valor negativo por lo que se presenta una correlación negativa baja entre los errores de las componentes (Ver figura 10). Si bien la correlación lineal es baja se puede ver en este gráfico la relación que existe entre las componentes (error en X negativo; error en Y positivo).

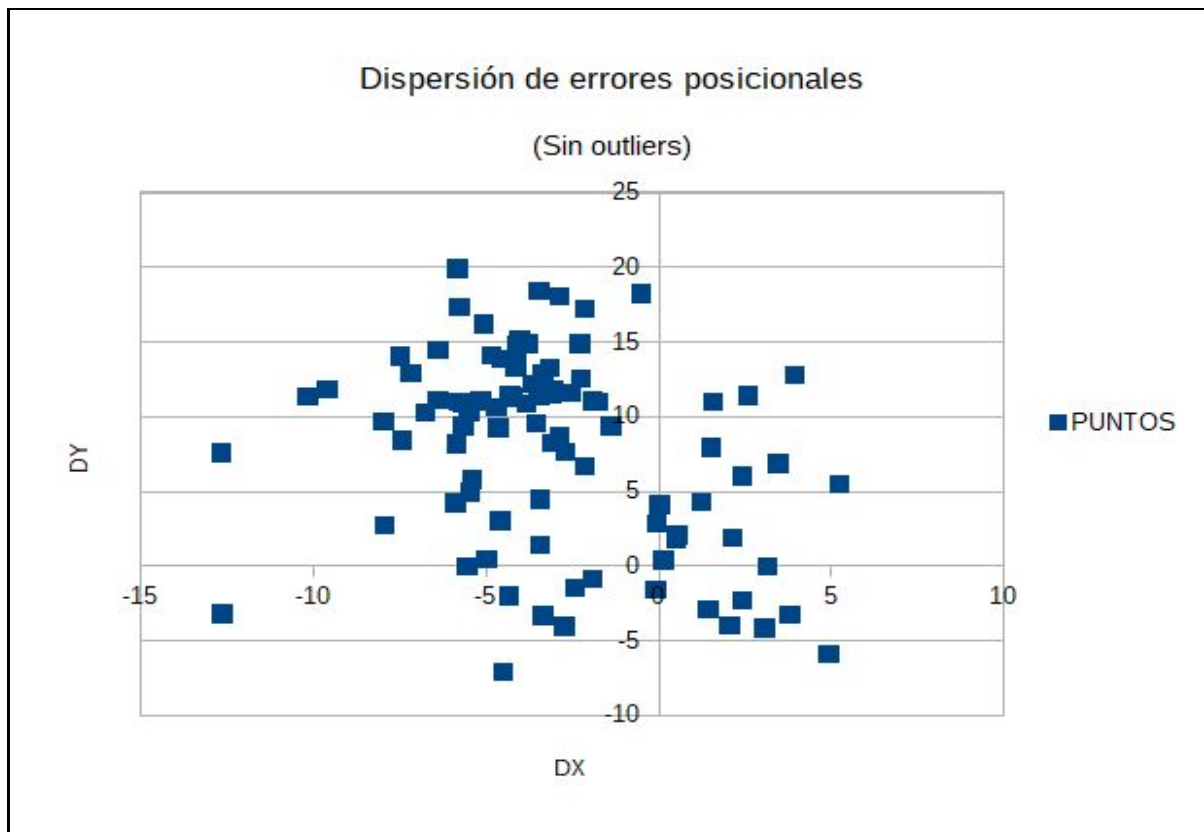


Figura 10 - Distribución de los errores posicionales para la caminería

Control de sistematismo

Para comprobar si el error sistemático es aceptable se realiza un test sobre la media con el supuesto de varianza poblacional desconocida para un 95% de nivel de confianza. Para esto las hipótesis son los siguientes y los resultados se muestran en la tabla 29:

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$
$ t $ menor o igual a $t_{n-1,\alpha}$

Tabla 29 - Parámetros test de control de sistematismos		
	Error X	Error Y
N	92	92
Media	-3,049	7,812
D. E. Muestral	3,618	6,506
Estadístico t	8,082	11,518
Estadístico $t_{0,05;92}$	1,986	
Cumplimiento	No cumple	No cumple

En resumen se tiene la siguiente información para los errores calculados a partir de la muestra (Ver tabla 30).

Tabla 30 - Resumen de resultados para el control de la caminería		
Para el 95% de confianza	Error X	Error Y
Detección de outliers	Si. Método de Chauvenet	
Normalidad	No cumple	No cumple
Aleatoriedad	Cumple	Cumple
Sistematismo	No cumple	No cumple
Independencia	El coeficiente de Pearson es de -0.3182	

National Standard for Spatial Data Accuracy (NSSDA)

Aunque no se cumplan algunas de las condiciones que se requieren (ver tabla 30), el resultado de aplicar el NSSDA es:

Tabla 31 - Resultados para el NSSDA para zonas censales		
EMC	EMCmin / EMCmax	NSSDA
EMCx = 4,717 metros	0,465	18,187 metros
EMCy = 10,144 metros		

Como se puede ver en la tabla 31 tampoco se cumple la condición de que la relación entre en EMC_{\min}/EMC_{\max} se encuentre entre 1 y 0,6. Esta es otra de las condiciones que no se cumple.

Resultados

Lo primero que se se destaca es el claro sistematismos de los errores como se puede ver en la figura 8 y 10. En la figura 8 se ve que la mayoría de los vectores están hacia el 4º cuadrante (dirección nor-oeste). Esto es especialmente claro para los puntos ubicados en los departamentos de San José, Florida y Canelones. En la lámina 10 se puede ver que la ubicación de los pares de errores también se ubican en el cuarto cuadrante de forma mayoritaria. Esto se confirma ya que para el test de detección de sistematismos se rechaza la hipótesis nula que indica la inexistencia de estos errores (tabla 29).

Este sistematismo se puede deber al procesamiento del relevamiento inicial para la generación de los datos ya que parece ser un error general del conjunto de datos.

Independientemente de lo anterior y que también se rechazó la normalidad de los datos, se aplicó el NSSDA. El resultado para este cálculo es de 18,187 metros. Si comparamos este resultado con la calidad declarada vemos que se está bastante apartada de la misma. La relación entre el valor del NSSDA calculado y la calidad declarada en los metadatos es de 3.7 veces mayor.

CATASTRO RURAL

Se trabajará con las parcelas rurales de la cuenca del Río Santa Lucía siendo 51372 las consideradas para este estudio. Los datos fueron descargados del Catálogo de Datos Abierto del Gobierno Uruguayo y la fecha de su actualización es el 16 de noviembre de 2018. La organización responsable de la generación y mantenimiento de este conjunto de datos es la Dirección Nacional de Catastro. Los puntos seleccionados para el estudio son clasificados en dos tipos, vértices de parcelario sobre camino y vértices de parcelario fuera de los caminos.

Para este conjunto de datos se cuenta con un estudio anterior de la exactitud posicional absoluta para el parcelario del departamento de Lavalleja (Faure et al, 2010). Parte de la superficie de trabajo coincide con el trabajo anterior por ocupar, este último, un área que forma parte de la cuenca del Río Santa Lucía.

En sus conclusiones este trabajo establece que *“Los resultados obtenidos son exclusivamente aplicables al parcelario digital del departamento de Lavalleja, no debiéndose extrapolar los mismos a otros parcelarios departamentales, esto es debido a la forma en que fueron confeccionadas en su momento las láminas 1/20.000 de la Dirección Nacional de Catastro.”* Independientemente de esto se considera un antecedente a tener en cuenta el resultado obtenido en este trabajo ya que se siguieron procedimientos estadísticos probados similares a los utilizados en este estudio.

En Faure et al 2010 realizó el control posicional a través del estándar NSSDA. Según este informe los errores posicionales no siguen una distribución normal y aleatoria pero de todas formas se continúa con los cálculos (según lo habilita el NSSDA) lo que da como resultado según el estándar de una exactitud de 211 metros para el 95% de nivel de confianza.

Si bien las conclusiones de este estudio son claras en cuanto a los resultados en los metadatos del conjunto de datos (<http://visor.catastro.gub.uy:8080/geonetwork/srv/spa/main.home>) se expresa que el denominador de escala del mismo es 20000. Seguramente este valor surge considerando el origen de los parcelarios rurales en las láminas catastrales producidas a partir del vuelo fotogramétrico del año 1966 a escala 1:20000 y no el resultado del antecedente considerado.

Resultados para el total de los puntos

En esta instancia no se eliminaron outliers ya que se revisó y se verificó que los puntos homólogos seleccionados estén bien identificados. La eliminación de outliers se entiende que pueden desvirtuar el resultados de la evaluación ya que son puntos con errores que son posibles de ser seleccionados en una muestra. Las diferencias de coordenadas surge de restar la coordenada del punto en el conjunto de datos geográficos menos la coordenada considerada como verdadera.

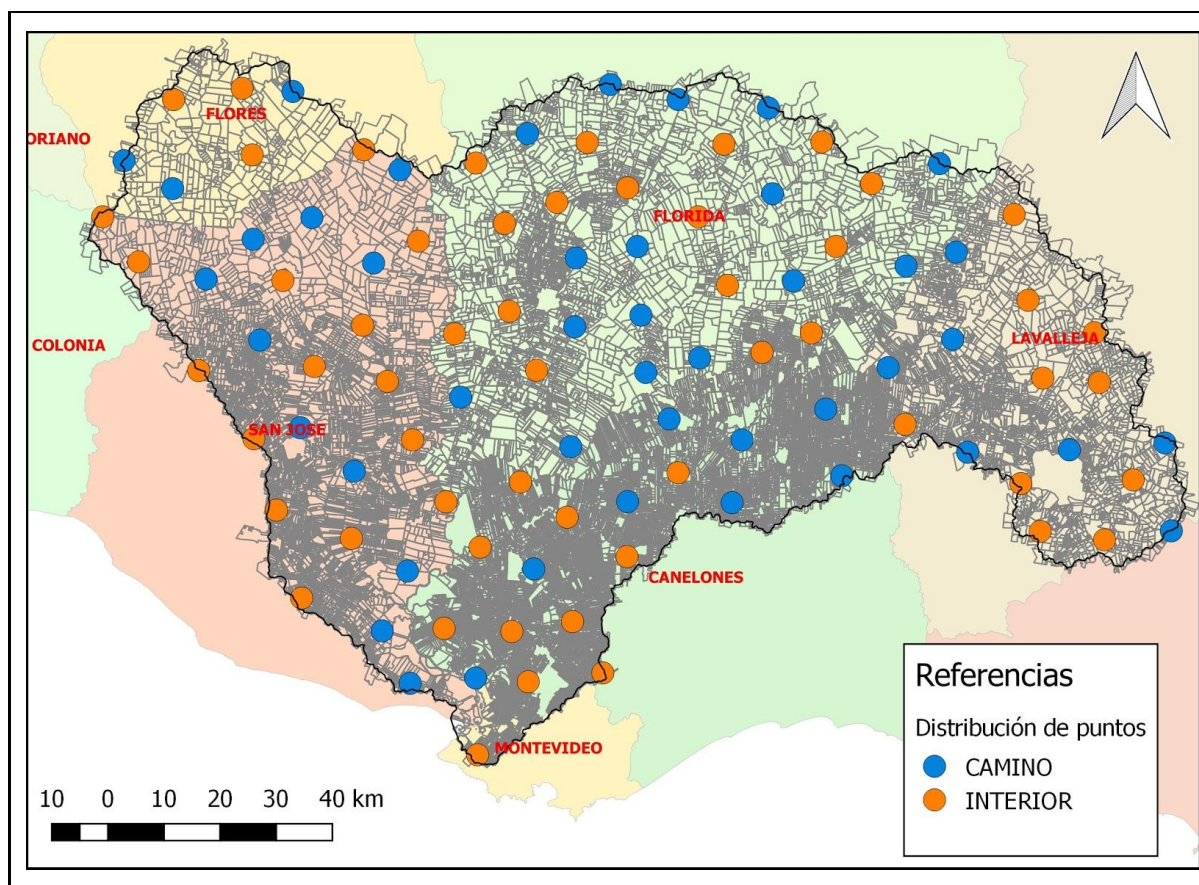


Figura 11 - Distribución de los puntos de control seleccionados para el catastro rural

Tabla 32 - Información básica para el total de puntos		
Medida	Diferencia X (metros)	Diferencia Y (metros)
Suma	897,788	-2193,913
Máximo	232,281	343,005
Mínimo	-196,058	-225,373
Media	8,978	-21,939
Mediana	9,927	-22,516
Error cuadrático medio	71,097	88,063

Como se mencionó anteriormente se tomaron puntos que se clasifican de dos maneras: vértices sobre camino y vértices interiores (ver imagen 11). Para diferenciar se muestran las estadísticas básicas en las tabla 32 y tabla 33. En las figuras 12 y 13 se pueden apreciar los módulos y direcciones de los vectores.

Tabla 33 - Información básica para los puntos estratificados				
Caminos (45 puntos)			Interiores (55 puntos)	
Medida	Diferencia X	Diferencia Y	Diferencia X	Diferencia Y
Suma	115,779	-1402,830	782,009	-791,083
Máximo	134,560	343,005	232,381	168,165
Mínimo	-196,058	-225,373	-165,174	-182,796
Media	2,573	-31,174	14,218	-14,383
Mediana	9,764	-37,901	11,183	-7,567
Error cuadrático medio	66,901	103,152	74,355	73,4467

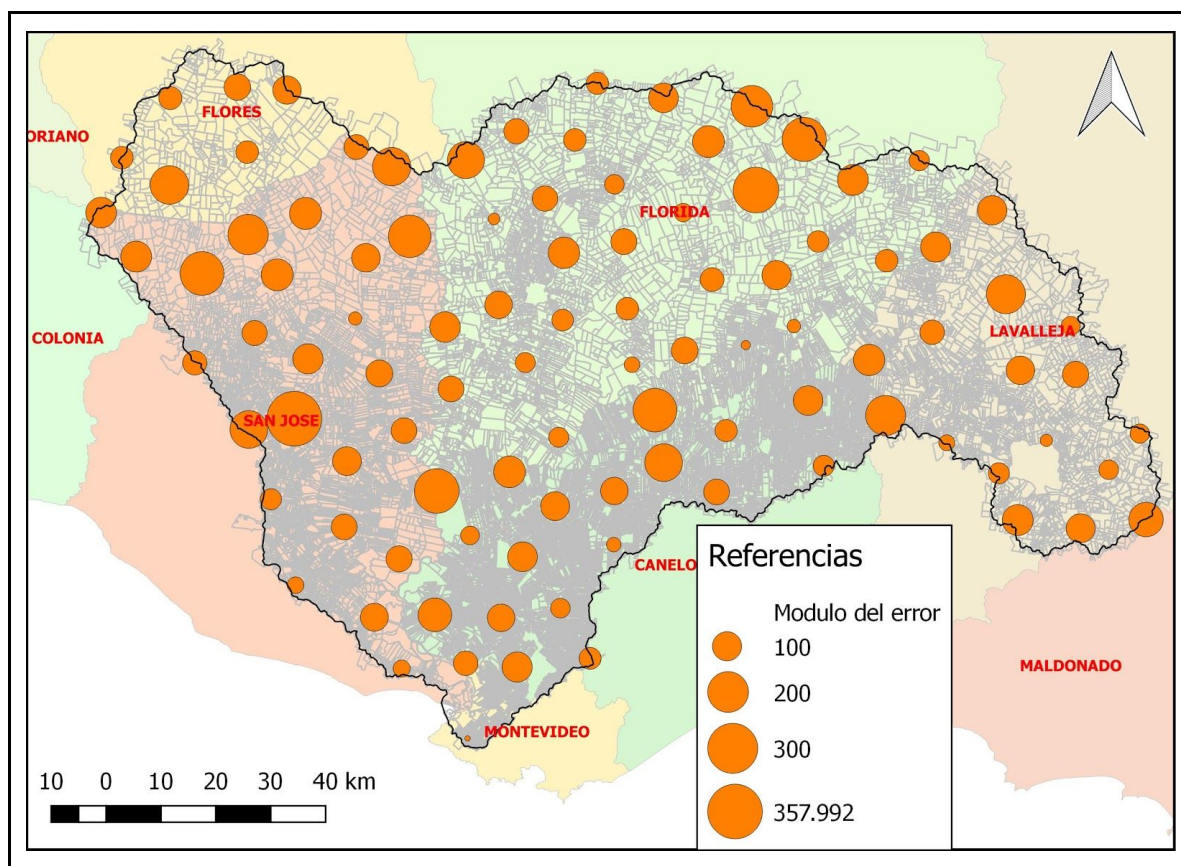


Figura 12 - Módulos del error posicional en cada punto para el catastro rural

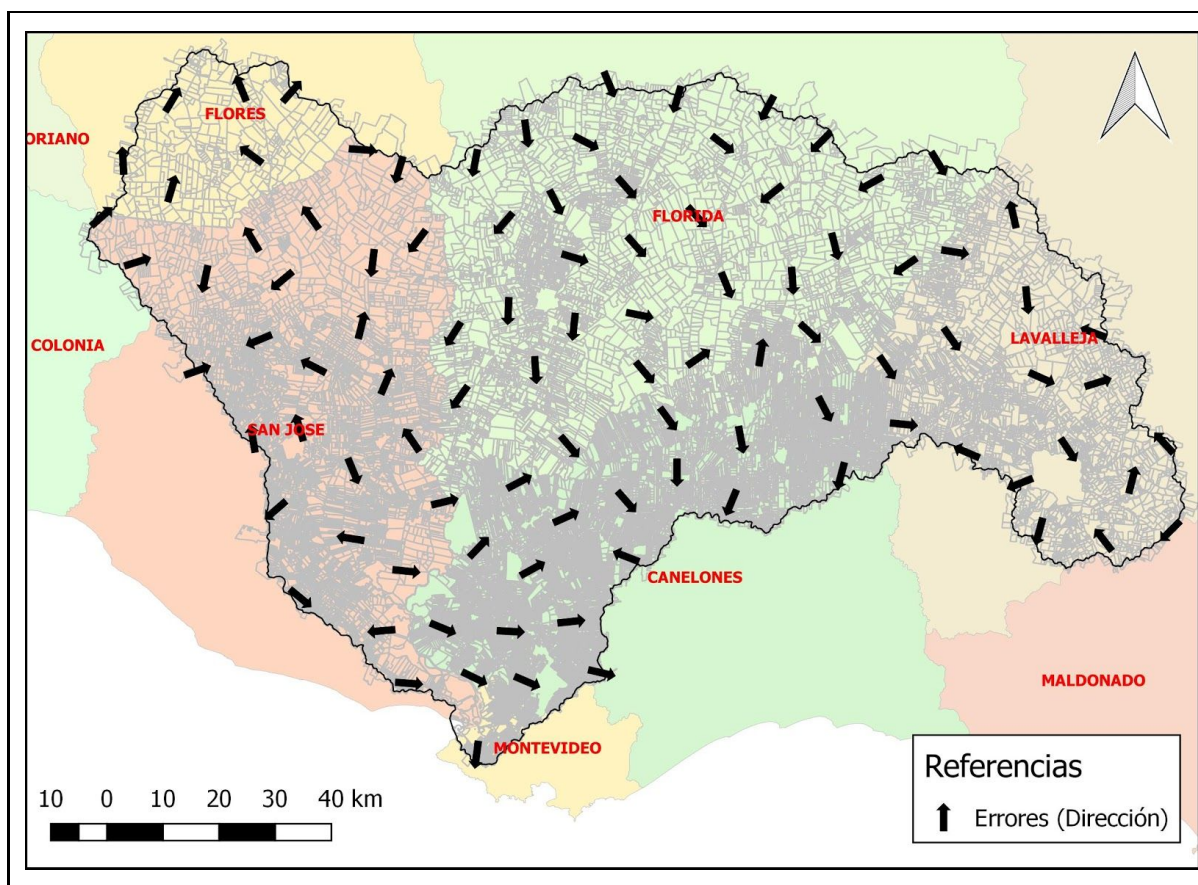


Figura 13 - Dirección de los errores en los puntos de control seleccionados para el catastro rural

Utilizando las medidas de la calidad establecidas en la norma ISO 19157 (Tabla 34 y 35).

Tabla 34 - Valor medio de las incertidumbres posicionales (Tabla D.29 Norma ISO 19157)		
1D X (metros)	1D Y (metros)	2D (metros)
55,071	65,292	94,052

Tabla 35 - Sesgo de las posiciones (Tabla D.30 Norma ISO 19157)		
ax (metros)	ay (metros)	ap (metros)
8,978	-21,939	23,705

Control posicional relativo

Se calcularon las 4950 combinaciones de distancias entre los puntos seleccionados y se realizó la diferencia entre distancias homólogas. La diferencia se expresa como porcentaje de la distancia considerada como verdadera.

A los efectos de caracterizar los valores de distancias, el valor máximo del universo de distancias trabajadas es de 198534,477 metros y el mínimo es de 7379,414 metros. Los valores de los principales parámetros se muestran en la tabla 36.

Tabla 36 - Información básica para los errores considerados en la exactitud relativa			
	Diferencia (m)	Diferencia (%)	Diferencia absoluta (%)
Suma	41769,837	61,553	808,468
Máximo	510,440	2,485	3,168
Mínimo	-453,401	-3,168	0,000
Media	8,438	0,012	0,163
Mediana	7,931	0,011	0,098
Error cuadrático medio	106,523	0,270	0,270

Para detectar outliers en las observaciones se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018).

P = 0,9999 → Error = $4,0 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 37 - Detección de outliers - Diferencia de distancias						
Iteración	Error	Tamaño muestra	Error máximo (+/- %)	Límite máximo (%)	Límite mínimo (%)	Cantidad de distancias eliminadas
0	$4,0 \times \sigma$	4950	1,080	1,093	-1,068	44
1	$4,0 \times \sigma$	4906	0,894	0,908	-0,880	29
2	$4,0 \times \sigma$	4877	0,843	0,857	-0,830	11
3	$4,0 \times \sigma$	4866	0,828	0,840	-0,816	9
4	$4,0 \times \sigma$	4857	0,816	0,820	-0,796	2
5	$4,0 \times \sigma$	4849	0,805	0,818	-0,793	0

Luego de quitar los outliers (ver tabla 37) la cantidad de observaciones es de 4849 pares de distancias. Se eliminaron el 2,04% de las observaciones.

Para determinar la normalidad de los errores en las distancias se utilizará el Test de Kolmogorov-Smirnov para un nivel de confianza del 95% (ver tabla 38).

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 38 - Parámetros del Test de Kolmogorov-Smirnov	
	Error X
N	4849
C_α	0,895
k(n)	69,637
D_n	0,076
D_α	0,013
Cumplimiento	No cumple

Para la aleatoriedad se utilizará el Test de Rachas (ver tabla 39) el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96.

Tabla 39 - Parámetros del Test de Rachas	
	Error X
N	4849
n_0	2275
n_1	2574
R	2081
μ_R	2416,282
σ_R	34,681
Z_R	9,667
Cumplimiento	No cumple

Luego de quitar los outliers si calculamos el error medio cuadrático 0,202%. Las diferencias entre las distancias no sigue una distribución normal por lo que en vez de calcular un intervalo de confianza se calcularon los percentiles 2,5% y el 97,5%. Los valores para estos parámetros son -0,423% y 0,479%.

Control posicional absoluto por puntos

Detección de outliers

Para aplicar el NSSDA primero se detectarán outliers en la componente X y luego en la componente Y (ver tablas 40 y 41). Para esto se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018). En la figura 14 se muestran los puntos detectados como outliers.

P = 0,995 → Error = $2,8 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 40 - Detección de outliers - Variable X						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Punto eliminado (metros)
0	$2,8 \times \sigma$	100	198,473	207,451	-189,495	31 y 69
1	$2,8 \times \sigma$	98	181,030	189,820	-172,240	45
2	$2,8 \times \sigma$	97	174,268	181,179	-167,358	Ninguno

Tabla 41 - Detección de outliers - Variable X						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Punto eliminado (metros)
0	$2,8 \times \sigma$	97	239,641	218,253	-261,029	65
1	$2,8 \times \sigma$	96	216,703	191,519	-241,886	Ninguno

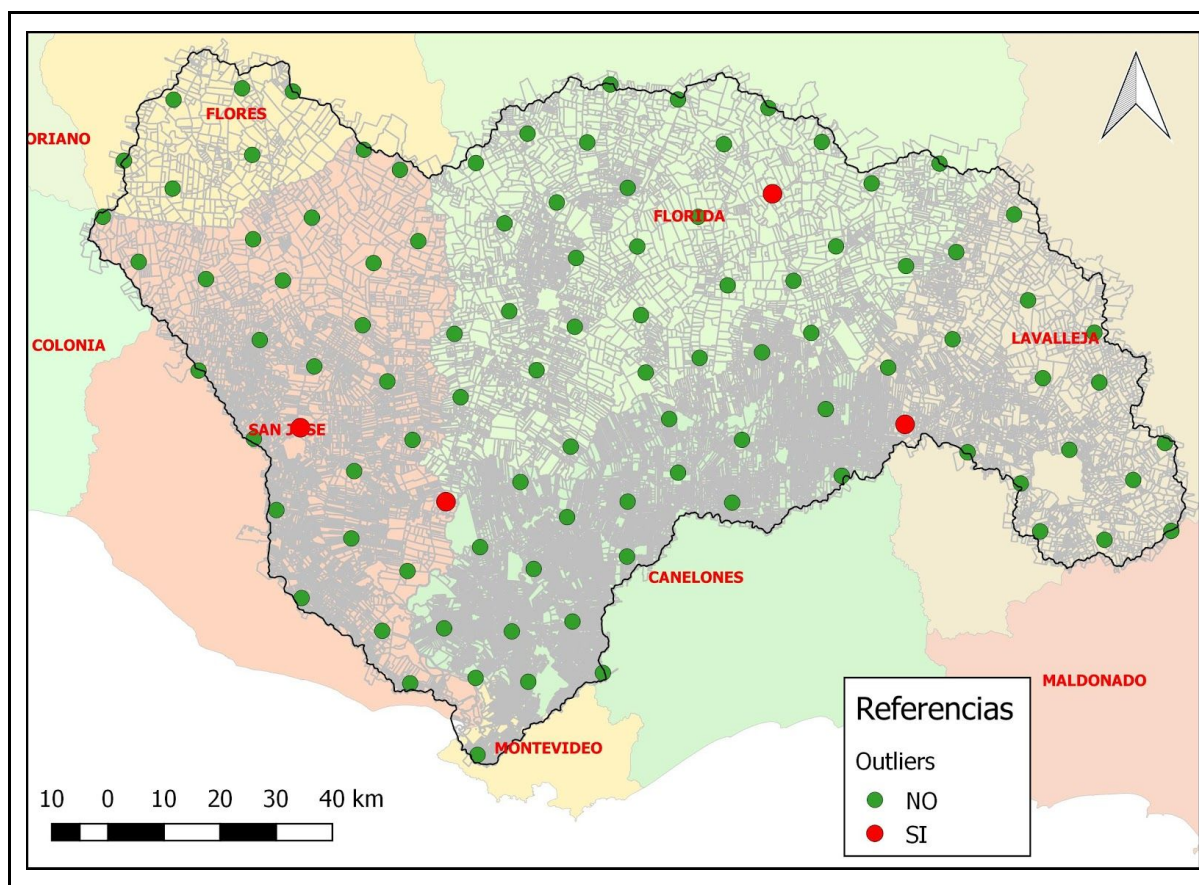


Figura 14 - Ubicación de los outliers para el parcelario rural

Control de normalidad

Para determinar la normalidad de los errores en las componentes estudiadas (X, Y) datos se utilizará el Test de Kolmogorov-Smirnov (ver tabla 42) para un nivel de confianza del 95%.

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 42 - Parámetros del Test de Kolmogorov-Smirnov		
	Error X	Error Y
N	96	96
C_α	0,895	0,895
k(n)	9,875	9,875
D_n	0,039	0,055
D_α	0,091	0,091
Cumplimiento	Cumple	Cumple

Control de aleatoriedad

Para la aleatoriedad se utilizará el Test de Rachas (ver tablas 43) el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96.

Tabla 43 - Parámetros del Test de Rachas		
	Error X	Error Y
N	96	96
n_0	40	64
n_1	56	32
R	52	47
μ_R	47,667	43,667
σ_R	4,736	4,326
Z_R	0,915	0,771
Cumplimiento	Cumple	Cumple

Control de independencia

Para la independencia se calculará el coeficiente de correlación de Pearson. El resultado para este coeficiente es de 0,1406, lo que da una baja correlación lineal. En la figura 15 se muestra la relación entre los errores en cada componente.

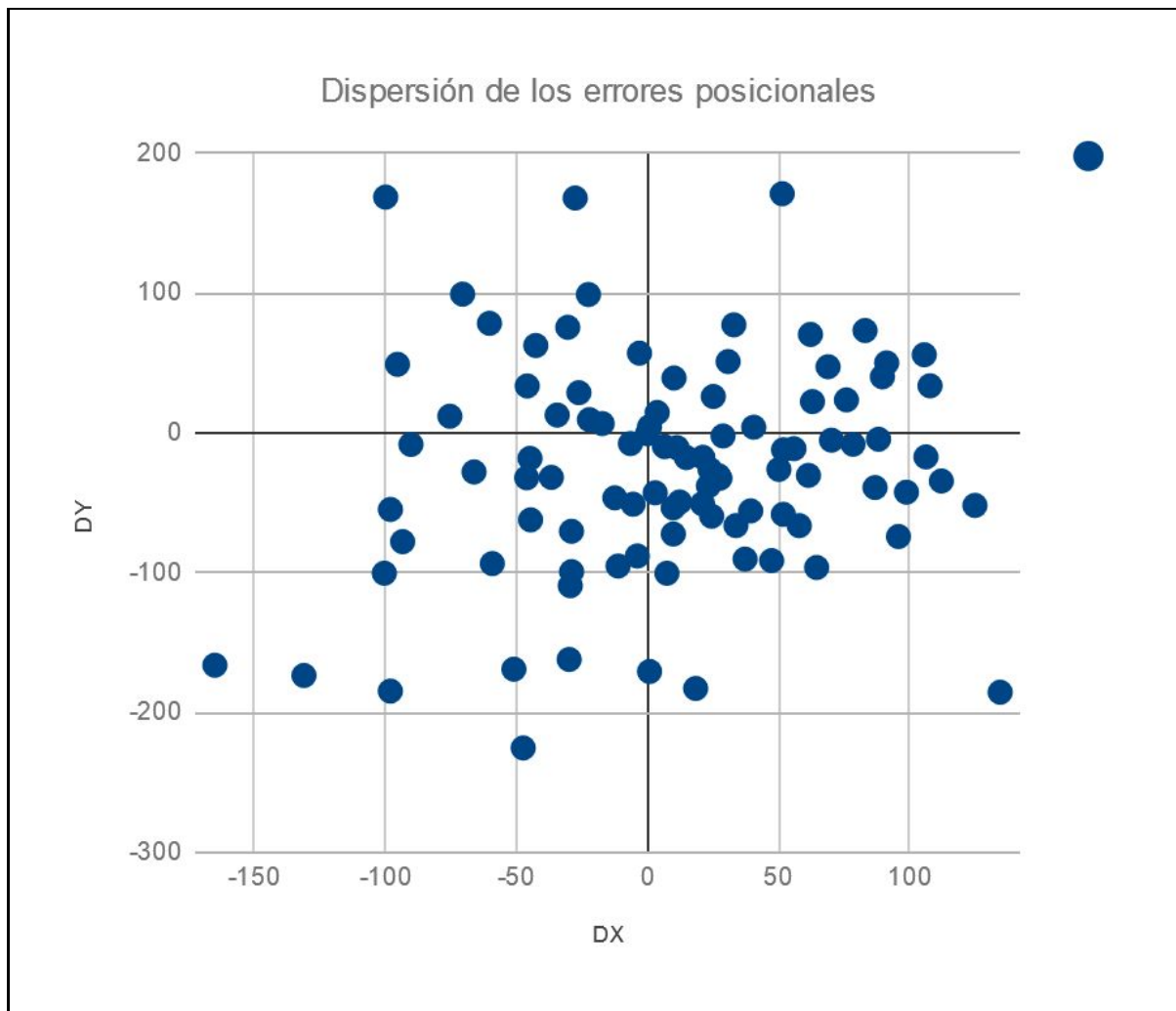


Figura 15 - Distribución de los errores posicionales para el parcelario rural

Control de sistematismo

Para comprobar si el error sistemático es aceptable se realiza un test sobre la media con el supuesto de varianza poblacional desconocida (ver tabla 44) para un 95% de nivel de confianza. Para esto las hipótesis son:

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$
$ t $ menor o igual a $t_{n-1,\alpha}$

Tabla 44 - Parámetros test de control de sistematismos		
	Error X	Error Y
N	96	96
Media	8,050	-25,184
D. E. Muestral	61,540	77,394
Estadístico t	1,282	3,188
Estadístico $t_{0,05;92}$	1,985	
Cumplimiento	Cumple	No cumple

Se verifica que los errores en la componente Y no pasa el test de sistematismos, por lo tanto se puede considerar que estos errores son importantes en esta componente para el parcelario rural y por lo tanto no se pueden considerar despreciables.

En resumen (ver tabla 45) se tiene la siguiente información para los errores calculados a partir de la muestra.

Tabla 45 - Resumen de resultados para el control del parcelario rural		
Para el 95% de confianza	Error X	Error Y
Detección de outliers	Si. Método de Chauvenet	
Normalidad	Cumple	Cumple
Aleatoriedad	Cumple	Cumple
Sistematismo	Cumple	No cumple
Independencia	El coeficiente de Pearson es de 0,1406	

National Standard for Spatial Data Accuracy (NSSDA)

Aunque no se cumplan algunas de las condiciones que se requieren (ver tabla anterior), el resultado de aplicar el NSSDA es:

Tabla 46 - Resultados para el NSSDA para el parcelario rural		
EMC	EMCmin / EMCmax	NSSDA
EMCx = 61,745 metros	0,762	174,703 metros
EMCy = 81,004 metros		

Resultados

Si se analizan los valores globales para los dos tipos de puntos tomados en consideración para este control se puede apreciar que para los puntos tomados sobre caminos presentan valores globales mayores sobre todo en la componente Y. Por ejemplo el rango de valores en esta componente para los puntos de camino es de 568 metros y para la componente X es de 428,439 o sea 25% menores.

Al comparar el resultado del trabajo de Faure et. al (2010) con el resultado del NSSDA para la cuenca del Río Santa Lucía con este trabajo resulta que los mismos son similares. Si bien para la cuenca del Santa Lucía es del orden de los 174 metros y el del trabajo mencionado es de 211 metros se puede ver que la diferencia entre ellos es del entorno del 17%. A los efectos de la aplicación del NSSDA hay que destacar que la componente Y no pasa el test de sistematismos.

PLAN CARTOGRÁFICO 1/50000

En este caso no se evaluará un conjunto de datos vectoriales como en los casos anteriores sino las láminas del Plan Cartográfico Nacional (PCN) 1:50000 del Instituto Geográfico Nacional en formato digital. Estas son producto del escaneo de las cartas originales papel por lo que están en formato raster y ya vienen georreferenciadas en el sistema SIRGAS ROU 98. Estas se pueden descargar desde <http://www.sgm.gub.uy/geoportal/cartografia/#pregunta2>

Según los metadatos de estas cartas (<http://www.sgm.gub.uy/geonetwork/srv/es/main.home>) no se especifica el procedimiento de escaneo ni la resolución a la que se hizo. Respecto a la georreferenciación de las mismas se utilizaron 9 puntos pero tampoco se especifica el procedimiento utilizado ni los errores obtenidos en este proceso para cada carta. La cantidad de cartas utilizadas para la zona de estudio fueron 33 lo que corresponde a aproximadamente el 10% del PCN.

Independientemente de esto, el proceso para el control de calidad que se utilizará será el mismo descrito en la metodología. Sin embargo hay que destacar que en este caso se presentarán otros factores de error en la determinación de los puntos homólogos, que se pasan a detallar:

El primero de ellos surge de la resolución en el escaneo de las cartas lo que determinará el tamaño del píxel una vez georreferenciadas las cartas. Por esto en el conjunto de datos a evaluar - por sus características - no tendremos puntos correctamente definidos y esto dependerá particularmente de la resolución con que fueron escaneadas las cartas 1/50000. Un aspecto vinculado a esto que también hay que considerar es la calidad del escaneo. Se detectaron algunas cartas donde algunos de los elementos (especialmente alambrados) que resultaron difícil de identificar por lo “claro” que están estos en la imagen. Ejemplo de esto es la carta H26 (figura 16).

El segundo error surge del hecho de que las cartas usan simbología para representar los elementos. El que las cartas cuenten con simbología que no respeta las dimensiones reales de los objetos hace que la representación afecte la evaluación de la calidad posicional.

El tercero de estos problemas, es producto de los errores de georreferenciación de las propias cartas. Se puede ver en varias de ellas que en la unión de dos cartas el canevas no coinciden. Esto revela un problema en la georreferenciación que afecta la calidad de este dato. También se detectaron zonas del territorio no cubierto por las cartas (figura 17).



Figura 16 - Problemas de escaneado en el PCN 1:50000 digitalizado y georreferenciado

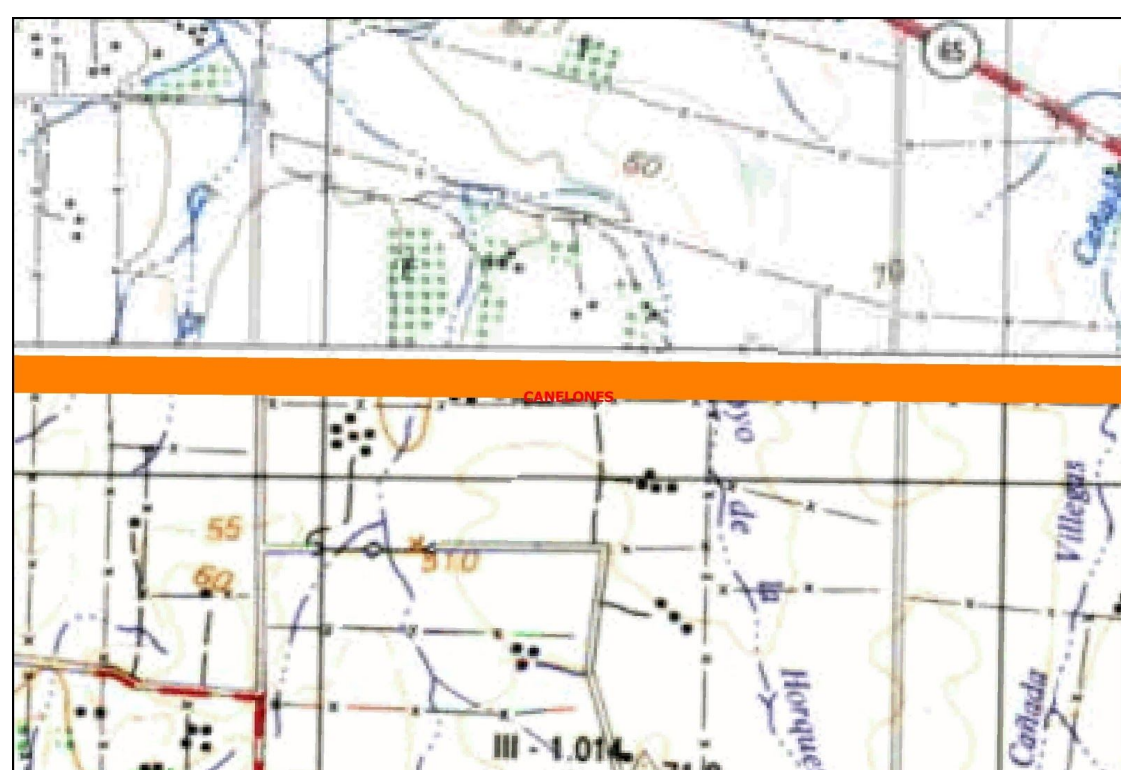


Figura 17 - Problemas de completitud en el PCN 1:50000 digitalizado y georreferenciado

También es importante destacar que las distintas cartas del PCN fueron confeccionadas en distintas épocas por lo que es esperable que no haya homogeneidad en los resultados de la evaluación posicional (Méndez y López, 2010).

En este mismo trabajo los autores establecieron el resultado del NSSDA para una carta (J28) en 100 metros considerando que los resultados eran adecuados para el procedimiento de creación de la misma.

Resultados para el total de los puntos

En esta instancia no se eliminaron outliers ya que se revisó y se verificó que los puntos homólogos seleccionados estén bien identificados. La eliminación de outliers se entiende que pueden desvirtuar el resultados de la evaluación ya que son puntos con errores que son pasibles de ser seleccionados en una muestra. Las diferencias de coordenadas surge de restar la coordenada del punto en el conjunto de datos geográficos menos la coordenada considerada como verdadera. La distribución de puntos y sus tipos se muestran en la figura 18 y en las figuras 19 y 20 los módulos y direcciones de los errores . En la tabla 47 se muestran las estadísticas básicas para estos puntos.

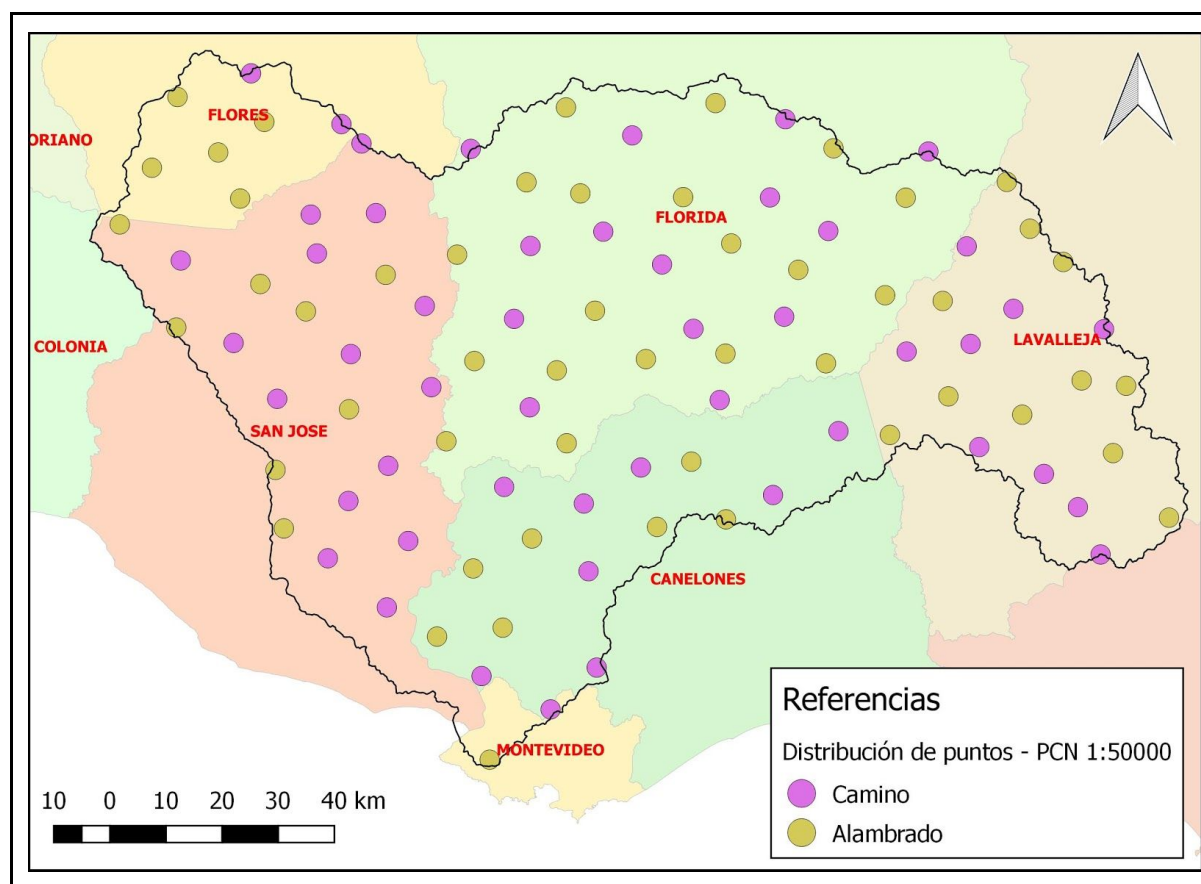


Figura 18 - Distribución de los puntos de control seleccionados para el PCN 1:50000

Tabla 47 - Información básica para el total de puntos		
Medida	Diferencia X (metros)	Diferencia Y (metros)
Suma	29,824	431,053
Máximo	106,093	217,911
Mínimo	-148,314	-110,601
Media	0,298	4,311
Mediana	6,112	1,000
Error cuadrático medio	45,820	44,177

Como se mencionó anteriormente se tomaron puntos de dos tipos: cruces de camino y alambrados representados en las cartas 1:50000. Se muestran las estadísticas básicas diferenciadas por tipo de punto (ver tabla 48).

Tabla 48 - Información básica para los puntos estratificados					
Cruces de caminos (49 puntos)			Alambrados (51 puntos)		
Medida	Diferencia X	Diferencia Y		Diferencia X	Diferencia Y
Suma	-81,912	-49,903		111,736	480,956
Máximo	72,274	68,651		106,093	217,911
Mínimo	-113,336	-82,212		-148,314	-110,601
Media	-1,672	-1,018		2,191	9,431
Mediana	2,488	1,475		7,671	-0,618
Error cuadrático medio	41,393	30,610		49,705	54,096

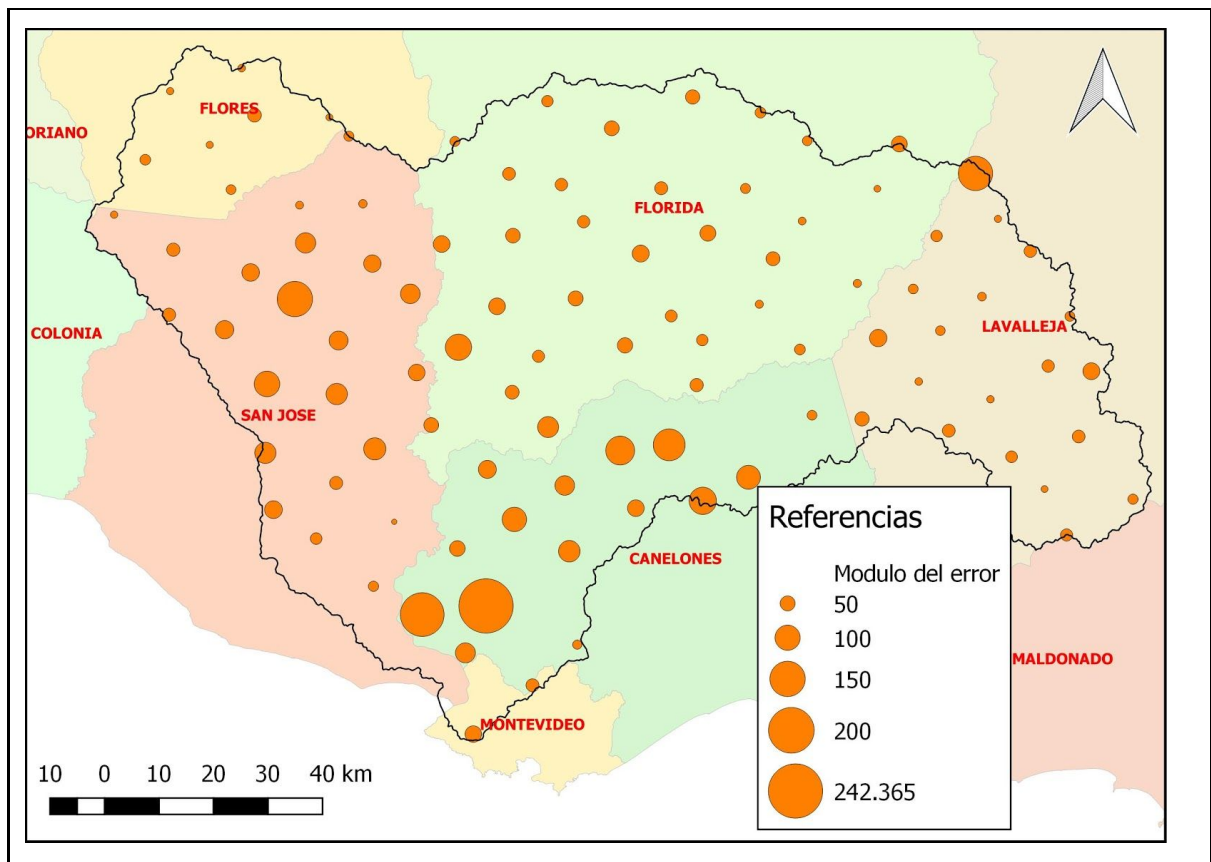


Figura 19 - Módulos del error posicional en cada punto para el PCN 1:5000

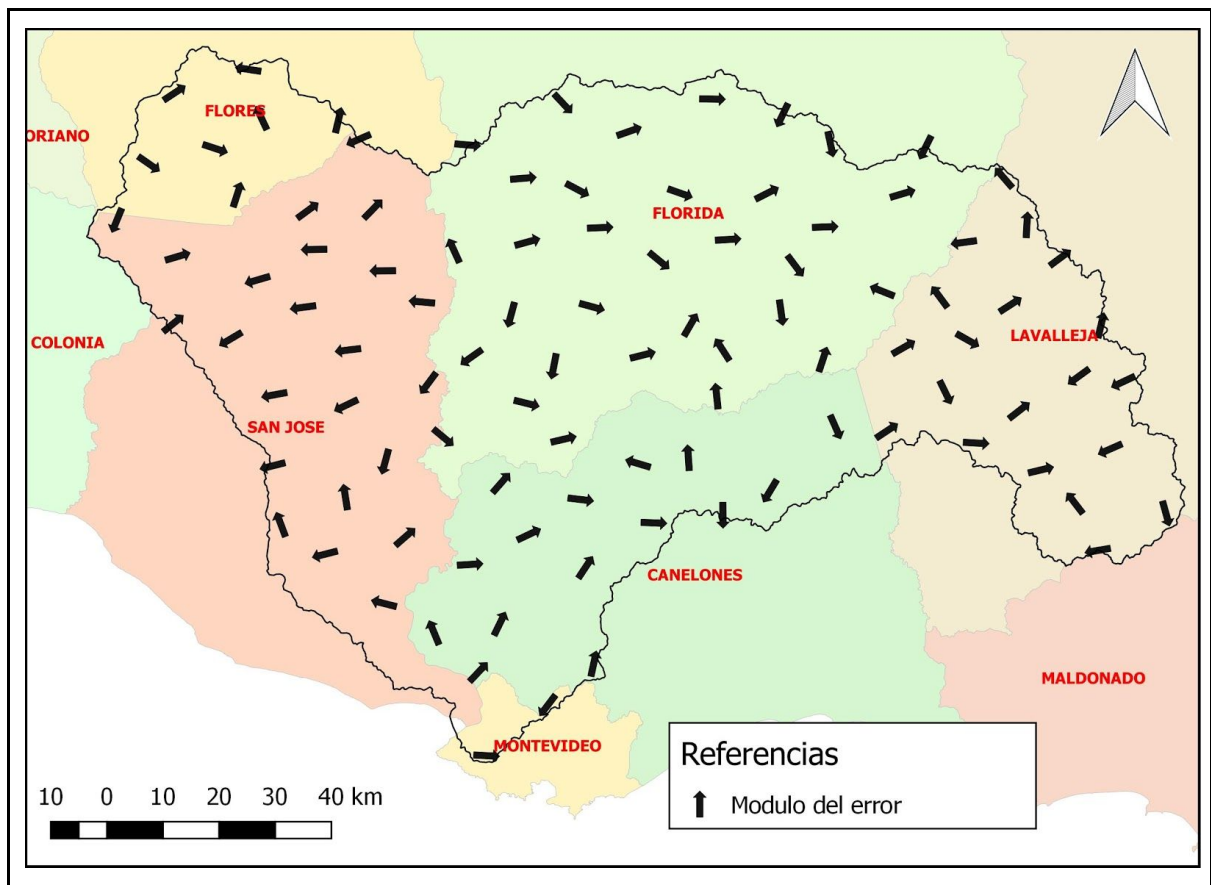


Figura 20 - Dirección de los errores en los puntos de control seleccionados para el PCN 1:50000

Utilizando las medidas de la calidad establecidas en la norma ISO 19157 (tablas 49 y 50).

Tabla 49 - Valor medio de las incertidumbres posicionales (Tabla D.29 Norma ISO 19157)		
1D X (metros)	1D Y (metros)	2D (metros)
35.949	27.837	50.980

Tabla 50 - Sesgo de las posiciones (Tabla D.30 Norma ISO 19157)		
ax (metros)	ay (metros)	ap (metros)
0.298	4.311	4.321

Control posicional relativo

Se calcularon las 4950 combinaciones de distancias entre los puntos seleccionados y se realizó la diferencia entre distancias homólogas. La diferencia se expresa como porcentaje de la distancia considerada como verdadera.

A los efectos de caracterizar los valores de distancias, el valor máximo del universo de distancias trabajadas es de 194033,473 metros y el mínimo es de 4980,281 metros. Además se muestran las estadísticas básicas en la tabla 51.

Tabla 51 - Información básica para los errores considerados en la exactitud relativa			
	Diferencia (m)	Diferencia (%)	Diferencia absoluta (%)
Suma	-7110,033	-10,219	489,827
Máximo	285,587	1,416	1,862
Mínimo	-224,552	-1,862	0,000
Media	-1,436	-0,002	0,099
Mediana	-1,223	-0,002	0,061
Error cuadrático medio	66,022	0,162	0,162

Para detectar outliers en las observaciones se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia entre las coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018).

$P = 0,9999 \rightarrow \text{Error} = 4,0 \times \sigma$ en todos los casos tanto para X como para Y

Tabla 52 - Detección de outliers - Diferencia de distancias						
Iteración	Error	Tamaño muestra	Error máximo (+/- %)	Límite máximo (%)	Límite mínimo (%)	Cantidad de distancias eliminadas
0	$4,0 \times \sigma$	4950	0,649	0,647	-0,651	40
1	$4,0 \times \sigma$	4910	0,541	0,538	-0,544	30
2	$4,0 \times \sigma$	4880	0,510	0,507	-0,514	9
3	$4,0 \times \sigma$	4871	0,503	0,499	-0,506	6
4	$4,0 \times \sigma$	4865	0,498	0,494	-0,502	4
5	$4,0 \times \sigma$	4861	0,495	0,491	-0,499	1
6	$4,0 \times \sigma$	4860	0,494	0,490	-0,498	0

Luego de quitar los outliers (ver tabla 52) la cantidad de observaciones es de 4860 pares de distancias. Se eliminaron el 1,82% de las observaciones.

Para determinar la normalidad de los errores en las distancias se utilizará el Test de Kolmogorov-Smirnov (tabla 53) para un nivel de confianza del 95%.

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 53 - Parámetros del Test de Kolmogorov-Smirnov	
	Error X
N	4860
C_α	0,895
k(n)	69,716
D_n	0,072
D_α	0,013
Cumplimiento	No cumple

Para la aleatoriedad se utilizará el Test de Rachas (tabla 54) el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0,025}$ y su valor es 1,96.

Tabla 54 - Parámetros del Test de Rachas	
	Error X
N	4860
n₀	2481
n₁	2379
R	1928
μ_R	2429,930
σ_R	34,838
Z_R	14,408
Cumplimiento	No cumple

Luego de quitar los outliers si calculamos el error medio cuadrático 0,124%. Las diferencias entre las distancias no sigue una distribución normal por lo que en vez de calcular un intervalo de confianza se calcularon los percentiles 2,5% y el 97,5%. Los valores para estos parámetros son -0,280% y 0,266%.

Control posicional absoluto por puntos

Detección de outliers

Para aplicar el NSSDA primero se detectarán outliers en la componente X y luego en la componente Y (ver tablas 55 y 56). Para esto se utilizará el criterio de Chauvenet el cual establece que se deben eliminar las observaciones cuyos errores aparentes (diferencia de coordenadas del punto - la media para la componente) tengan una probabilidad menor a $1-1/2n$ (Albagli y Martín, 2018). La ubicación de los outliers se muestra en la figura 21.

P = 0,995 → Error = 2,8 x σ en todos los casos tanto para X como para Y

Tabla 55 - Detección de outliers - Variable X						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Punto eliminado (metros)
0	2,8 x σ	100	128.941	129.239	-128.643	2
1	2,8 x σ	99	122.518	124.318	-120.719	Ninguno

Tabla 56 - Detección de outliers - Variable Y						
Iteración	Error	Tamaño muestra	Error máximo (+/- metros)	Límite máximo (metros)	Límite mínimo (metros)	Punto eliminado (metros)
0	$2,8 \times \sigma$	99	124.169	128.720	-119.618	45, 61, 65
1	$2,8 \times \sigma$	96	90.216	89.424	-91.008	18, 98
2	$2,8 \times \sigma$	94	79.174	78.375	-79.973	4, 43
3	$2,8 \times \sigma$	92	72.566	73.516	-71.616	Ninguno

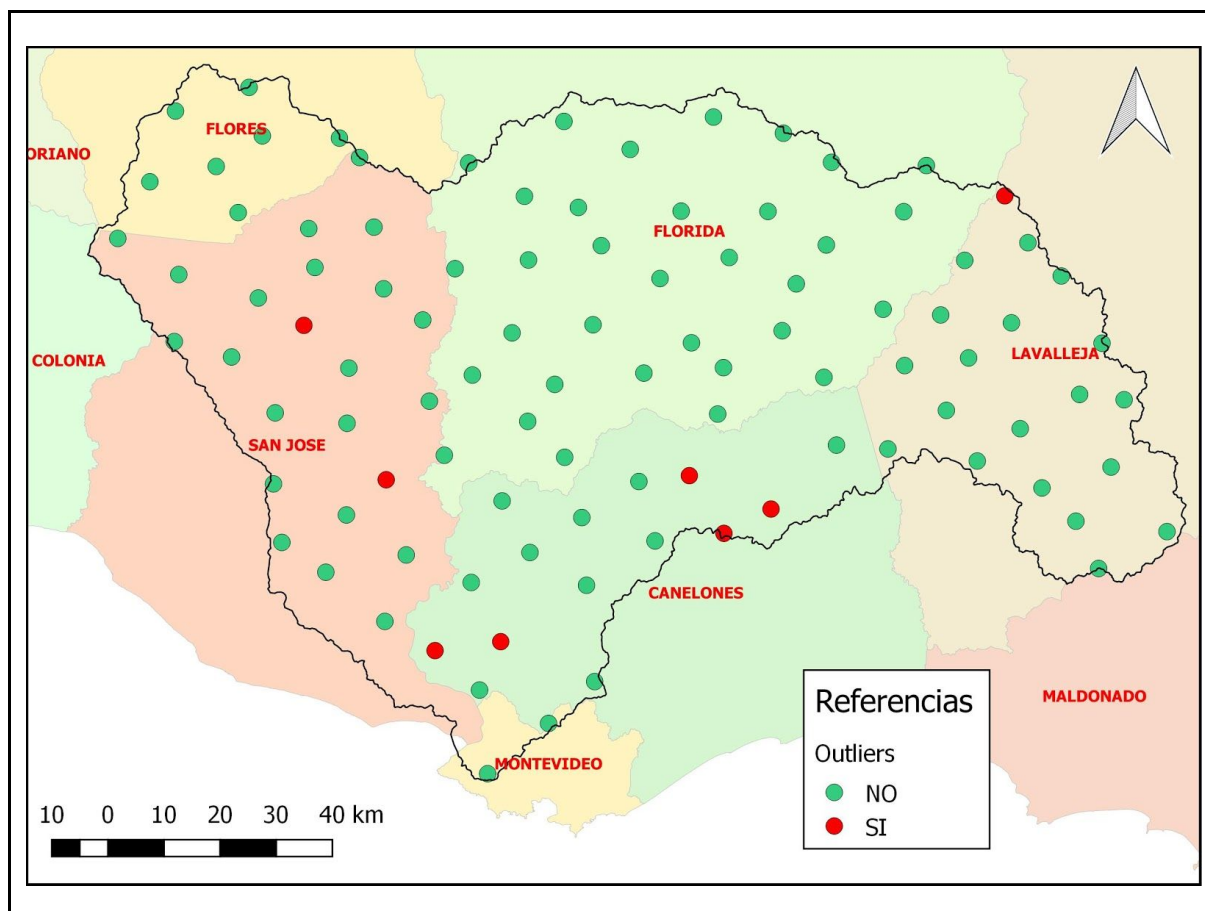


Figura 21 - Ubicación de los outliers para el PCN 1:50000

Control de normalidad

Para determinar la normalidad de los errores en las componentes estudiadas (X, Y) datos se utilizará el Test de Kolmogorov-Smirnov (tabla 57) para un nivel de confianza del 95%.

H_0 : Los datos de la muestra se ajustan a una distribución normal
H_1 : Los datos de la muestra no se ajustan a una distribución normal
$ D_n $ menor o igual a D_α

Tabla 57 - Parámetros del Test de Kolmogorov-Smirnov		
	Error X	Error Y
N	92	92
C_α	0,895	0,895
k(n)	9,875	9,875
D_n	0.076	0,069
D_α	0,093	0,093
Cumplimiento	Cumple	Cumple

Control de aleatoriedad

Para la aleatoriedad se utilizará el Test de Rachas (tabla 58) el cual se verificará a un 95% de nivel de confianza por lo que el estadístico de contraste es $Z_{1-0.025}$ y su valor es 1,96.

Tabla 58 - Parámetros del Test de Rachas		
	Error X	Error Y
N	92	92
n₀	36	45
n₁	56	47
R	42	50
μ_R	44.826	46.978
σ_R	4.542	4.767
Z_R	0,622	0,634
Cumplimiento	Cumple	Cumple

Control de independencia

Para la independencia se calculará el coeficiente de correlación de Pearson. El resultado para este coeficiente es de 0.2783, lo que da una baja correlación lineal. En la figura 22 se muestra la distribución de los errores.

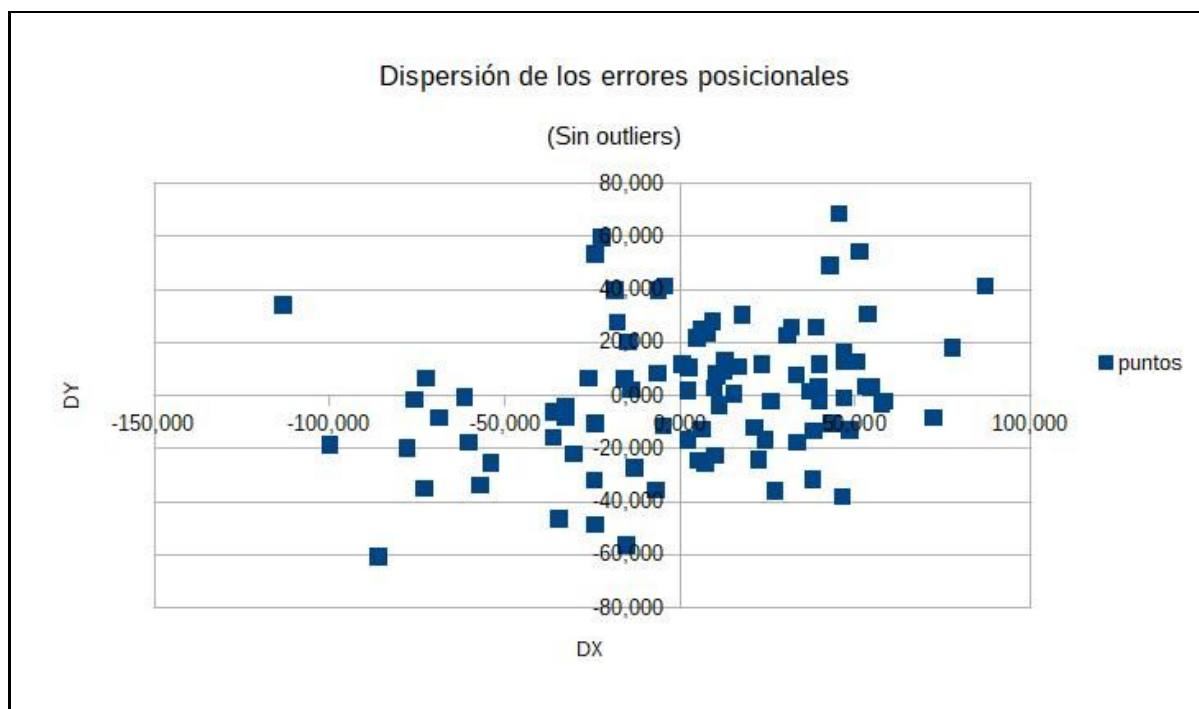


Figura 22 - Distribución de los errores posicionales para el PCN 1:50000

Control de sistematismo

Para comprobar si el error sistemático es aceptable se realiza un test sobre la media con el supuesto de varianza poblacional desconocida (tabla 59) para un 95% de nivel de confianza. Para esto las hipótesis son:

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$
$ t $ menor o igual a $t_{n-1,\alpha}$

Tabla 59 - Parámetros test de control de sistematismos		
	Error X	Error Y
N	92	92
Media	3.467	0.950
D. E. Muestral	41.707	25.916
Estadístico t	0.797	0.352
Estadístico $t_{0.05;92}$	1,986	
Cumplimiento	Cumple	Cumple

En resumen (ver tabla 60) se tiene la siguiente información para los errores calculados a partir de la muestra.

Tabla 60 - Resumen de resultados para el control del PCN 1:50000		
Para el 95% de confianza	Error X	Error Y
Detección de outliers	Si. Método de Chauvenet	
Normalidad	Cumple	Cumple
Aleatoriedad	Cumple	Cumple
Sistematismo	Cumple	Cumple
Independencia	El coeficiente de Pearson es de 0.2783	

National Standard for Spatial Data Accuracy (NSSDA)

En este caso se cumplen todas las condiciones que se requieren (ver tabla 60), por lo que el resultado de aplicar el NSSDA es:

Tabla 61 - Resultados para el NSSDA para el parcelario rural		
EMC	EMCmin / EMCmax	NSSDA
EMCx = 41.624 metros	0.620	82.508 metros
EMCy = 25.793 metros		

Si bien se cumple la condición para EMCmin/EMCmax se puede ver que la misma está en el límite.

Resultados

Como primer análisis se puede ver que para el caso del Plan Cartográfico Nacional 1:50000 los errores pasan todos los test a la que se sometieron. De todas formas se puede ver en la figura xx que en algunas zonas hay grupos de vectores que apuntan en la dirección.

El resultado del NSSDA es del orden de los 82 metros. En el caso del antecedente consultado el error es de 100 metros lo que hace que los valores sean similares o se encuentre dentro de lo esperado.

Algo a destacar es que si se analizan los valores para los distintos tipos de puntos resulta que el rango de los errores es mucho mayor para los puntos tomados sobre alambrados que los tomados sobre caminos. Esto es más significativo para el caso de la componente Y donde la diferencia en el rango de los valores es de más del doble que la componente X. Puede considerarse este resultado interesante para seguir investigando la causa del mismo ya que es esperable que los errores en los alambrados sean menores que en los caminos básicamente debido a que su representación en la carta se podría asimilar a una línea sin espesor.

Contrario a lo que pasaba en otros casos estudiados los outliers no estaban solo dentro de uno de los tipos de puntos aunque si hay una predominancia de los errores de outliers que estaban ubicados sobre límites de alambrados. Si bien se revisaron los puntos detectados como outliers para verificar su correcta ubicación, en algunos de los casos podría existir cierta duda ya que no todos los alambrados aparecen en la carta 50000. Por otro lado la diferencia de tiempo entre la producción de las cartas 50000 y el vuelo hace que las realidades sean muy distintas lo que dificulta la elección de puntos homólogos.

ANÁLISIS GENERAL

Se presenta a continuación un resumen de los principales indicadores calculados para los distintos conjuntos de datos geográficos. El primero de ellos es el que surge de aplicar la medida “Valor medio de las incertidumbres posicionales” que surge de la norma ISO 19157 tabla D.29. Los resultados para los 4 conjuntos de datos evaluados son:

Tabla 62. Valor medio de las incertidumbres posicionales			
Conjunto de datos	1D X (metros)	1D Y (metros)	2D (metros)
Zonas censales	34	28	47
Caminería	5	9	11
Parcelario rural	55	65	94
PCN 1/50000	36	28	51

Resulta llamativo la similitud de los valores medios de las incertidumbres para las zonas censales y el PCN 1/50000. Consultada la responsable de la cartografía del INE resulta que el conjunto de datos de las zonas censales se construyó utilizando como base la digitalización de las cartas del PCN. Esta vinculación entre conjuntos de datos explica los valores similares.

Tabla 63. Sesgo de las posiciones			
Conjunto de datos	1D X (metros)	1D Y (metros)	2D (metros)
Zonas censales	4	4	6
Caminería	-3	7	8
Parcelario rural	9	-22	24
PCN 1/50000	0	4	4

LLama la atención el fuerte sesgo en dos de las capas evaluadas, esto se ve confirmado al ver la tabla 65 donde se muestra que ambas capas no pasan el test de sistematismos.

La primera de ellas es la capa de caminería. Viendo la figura 8 se puede ver gráficamente esta tendencia. Si bien 8 metros en el valor del sesgo puede parecer poco, comparado con el valor del NSSDA para esa capa resulta que el módulo del sesgo ronda el 45% del NSSDA.

En el caso del parcelario si bien la relación entre el NSSDA y el valor del sesgo no es tan elevada como en el caso de la caminería, igual se verifica que existe un sesgo importante en la componente Y (ver figura 13).

Tabla 64. Exactitud relativa				
Conjunto de datos	Distancia media (metros)	ECM (%)	P2.5 (%)	P97.5 (%)
Zonas censales	-7	0,116	-0,268	0,252
Caminería	-1	0,016	-0,040	0,030
Parcelario rural	8	0,202	-0,423	0,479
PCN 1/50000	-1	0,124	-0,280	0,266

En el caso de la exactitud posicional vale hacer la aclaración de que los porcentajes mostrados en la tabla 64, no pueden ser aplicados directamente a cualquier distancia que se mida en el conjunto de datos correspondiente ya que estos valores son generales y no representan toda la gama de posibles distancias que se puedan obtener.

Como se puede ver la caminería es la que presenta los valores más pequeños. Esto se puede deber a la forma de generación del conjunto de datos ya que los mismos fueron generados a partir de un relevamiento directo. Nuevamente en este caso los valores del ECM y de los percentiles para las capas de zonas censales y del PCN 1/50000 son similares lo que se puede deber a la relación que existe entre ellos (las zonas censales utiliza para su creación los elementos digitalizados del PCN).

En todos los casos estudiados los errores encontrados fueron superiores a los establecidos o declarados por los proveedores de los datos (ver tabla 66). En el caso de los datos de zonas censales y parcelario esto se puede deber al proceso de creación en el cual se usaron conjuntos de datos analógicos existentes. También en la caminería a pesar de ser un producto en su mayor parte resultado del relevamiento directo se aprecia que el error declarado es superado en la evaluación.

En el caso del PCN la diferencia se puede deber al proceso de digitalización y a la georreferenciación de las cartas. No se muestran en los metadatos de esas cartas los errores resultantes del ajuste de cada carta.

Solo en el caso del PCN 1:50000 se cumplen todas las condiciones para poder aplicar el estándar NSSDA. Esto podría estar indicando algún vicio en el proceso de creación de aquellos datos que no cumplen con estas condiciones.

Tabla 65 - Resumen de controles previos a aplicar el NSSDA							
	X			Y			Pearson (X,Y)
	Normalidad	Aleatoriedad	Sistematismo	Normalidad	Aleatoriedad	Sistematismo	
Zonas censales	No cumple	Cumple	Cumple	Cumple	Cumple	Cumple	-0,0597
Caminería	No cumple	Cumple	No cumple	No cumple	Cumple	No cumple	-0,3182
Parcelario	Cumple	Cumple	Cumple	Cumple	Cumple	No cumple	0,1406
PCN 1/50000	Cumple	Cumple	Cumple	Cumple	Cumple	Cumple	0,2783

Tabla 66 - Resumen de resultados del NSSDA		
Conjunto de datos	EMCmin/EMCmax	Resultado (en metros)
ZONAS CENSALES RURALES	0,898	77
CAMINERÍA	0,465	18
PARCELARIO CATASTRAL RURAL	0,762	175
PLAN CARTOGRÁFICO 1/50000	0,620	83

BIBLIOGRAFÍA

- Albagli, Karen Janet; Martín, Benjamín Andrés (2018). "Control de Calidad posicional de cartografía digital con métodos estadísticos, según normas ISO 19100". Facultad de Ingeniería, Universidad Nacional de San Juan, Argentina. Tesis de grado.
- Ariza López, Francisco Javier y Atkinson Gordo, Alan David (2008). "*Variability of NSSDA Estimations*". Journal of Surveying Engineering. Volumen 134, número 2, páginas 39 a 44.
- Ariza F.J. (2009). Calidad de la IG: Introducción. En la 2ª Edición del Curso de Experto Universitario en Evaluación de la Información Geográfica, Universidad de Jaén, Jaén.
- Ariza López, Francisco Javier y García Balboa, José Luis (2010). "Evaluación de las componentes de la calidad de la información geográfica". En Título de Experto Universitario en Evaluación de la Calidad de la Información Geográfica (2ª Edición). Universidad de Jaén, España.
- Faure, Jorge et al (2010). "Evaluación de la exactitud posicional horizontal absoluta del parcelario rural vectorial del Departamento de Lavalleja". Instituto de Agrimensura, Facultad de Ingeniería, Udelar y Dirección Nacional de Catastro. Informe técnico.
- FGDC, Federal Geographic Data Committee (1998). *Geospatial Positioning Accuracy Standards. Part 3: National Standard for Spatial Data Accuracy*. National Spatial Data Infrastructure.
- Instituto Geográfico Nacional. www.sgm.gub.uy. Sitio visitado en mayo de 2019.
- Instituto Nacional de Estadísticas. www.ine.gub.uy. Sitio visitado en mayo de 2019.
- ISO 19157:2013 (2013). Información Geográfica. Calidad de datos. International Organization for Standardization. Versión oficial, en español, de la Norma Europea EN ISO 19157:2013, que a su vez adopta la Norma Internacional ISO 19157:2013.
- Méndez Baillo, Rodolfo y López Vázquez, Carlos (2010). "Evaluación de exactitud posicional horizontal y vertical de la Cartografía Oficial a escala 1:50000". I Congreso Uruguayo de Infraestructura de Datos Espaciales. ISBN 978-9974-8191-9-1.
- MTOP. Infraestructura de Datos Espaciales del MTOP, Metadatos. <http://geoportal.mtop.gub.uy>. Sitio visitado en mayo de 2019.