

Metodología para procesar datos LiDAR con ArcGIS Desktop 10

Antonio San José Albacete¹, Francisco Mauro Gutiérrez²

¹Universidad Complutense de Madrid, ²Universidad Politécnica de Madrid

sanjosealbacete@yahoo.es

Resumen

El presente trabajo pretende mostrar el procesamiento de los datos LiDAR mediante el uso de ArcGIS Desktop 10.

Para ello se realiza una visión del tratamiento de los datos en ArcGIS Desktop 10: acceso a la información, visualización y herramientas disponibles. En este punto se ha desarrollado un modelo de geoprocésamiento que permite realizar una primera depuración automática de los datos y crear dos productos básicos: el Modelo Digital de Superficies y el Modelo Digital de Elevaciones, así como la normalización de los datos.

Finalmente, y a partir de los archivos .LAS facilitados por el Grupo de Investigación, Tecnologías y Métodos para la Gestión Sostenible, se ha elaborado un procedimiento y un modelo de geoprocésamiento para procesar los datos y permitir su posterior uso en aplicaciones forestales: cálculo del volumen forestal, estimación de biomasa forestal u otras variables de interés.

Palabras clave: LiDAR, ASPRS (American Society for Photogrammetry and Remote Sensing), ArcGIS Desktop 10, archivos .las, shapefile, MDT, MDE, MDS, normalización, biomasa.

1 Introducción

1.1 Planteamiento y justificación del trabajo

El planteamiento inicial de este trabajo se centra en la necesidad de procesar los datos LiDAR con el software ArcGIS Desktop 10. El tratamiento de datos LiDAR con este software suscita un problema fundamental: la visibilidad de los campos de información. Debido a su codificación binaria, ArcGIS es incapaz de mostrar todos los campos con la información inicial de los archivos .LAS. Esta información es esencial para el tratamiento y la elaboración de productos derivados de los datos LiDAR.

La necesidad del Grupo de Investigación, Tecnologías y Métodos para la Gestión Sostenible es simplificar los procesos de tratamiento de datos LiDAR y facilitar el acceso a su información usando un único software. Esto permitiría a los operadores centrarse en los análisis forestales, evitando el manejo de diferentes manuales y tipos de software para extraer la información necesaria.

1.2 Hipótesis y objetivos

Los objetivos del presente trabajo se centran en el uso de ArcGIS 10 para transformar los datos LiDAR en archivos shapefile, mostrando todos los campos de información y permitiendo su uso sin restricciones.

El trabajo se centra en desarrollar modelos de procesado y de normalización para estandarizar y simplificar estas operaciones.

Hay que tener en cuenta la enorme difusión que tiene el uso del software ArcGIS entre los profesionales dedicados a estudios y aplicaciones forestales. Estos usuarios forestales se encuentran con numerosas dificultades a la hora de realizar estos estudios, debido a que gran parte de los programas de tratamiento de datos LiDAR son poco amigables y presentan una escasa interfaz gráfica.

A continuación se plantea la necesidad de desarrollar un modelo de geoprocusamiento orientado a las aplicaciones forestales del Grupo de Investigación, Tecnologías y Métodos para la Gestión Sostenible. Este método se desarrolla para estimar variables de interés forestal como la biomasa, el volumen maderable u otras en un área concreta, a partir de la comparación de los datos LiDAR con los datos obtenidos en las mediciones y trabajos de campo.

2 Desarrollo y evolución de los datos LiDAR

El sensor LiDAR (Light Detection And Ranging) es un sensor activo, él mismo genera la energía (el pulso láser) que permite hacer los cálculos y las medidas. Envía una señal a la superficie terrestre desde un sensor aerotransportado, que

regresa al transporte aéreo; el tiempo que tarda en llegar hasta la superficie terrestre y regresar al transporte aéreo determina la distancia hasta los elementos de la superficie. El resultado es una densa red de puntos de elevación georreferenciados de elevada precisión, denominada "nube de puntos".

La habilidad del sensor LiDAR para capturar información bajo la cubierta vegetal es una de sus principales ventajas. La densidad de la nube de puntos es suficiente para crear un Modelo Digital de Superficies muy preciso.



Figura 1. Esquema de retorno de los pulsos láser.

El proceso de adquisición de datos LiDAR consiste básicamente en la sincronización de los datos del sensor láser con la información de navegación GPS (x, y, z) e INS (omega, phi y kappa). Es necesario realizar un estudio de planeamiento previo antes de iniciar el proceso de obtención de datos.

Los datos LiDAR se almacenan en formato binario. Los datos en bruto se agrupan en una sola clase que recoge todas las respuestas de cada disparo láser.

El formato .LAS es el más utilizado; contiene datos binarios que consisten en archivo cabecera, registros de longitud variable y datos de puntos. Este formato presenta a su vez diferentes formatos que han evolucionado con el tiempo para incluir más información [1].

3 Desarrollo de la investigación

3.1 Tratamiento de datos LiDAR en ArcGIS 10

En la mayor parte de los análisis con datos LiDAR no se usan todos los atributos de forma simultánea ya que, en función del tipo de análisis o estudio que se quiera realizar, este se apoyará en unos atributos determinados.

De este modo, los retornos son necesarios para análisis forestales y urbanos, aunque hay opciones que permiten realizar esos análisis con los atributos de intensidad. Las clasificaciones son fundamentales, aunque es posible realizar otro tipo de clasificaciones en función de los retornos. Lo cierto es que cuanto mayor y más densa sea la información que nos proporcionan los datos LiDAR, mejores y

más aproximados a la realidad serán los productos derivados de los posteriores análisis.

Sin embargo hay una serie de atributos que son comunes y necesarios para todos los análisis. Se trata de la posición y la altura (x, y, z). El resto de atributos de los datos LiDAR son: intensidad, retorno, número de retornos, clasificación, ángulo de escaneado, identificador, datos de usuario, dirección de escaneado, dirección de vuelo e información GPS.

El programa ArcGIS 10 no permite manipular archivos .LAS directamente ya que almacena los atributos en formato binario: BLOB (conjunto de datos binario almacenados como una única entidad en una base de datos que permite su compresión).

El elevado número de retornos que dan lugar a la nube de puntos hace que ArcGIS sea incapaz de cargar todos los atributos y administrarlos. Se almacenan como entidades multipunto: clúster que agrupan gran cantidad de puntos individuales en un solo registro de la tabla de atributos.

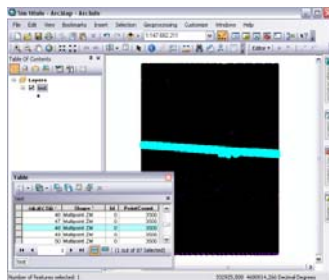


Figura 2. Visualización de entidades de multipunto.

3.2 Modelo de geoprocesamiento de datos LiDAR

Los datos LiDAR requieren superar una serie de filtros antes de poder trabajar con ellos en cualquier tipo de aplicación. Por lo general, los datos LiDAR sufren una serie de procesos de depuración antes de ser almacenados en formato .LAS. Este primer tamizado, limpia todos los datos excéntricos que son consecuencia de errores, tanto del sensor como de los datos erróneos.

Un segundo filtro de los datos, antes de almacenarse en formato .LAS, es el que permite separar el suelo de la vegetación y de los objetos sobre la superficie terrestre. Este permite rellenar el atributo de la clasificación para posteriores procesos de selección.

Normalmente los datos empleados para realizar análisis ya han pasado estos procesos, por lo que el presente trabajo se centra en el procesamiento de los archivos .LAS clasificados.

Los modelos se centran en el flujo de trabajo orientado al almacenamiento y la depuración de los datos LiDAR para su posterior aplicación en estudios forestales. En primer lugar se presenta el modelo de almacenamiento de datos LiDAR y, a continuación se muestra el modelo de carga de datos LiDAR en ArcGIS 10.

3.2.1 Modelo de creación del espacio de trabajo

Hay que tener en cuenta que es necesario almacenar todos los datos en una geodatabase. Esto permite manejar los datos LiDAR y tratar la información de una manera más ágil y compacta, recogiendo todos los shapefiles dentro de un conjunto de datos de entidades (Feature Dataset).

La elección del tipo de geodatabase depende de las características del estudio. Debido a que las muestras de datos LiDAR suponen gran cantidad de puntos almacenados en archivos .LAS, es recomendable elegir una geodatabase de archivo (File Geodatabase), que permite almacenar desde 1 TB hasta 256 TB.

3.2.2 Modelo de carga de datos LiDAR en ArcGIS 10

El modelo de carga de datos LiDAR permite transformar los archivos .LAS en shapefile para visualizar la nube de puntos en ArcGIS. El modelo mostrado se basa en herramientas de ArcGIS apoyadas en el script ExplodeLASMultipoint [3].

3.2.3 Modelo de carga de datos con herramientas de ArcGIS y script de ESRI Support

Este modelo permite obtener de manera rápida y eficaz todos los atributos de los archivos .LAS, incluyendo el valor de elevación, en un solo paso mediante un script. Para ello es necesario tener un conocimiento previo de los campos que contienen los archivos .LAS para poder filtrarlos en ArcGIS. La mejor solución es usar un editor gratuito de datos LiDAR y comprobar los campos de sus atributos.

El flujo de trabajo es el siguiente:

- Comprobar los atributos del archivo .LAS en un editor (por ejemplo, LASEdit).
- Ejecutar la herramienta LAS To Multipoint desde ArcGIS, incluyendo aquellos atributos presentes en el archivo .LAS que sean necesarios.
- Ejecutar el script ExplodeLASMultipoint desde el editor de Visual Basic de ArcMap.

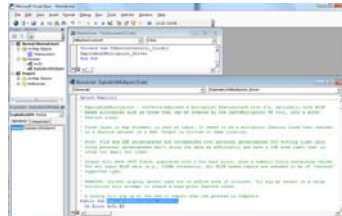


Figura 3. Código fuente del botón ExplodeLASMultipoint en el editor de Visual Basic.

Este proceso da como resultado un archivo shapefile de puntos con los valores de elevación y los atributos seleccionados. Estos atributos no son visibles en la tabla de atributos de la entidad de multipuntos debido a que están almacenados en formato BLOB. Sin embargo, el script se encarga de transformarlos y asignarlos a cada uno de los puntos de forma individual.

ID	OID*	Shape*	Intensity	ScanAngle	Return	NumReturns	GeoTime	Class	PointCount
1	1	Multipoint Z	Blob	Blob	Blob	Blob	Blob	Blob	1691
2	2	Multipoint Z	Blob	Blob	Blob	Blob	Blob	Blob	1891
3	3	Multipoint Z	Blob	Blob	Blob	Blob	Blob	Blob	1877
4	4	Multipoint Z	Blob	Blob	Blob	Blob	Blob	Blob	1814

ID	OID*	Shape*	Intensity	ScanAngle	Return	NumReturns	GeoTime	Class
1	1	Point	100.11	60	14	1	20120120000000	4
2	1	Point	100.11	60	14	1	20120120000000	4
3	1	Point	100.11	60	14	1	20120120000000	4
4	1	Point	100.11	60	14	1	20120120000000	4

Figura 4. Comparación de las tablas de entidades de multipuntos y puntos individuales.

Mediante las opciones de personalización de ArcGIS 10 es posible configurar un botón para realizar las mismas operaciones del script ExplodeLASMultipoint [4].

3.3 Aplicaciones forestales con datos LiDAR

El Grupo de Investigación, Tecnologías y Métodos para la Gestión Sostenible se dedica a estudios forestales mediante datos LiDAR. Uno de sus proyectos está orientado a medir el volumen maderable y la biomasa de determinadas zonas forestales mediante la comparación de datos de campo y datos LiDAR.

Los datos LiDAR con los que trabajan llegan pre-procesados, es decir, reciben un primer filtro para eliminar los datos erróneos y se rellenan los atributos de la clasificación separando el suelo de la vegetación (pero sin tocar los atributos de elevación).

Una vez recibidos los datos LiDAR, es necesario filtrar los datos de suelo y los datos de vegetación a través de su atributo de elevación, en el caso de que los datos no estén clasificados. Este proceso se denomina Ground Filter dentro del software FUSION, aunque también se conoce como proceso de normalización en el ámbito de las aplicaciones forestales de datos LiDAR. Permite extraer los puntos que corresponden al suelo y generar un archivo .LAS exclusivamente con los datos a nivel de suelo.

Para este proceso es fundamental entender la diferencia entre altura y altitud:

- La altitud se refiere a la elevación del conjunto de puntos sobre el nivel del mar, tanto del suelo como de los objetos sobre la superficie terrestre.
- La altura se refiere a la elevación de los objetos sobre el nivel del suelo, por lo que los datos de suelo estarán en valores entorno a cero.

El flujo de trabajo es el siguiente:

- Filtrado de datos (pre-procesado).

- Separación entre puntos clasificados como suelo y puntos clasificados como vegetación u otros (pre-procesado).
- Creación del MDT con los puntos clasificados como suelo (procesado).
- Extracción de alturas por diferencia entre puntos de suelo y puntos de vegetación (procesado).

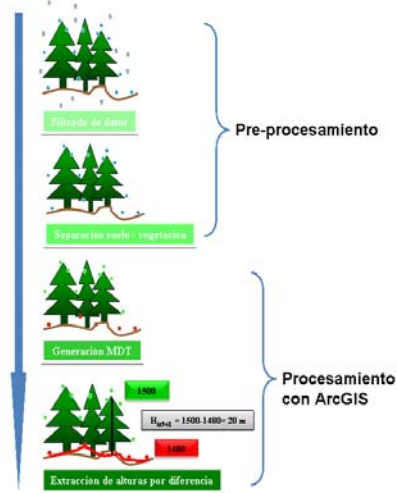


Figura 5. Esquema del proceso de normalización.

3.3.1 Modelo de normalización

El modelo de geoprocésamiento para normalizar los datos LiDAR pasa por el modelo de carga de datos. Una vez cargados los datos en una geodatabase como una entidad de puntos individuales con todos sus atributos es posible realizar el proceso de normalización. El flujo de trabajo es el siguiente:

- Crear una capa de la entidad de puntos.
- Seleccionar los puntos de la capa cuya clase sea suelo (los valores son 2 y 7, según la clasificación de la ASPRS).
- Crear un MDT o un TIN mediante la herramienta de interpolación (IDW).
- Crear una nueva entidad de puntos con los valores de elevación de cada punto y los valores de elevación calculados a nivel de suelo en el MDT.
- Calcular en un nuevo campo los valores de elevación normalizados (altura sobre el nivel del suelo), restando los valores del campo raster value a los valores de elevación sobre el nivel del mar del conjunto de puntos.

De este modo, y partiendo de una entidad de puntos con su atributo de elevación referido al nivel del mar, obtendremos un nuevo campo que nos permitirá trabajar con elevaciones sobre el nivel del suelo.

Una vez aplicado este modelo de normalización a los datos LiDAR es posible realizar análisis concretos referidos a determinadas entidades sobre la superficie terrestre. En este caso se trata de analizar la biomasa de la cubierta vegetal.

3.4 Aplicación forestal

La estimación de variables forestales a partir de datos LiDAR suele realizarse ajustando modelos de regresión que relacionen variables derivadas de los datos LiDAR con variables de interés forestal. Para realizar el ajuste de estos modelos se emplea un conjunto reducido de parcelas en las cuales se conocen tanto las variables forestales (variables a predecir) como las variables LiDAR (variables predictoras). Las variables LiDAR son variables derivadas de la nube de puntos normalizada y se conocen para cualquier punto cubierto en el vuelo.

Una vez ajustado un modelo de regresión, éste se aplica a puntos donde solo se conocen las variables LiDAR (no se conocen las variables forestales a estimar) y se obtienen las estimaciones de las variables forestales [5].

Las variables predictoras derivadas del LiDAR son, generalmente, variables estadísticas derivadas de la distribución de alturas de los puntos LiDAR en una determinada superficie (la altura máxima, mínima, media, la desviación estándar, los percentiles, los cuantiles). La superficie de referencia es la superficie de las parcelas medidas en campo que se emplean para ajustar los modelos de regresión. Generalmente este proceso se realiza con el software FUSION.

Una vez que se conocen las variables a predecir (mediciones de campo) y las variables predictoras (variables calculadas en el paso anterior), se introducen ambas en un software estadístico que permite ajustar los modelos [6].

Con el paquete estadístico se realizarán las siguientes operaciones:

- Selección de variables independientes o auxiliares óptimas y Ajuste del modelo de regresión.
- Validación del modelo de regresión y análisis de los errores para las predicciones realizadas con el modelo.

3.5 Aplicaciones a un caso de estudio

El área de estudio seleccionada es una zona de estudio del pinar de Valsaín, en el municipio de San Ildefonso o La Granja, en la provincia de Segovia. Se trata de un área forestal de 303.356 Ha (3.03356 km²).

3.5.1 Carga de datos

Se ha optado por crear una geodatabase de archivo (.gdb), con mayor capacidad de almacenamiento (de 1TB a 256TB). El modelo comprende la creación de un conjunto de entidades (Feature Dataset) con un sistema de coordenadas geográfico WGS 1984 UTM Zona 30 Norte.

En primer lugar extrae los datos LiDAR de las parcelas, calcula las variables estadísticas del atributo de elevación y añade un campo para calcular el modelo de predicción.

A continuación carga los datos de biomasa de las parcelas calculados en el trabajo de campo, los une a la tabla de las parcelas y calcula el modelo de predicción de biomasa.



Figura 7. Modelo de análisis forestal aplicado a las parcelas.

El análisis forestal del área de estudio extrapola los valores obtenidos con el modelo anterior aplicados al área de estudio.

En primer lugar crea una malla con tamaño de celda igual al tamaño de las parcelas. En este caso se trata de una malla de 2.303 celdas de 35,45 x 35,45 metros (1256,7 m²) cada una. Estas celdas se cruzan con la nube de puntos mediante un recorte de la entidad de puntos.

A continuación se calculan las variables estadísticas del atributo de elevación para cada celda de la malla.

Finalmente se añaden dos nuevos campos en los que se aplican los modelos de predicción, tanto el de volumen forestal como el de biomasa, realizados por el Grupo de Investigación, Tecnologías y Métodos para la Gestión Sostenible mediante el software Statgraphics.



Figura 8. Modelo de análisis forestal aplicado al área de estudio.

En este caso, los modelos de predicción aplicados son los siguientes:

- $\text{Volumen (m}^3/\text{ha)} = -286.1671564561 + 60.5713592113 * \text{MEAN_ALTUR}$

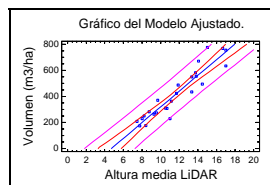


Figura 9. Gráfico de regresión del modelo de predicción de volumen.

- $\text{Biomasa (Kg/ha)} = 16646,04250884 + 3014,86872173 * \text{MEAN_ALTUR} + -1007,84640200 * \text{MAX_ALTURA}$.

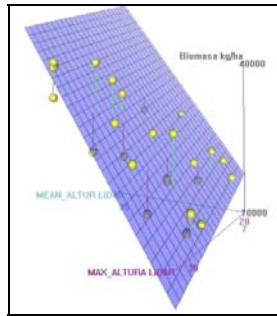


Figura 10. Gráfico de regresión del modelo de predicción de biomasa.

3.5.4 Mapas de volumen forestal y estimación de biomasa

El resultado final de todo este proceso de análisis forestal es el mapa de volumen forestal y el mapa de estimación de biomasa. Estos se representan a través de los atributos volumen y de biomasa calculados con el modelo de predicción en el último análisis.

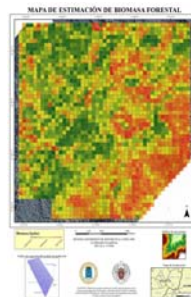


Figura 11. Mapa de estimación de biomasa forestal.

4 Conclusiones

4.1 Evaluación de los objetivos

El desarrollo de la investigación ha cubierto las necesidades planteadas y se ha elaborado un proceso y una metodología que permiten realizar análisis forestales con datos LiDAR en el entorno de ArcGIS 10.

Este trabajo sirve de precedente y modelo de desarrollo para futuras aplicaciones orientadas al tratamiento de datos LiDAR con ArcGIS 10, así como

una guía orientativa en la introducción al tratamiento de datos LiDAR. Del mismo modo, es muy útil su aplicación didáctica para comprender el flujo de trabajo y explicar la estimación de variables forestales con datos LiDAR con ArcGIS 10 [7].

4.2 Síntesis del modelo propuesto y aportaciones originales

La metodología empleada se basa en el uso de Model Builder para generar procesos automáticos que hagan del tratamiento de datos LiDAR una tarea lo más ágil y simple posible.

La organización del modelo de desarrollo propuesto comprende los siguientes puntos concretos:

- Creación de un espacio de trabajo.
- Carga de datos LiDAR.
- Normalización de los datos.
- Preparación de datos para el ajuste de modelos de regresión.
- Preparación de datos para la aplicación del modelo de regresión y aplicación del modelo en sí.

5 Referencias bibliográficas

- [1] The American Society for Photogrammetry and Remote Sensing (ASPRS), 2010, LAS SPECIFICATIONS VERSION 1.3.
- [2] McGaughey, R.J., 2010, FUSION/LDV: software for LiDAR Data Analysis and Visualization, FUSION v.2.90.
- [3] Crawford, C., 2010, LiDAR Solutions in ArcGIS, ESRI International User Conference.
- [4] Chang, K-T., 2007, Programming ArcObjects with VBA. A Task-Oriented Approach, CRC Press, Taylor and Francis Group. ISBN 13: 978-1-58488-580-1.
- [5] Palomino Palomino, P., 2010, Algoritmo para la localización y estimación de masa forestal a partir de imágenes LiDAR, Trabajo Fin de Máster de Investigación en Sistemas Inteligentes, Facultad de Informática, UCM. <http://eprints.ucm.es/9912>
- [6] Peña Sánchez de Rivera, D., 2002, Regresión y diseño de experimentos, Alianza Editorial, Madrid. ISBN: 84-206-8695-6.
- [7] San José Albacete, A., 2011, Procesamiento de datos LiDAR con ArcGIS Desktop 10, Trabajo Fin de Máster de Tecnologías de Información Geográfica, Facultad de Geografía e Historia, UCM. <http://eprints.ucm.es/13981/>