

## PROCESADO SIG DE DATOS LIDAR PARA LA CREACIÓN Y EVALUACIÓN DE MDTs Y FILTRADO DE PUNTOS OUTLIERS

J. L. Gil y A. Fernández-Sarria.

*Universidad Politécnica de Valencia. Camino de Vera s/n, 46022 Valencia. jogiye@gmail.com*

### RESUMEN

Las nuevas tecnologías LIDAR han obligado a abordar nuevas y distintas estrategias de análisis, procesamiento y filtrado debido a la numerosa cantidad de información proporcionada por esta tecnología. Este artículo trata sobre dicha metodología de trabajo testando los nuevos algoritmos desarrollados sobre SIG GRASS para la automatización de dicho filtrado de las observables LIDAR así como la eliminación de los observables erróneos o puntos outlier y comparando los modelos digitales de terreno resultantes tras dichos procesos.

### ABSTRACT

New LIDAR technologies have obliged to deal with new and different strategies to analyze, processing and filtering due to the big quantity of information proportioned by this technology. This article talks about this kind of work methodology testing the new algorithms developed in SIG GRASS for the filtering automation of this LIDAR information and the removal of the errors or outlier point and doing the comparison between the resulting digital terrain models at the end of the process.

**Palabras clave:** LIDAR, SIG, MDT, filtrado, outlier.

### INTRODUCCIÓN

Cada vez más, la tecnología LIDAR (Wehr y Lohr 1999) forma parte de los nuevos desarrollos cartográficos tanto para la creación de cartografía como para estudios en los que ésta interviene.

Las nuevas técnicas LIDAR tienen la problemática de no poder distinguir entre los puntos terreno y los puntos no pertenecientes a éste. La generación del MDT necesita un método para realizar esta distinción. Tomando como partida los algoritmos utilizados en proyectos SIG del Politécnico de Milano, se permitió desarrollar el software para la filtración y elaboración de los datos LIDAR para la posterior creación de MDTs de zonas urbanas (Brovelli, Cannata 2002). Estos algoritmos que se utilizaron para la generación de los MDTs y del filtrado de puntos LIDAR son los mismos que se utilizan para el desarrollo de este trabajo.

En este artículo se describirán metodologías y criterios de filtrado de los datos observados con LIDAR realizando una introducción metodológica al estudio. Tras esta pequeña introducción se realizará un estudio de los pasos spline y un estudio del threshold para eliminar los puntos outlier, utilizando en este apartado dos metodologías diferentes.

### DESARROLLO

#### Zona de estudio

La zona de estudio pertenece a la ciudad de Olbia, Italia, con una superficie de estudio de 4 Km<sup>2</sup> dispuesta en un cuadrado de 2 Km. de lado. Esta región corresponde a una zona cercana a la costa, por lo que el terreno no presenta mucha altura y es un terreno bastante liso sin cambios bruscos en su pendiente. En contraste, gran parte de los datos corresponden a superficies urbanas con gran densidad de puntos de último pulso debido a torres de comunicación, construcciones y tejados con cubiertas poco comunes. Estas zonas son zonas puntuales y concretas. A las afueras se encuentran áreas de masas forestales.

#### Datos LIDAR iniciales

Los datos para el estudio provienen de observaciones LIDAR en la proyección UTM, huso 32 norte con elipsoide WGS84. El total de puntos medidos es de 9519892, tanto de primer como de último pulso, y su densidad en el terreno es de 2,4 puntos por metro cuadrado. Esto permite desarrollar un modelo con una resolución de 2 metros para la posterior comparación con los MDTs ya existentes de la zona. Los datos LIDAR han sido proporcionados por la Universidad Politécnica di Milano, y las características técnicas del láser son: una frecuencia de 70.000 Hz, un ángulo máximo de escaneado de 20 deg., una frecuencia de escaneado

de 36 Hz., un ancho de banda de 1380 metros y una longitud de pasada de 1050 metros.

Para la comparación de los MDTs procesados, se dispone de otro modelo digital del con una resolución de píxel de 2 metros también proporcionado por la Politecnica di Milano y una ortofoto con resolución de 0,125 metros para un mejor estudio y conocimiento de la zona.

#### Algoritmos empleados para filtrado

El algoritmo estudiado trabaja sobre SIG GRASS y actúa dividido en varios comandos. En un principio este algoritmo de filtrado fue diseñado para datos correspondientes a zonas urbanas sin información de reflectancia en el pulso (Antolin y Brovelli 2006). Este artículo se centra en un comando del algoritmo llamado '*v.outlier*' que elimina puntos erróneos de las mediciones LIDAR.

#### Introducción metodológico al estudio de filtrado

El filtrado de los datos LIDAR comienza con la eliminación de puntos outlier. Para ello, se ha utilizado el comando '*v.outlier*' comentado en el apartado anterior.

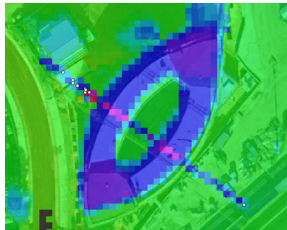


Figura 1.- Edificio en obras, puntos first 10 x 10 con outliers.

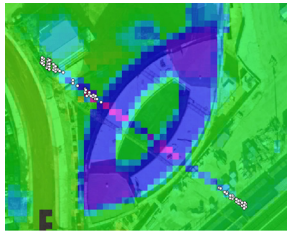


Figura 2.- Edificio en obras, puntos first 20 x 20 con outliers.

El comando se ejecuta fijando el valor del umbral en un valor estándar de 30 metros y utilizando valores de spline iniciales de 10 x 10 y 20 x 20 metros en direcciones norte-sur y este-oeste con un valor de parámetro regularizador de Tikhonov de 0.01 y 0.001 tanto para los archivos de puntos first como last. En las imágenes (Figura 1 y Figura 2) se muestra algún ejemplo de la diferencia que se

produce en la captación de puntos outlier con los diferentes pasos de spline.

Una vez que el archivo ha sido eliminado de puntos outlier se generan interpolaciones para realizar las diferencias entre los archivos "limpios" de errores y los archivos originales, con el posterior estudio estadístico con sus errores medios cuadráticos y los intervalos correspondientes a 3 veces la desviación típica tomando estos como valores máximos dentro de una distribución normal. Esto se realiza para cuantificar el número de celdas de la imagen que se sitúan fuera de este dominio y observar si se ha realizado la eliminación de puntos outlier correctamente.

Un error de 3 sigma, corresponde a la curva estadística de la distribución normal con un 99.73%, lo que permite un error del 0.27%. Con este estudio se buscan los parámetros adecuados que se deben introducir al comando '*v.outlier*' para que el error cometido no sea superior a este porcentaje.

#### Estudio del paso de spline

Para continuar con el filtrado de puntos LIDAR se prueban diferentes pasos de spline únicamente en los archivos de puntos de primer pulso ya que es en este archivo donde se encuentra un mayor número de puntos outlier. A medida que se aumenta el paso de spline en direcciones norte-sur y oeste-este el algoritmo suaviza y alisa cada vez más la superficie con la que interactúan los outlier.

Tabla 1.- Número de puntos outlier.

sie-sin	$\lambda$	Threshold	pts_outlier
10x10	0,01	30	138
20x20	0,01	30	226
30x30	0,01	30	229
40x40	0,01	30	234
50x50	0,01	30	249

Como se puede observar en la tabla y en las imágenes (Tabla 1 y Figuras 1 y 2), al aumentar el paso de spline aumenta el número de puntos outlier situándose estos en las mismas zonas. Gracias al análisis visual de las imágenes y de las ortofotos en dichas zonas donde aumenta la densificación, se ha comprobado que la mayor parte de los puntos pertenecen al modelo digital de elevaciones. Es por esto que cuanto más se aumente el paso de spline mayor número de puntos MDS será eliminado, por lo que se toma un paso de spline de 10 y de 20 metros para continuar con el estudio. (Tabla 2).

Al fijar el valor del paso de spline en 20 metros se vuelve a efectuar el estudio con las interpolaciones realizadas para calcular el error que se comete cuantificando el número de celdas píxel

fuera del intervalo de la distribución normal, modificando esta vez el intervalo para ver si se encuentra un error tolerable para la distribución.

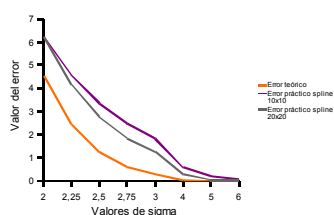


Figura 3.- Gráfico de porcentaje de errores.

A pesar de disminuir o aumentar los intervalos los errores que se cometen siguen siendo elevados (Figura 3) y el error teórico no coincide en ningún momento con el error empírico.

Vistos los resultados de la gráfica (Figura 3), se continuará calculando y concretando parámetros para que los errores sean tolerables.

#### Estudio del parámetro threshold

Otro elemento a estudiar es el valor del umbral, el cual se aborda con dos procedimientos diferentes y en ambos casos, se estudian los resultados raster y vectorial. Con GRASS se procesa de forma raster. Debido a la gran cantidad de puntos no se ha podido utilizar SIG GRASS para el análisis vectorial, por lo que el estudio vectorial se realiza con software Matlab programando funciones y programas adecuados para esta labor.

El primer procedimiento consiste en realizar una interpolación bicúbica de los puntos observados y posteriormente realizar otro spline con el archivo de puntos libres de outliers procesados con el comando '*v.outlier*' fijando los parámetros de Tikhonov y spline y cambiando el parámetro del umbral. Tanto en el procesamiento vectorial como en el raster en este primer método, con los archivos de puntos se realiza una interpolación para su diferenciación entre original y libre de outliers para los diferentes valores de umbral. A partir de las diferencias se calculan los errores cuantificando el número de puntos o celdas raster fuera de la curva de la distribución normal fijando un estándar de 3 sigma o error del 0.27%.

El segundo procedimiento consiste en hacer la diferencia entre los puntos observados libres de outlier con el spline correspondiente a dicho archivo de puntos sin outlier. Esto mismo se realiza con el archivo que contiene todos los puntos para poder contrastar el resto de diferencias con una diferencia inicial. Para el proceso raster se procede de la misma

forma que en el primer procedimiento haciendo las diferencias entre los valores observados y los valores limpios de outliers e interpolados para diferentes valores de threshold, y lo mismo se aplica al procesamiento vectorial calculando el número de puntos fuera del 99.73% de la distribución normal.

Tras el estudio de los mejores parámetros convenientes para la zona de trabajo y el análisis tanto visual como estadístico, se generan los MDTs con puntos first y last, siendo estos últimos los que se utilizan normalmente para el cálculo de los MDTs con un tamaño de píxel de 2 metros. Estos son creados a partir de software diferentes como son GRASS, ArcGis y Gstat utilizado para técnicas de kriging. Tras la creación de los modelos se realiza un análisis visual y estadístico de las diferencias entre los diferentes MDTs procesados a lo largo del trabajo y los modelos iniciales procedentes de técnicas fotogramétricas.

## RESULTADOS

A lo largo del trabajo se han obtenido distintos resultados que han condicionado tanto los estudios como las directrices a seguir para conseguir el objetivo final de dicho trabajo. Si hay que destacar algún resultado serían los parámetros finales de los estudios realizados (Tabla 2) así como los MDTs de primer y último pulso.

Tabla 2.- Parámetros finales de estudio.

Valor de regularización de Tikhonov-lambda	0.01
Paso de spline dirección N-S y E-W	20x20 m.
Threshold o umbral	35 m.

Para el estudio inicial se obtiene por cada archivo de puntos un número aproximado de outliers con unos parámetros iniciales y las celdas del MDS consideradas como errores fuera de la distribución normal. Con los resultados obtenidos y las imágenes se observa que las zonas de outliers son siempre las mismas y se consideran zonas conflictivas.

Para los diferentes pasos de spline se calculan los diferentes números de outliers. Cuanto más aumenta el valor del spline más aumenta el número de outliers y las zonas donde estos se sitúan siguen siendo las zonas conflictivas. En este procedimiento también se calculan las celdas fuera del intervalo sigma con diferentes valores para ver si en algún caso el valor del error práctico coincide con el error teórico (Figura 3).

Para el primer procedimiento del estudio del umbral, tanto raster como vectorial, los valores

de los errores teóricos y los errores calculados se intersectan en el intervalo que corresponde a un valor de threshold de entre 30 y 35 metros (Tabla 3). En cambio para el segundo procedimiento los valores de los errores calculados son muy altos comparados con el valor del error teórico al que se quiere llegar (Tabla 4). Al comparar las imágenes de puntos outlier se puede ver que con este segundo procedimiento los puntos no siguen ningún patrón de comportamiento como sucede en el primer procedimiento. Si se analizan los valores de los resultados, el valor calculado de los errores es muy elevado debido a que las superficies difieren más de lo que se esperaba, por lo que se aumenta el valor de threshold. Aún en el caso vectorial el número de puntos sigma sigue siendo elevado y la situación de los puntos sigue siendo. Después de ver las imágenes y el comportamiento de los puntos se puede decir que con este segundo procedimiento se están eliminando puntos pertenecientes al MDS. Si se considera que en las observaciones LIDAR la existencia de puntos outlier es mínima, los resultados de este procedimiento son coherentes y es por eso mismo que en el primer procedimiento se obtienen valores aceptables para los parámetros estudiados con resultados correctos.

Con respecto a la comparación de los MDTs. Al comprar los MDTs no procesados en este trabajo con los MDTs calculados se encuentran semejanzas ya que sus mayores diferencias se producen en las zonas conflictivas sobre la superficie urbana. Cabe destacar que al comparar el MDT fotogramétrico con los resultantes del estudio, a pesar de que el rango de valores de las diferencias es pequeño, se sigue observando que las zonas conflictivas siguen siendo las mismas y que el resto de las diferencias más grandes se sitúan en estas zonas puntuales a pesar que se destaquen más en la comparación hecha con del MDT fotogramétrico de 10 metros sin contar los efectos de borde.

## CONCLUSIONES

La zona que se ha estudiado es una superficie llana sin ningún cambio brusco de altura. Por otra parte, se encuentra una zona urbana con elementos de gran altura. Si la superficie no hubiera presentado zona urbana, se podría haber utilizado spline más elevado que el que se ha utilizado, así como valores de umbral más bajos debido a la ausencia de grandes cambios de altura. Este mismo caso se repetiría en zonas forestales. Dependiendo de la vegetación se utilizarían unos valores de umbral altos para vegetación de poca altura y un valor pequeño para vegetación de mayor altura.

Con respecto al filtrado con GRASS, se puede considerar que los algoritmos funcionan correctamente. No obstante, en el terreno se pueden observar zonas conflictivas ya que al automatizar los procesos de filtrado no se tienen en consideración algunas de las características de estas zonas, produciendo errores que alteran las estadísticas. Es por esto, que cuando se busca la automatización del trabajo de filtrado y de procesado de puntos es importante el criterio de un operador para controlar dicho proceso de filtrado y controlar este tipo de situaciones, ya que la automatización busca unos parámetros acordes con todos los puntos que intervienen en el proceso en las diferentes zonas de estudio.

A efectos prácticos, para la realización de un MDT por procesos fotogramétricos en comparación con procedimientos LIDAR, el usuario interviene en una mayor medida obteniendo unos resultados muy similares. No hay duda de que en un futuro los algoritmos de filtrado de puntos LIDAR serán más rigurosos y serán capaces de abordar estas zonas sin la intervención del usuario y con una mejor y mayor precisión que ya se ha observado en comparación con los métodos fotogramétricos todavía no desfasados.

**Tabla 3.-** Errores primer procedimiento threshold.

Threshold	pts outlier	cells 3sigma	total cells	% calculado
20	1919	7485	1000000	0,75
25	1047	5551	1000000	0,56
30	226	4402	1000000	0,44
35	27	2547	1000000	0,25

**Tabla 4.-** Errores segundo procedimiento threshold.

Threshold	pts outlier	cells 3sigma	total cells	% calculado
20	1919	12647	999461	1,2654
25	1047	12648	999461	1,2655
30	226	12662	999461	1,2669
35	27	12641	999461	1,2648

## BIBLIOGRAFÍA

Antolin R and Brovelli M.A.2006. LIDAR data filtering with GRASS GIS for the determination of digital terrain models. *Servei de sistemes d'informació geogràfica i teledetecció, Universitat de Girona. I Jornades de Sig Libre.*

Brovelli M.A and Cannata. M 2002 Digital Terrain model reconstruction in urban areas from airborne laser scanning data: the method and the example of the town of Pavia (Northern Italy). *International Archives of Photogrammetry and Remote Sensing.* Xian, China, XXXIV: 43-48.

Wehr A. and Lohr U.1999. Airborne Laser Scanning- an introduction and overview. *ISPRS Journal of photogrammetry & Remote Sensing* 54. 68-82.