

**CENTRO DE INVESTIGACIÓN EN GEOGRAFÍA Y GEOMÁTICA**

**“ING. JORGE L. TAMAYO”, A. C.**

**CentroGeo**

**Centro Público de Investigación CONACYT**

**ESTIMACIÓN DE ALMACENAMIENTO DE CARBONO  
EN EL SUELO DE CONSERVACIÓN DEL DISTRITO FEDERAL  
MEDIANTE EL USO DE DATOS LIDAR AEROTRANSPORTADO**

**TESIS**

Que para obtener el grado de Maestro en Geomática

Presenta

**Juan Andrés Almazán González**

Supervisor Principal:

**Dr. José Luis Silván Cárdenas**

Comité Supervisor:

**Dr. Gustavo Manuel Cruz Bello**

Examinador Externo:

**Dr. Stéphane Robert André Couturier**

México, D. F., 24 de abril de 2013

© CentroGeo. Derechos reservados. El autor otorga a CentroGeo el permiso de reproducir y distribuir copias de esta tesis en su totalidad o en partes.

## RESUMEN

La administración de áreas forestales es una labor utilizada en la mitigación de gases de efecto invernadero, pues la vegetación almacena carbono, que es uno de estos gases, reduciendo su presencia en la atmósfera. En la ejecución de esta administración, uno de los primeros pasos es tener una estimación de la cantidad de carbono presente en estas áreas.

En el presente trabajo se utilizan datos de lidar aerotransportado para la estimación de carbono en el Suelo de Conservación del Distrito Federal, estimando también otras variables forestales: número, altura media y diámetro normal promedio de los árboles, por área. Se emplean dos métodos: la segmentación de árboles individuales mediante un algoritmo de identificación por máximos locales y agrupación k-means, para estimar el número de árboles y altura media; y la construcción de modelos de predicción utilizando la regresión stepwise, a partir de variables obtenidas de la distribución y características de la nube de puntos, para estimar carbono, altura media, diámetro normal promedio y número de árboles. El objetivo es conocer la exactitud de estos métodos para determinar la viabilidad de incorporarlos a una metodología de estimación de variables forestales.

La exactitud de la identificación de árboles en la nube de puntos fue inferior a la de otros trabajos de detección, y no se considera admisible para llevar a cabo estimaciones. Por otro lado, los modelos de predicción de carbono para pino y altura media para oyamel entregaron resultados que los hacen aceptables para estimar estas variables.

*A mi madre, mi hermana y mi abuela.*

*A las que amo con todo mi ser.*

**Reconocimientos:**

A los Doctores José L. Silván, Gustavo M. Cruz y Stéphane Couturier, y al Biólogo Armando Peralta, cuya disposición y comentarios fueron un pilar de apoyo en la construcción del presente trabajo.

## TABLA DE CONTENIDO

<b>RESUMEN.....</b>	<b>ii</b>
<b>DEDICATORIA.....</b>	<b>iii</b>
<b>RECONOCIMIENTOS.....</b>	<b>iv</b>
<b>TABLA DE CONTENIDO.....</b>	<b>v</b>
<b>LISTA DE TABLAS.....</b>	<b>vii</b>
<b>LISTA DE FIGURAS.....</b>	<b>viii</b>
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>2. ESTIMACIÓN DE VARIABLES FORESTALES.....</b>	<b>5</b>
2.1 INTRODUCCIÓN.....	5
2.2 ESTIMACIÓN DE VARIABLES FORESTALES MEDIANTE LIDAR AEROTRANSPORTADO.....	7
2.2.1 Estimación de variables forestales en general.....	7
2.2.2 Estimación de biomasa.....	11
2.2.3 Segmentación de copas.....	12
2.2.4 Identificación de especies.....	14
2.3 EL SISTEMA LIDAR AEROTRANSPORTADO.....	15
2.3.1 Medición de distancias por tiempo de vuelo del pulso.....	18
2.3.2 Medición de distancias por diferencia de fase.....	20
2.3.3 Partes que componen el sistema lidar aerotransportado.....	23
2.3.3.1 Unidad básica de medición laser.....	24
2.3.3.2 Unidad de distribución de los pulsos.....	25
2.3.3.3 Unidad de proceso y control (hardware/software).....	28
2.3.3.4 Unidad de posicionamiento y orientación.....	29
2.3.3.5 Unidad de obtención de imágenes.....	29
2.3.3.6 Control terrestre.....	30
2.3.3.7 Unidad de sincronización y registro.....	30
2.3.4 Corrección del error de altura sistemático de la nube de puntos.....	32
2.3.5 Divergencia del rayo y huella.....	32
2.3.6 Reflectividad.....	33
2.3.7 Procesamiento de los datos.....	34
2.3.7.1 Visualización.....	34
2.3.7.2 Transformaciones.....	34

2.3.7.3 Segmentación y filtrado.....	35
2.3.7.4 Gridding.....	35
2.3.8 Aplicaciones.....	36
2.3.8.1 Aplicaciones en la administración de recursos naturales y desastres.....	37
2.3.8.2 Aplicaciones en áreas urbanas.....	38
2.3.8.3 Aplicaciones en ingeniería civil no urbana.....	38
2.3.9 Ventajas y desventajas de utilizar lidar.....	38
<b>3. METODOLOGÍA .....</b>	<b>40</b>
3.1 INTRODUCCIÓN.....	40
3.2 DESCRIPCIÓN DEL ÁREA DE ESTUDIO.....	41
3.3 DATOS.....	42
3.3.1 Datos de campo.....	42
3.3.2 Datos lidar.....	44
3.4 OBTENCIÓN DE ALTURAS SOBRE EL TERRENO.....	45
3.5 IDENTIFICACIÓN Y SEGMENTACIÓN DE ÁRBOLES INDIVIDUALES.....	50
3.6 ESTIMACIÓN DE CARBONO POR DISTRIBUCIÓN DE ALTURAS DEL DOSEL.....	54
3.7 EVALUACIÓN Y VALIDACIÓN DE LAS ESTIMACIONES.....	62
3.7.1 Identificación de árboles individuales.....	62
3.7.2 Estimación de carbono por distribución de alturas del dosel.....	64
<b>4. RESULTADOS Y DISCUSIÓN.....</b>	<b>66</b>
4.1 IDENTIFICACIÓN DE ÁRBOLES INDIVIDUALES.....	66
4.2 ESTIMACIÓN DE CARBONO POR DISTRIBUCIÓN DE ALTURAS DEL DOSEL.....	78
<b>5. CONCLUSIONES GENERALES.....</b>	<b>90</b>
5.1 IDENTIFICACIÓN DE ÁRBOLES INDIVIDUALES.....	90
5.2 ESTIMACIÓN DE CARBONO POR DISTRIBUCIÓN DE ALTURAS DEL DOSEL.....	91
<b>6. BIBLIOGRAFÍA.....</b>	<b>93</b>

## LISTA DE TABLAS

Tabla 3.1. Parámetros forestales que caracteriza cada variable predictora utilizada.....	57
Tabla 3.2. Distribución por tipo de vegetación de los sitios utilizados en los segmentos de entrenamiento y prueba.....	61
Tabla 4.1. Resultados de la identificación de árboles, por tipo de vegetación.....	66
Tabla 4.2. Estudios forestales que incluyen la identificación de árboles individuales, ordenados en función de la exactitud de la detección.....	77
Tabla 4.3. Indicadores de la calidad de los modelos de predicción para todos los tipos de vegetación.....	85
Tabla 4.4. Indicadores de la calidad de los modelos de predicción de carbono para los tipos de vegetación oyamel y pino.....	86
Tabla 4.5. Estudios de estimación de biomasa basados en la distribución de las alturas del dosel, ordenados de acuerdo al valor más alto de su coeficiente de determinación.....	88

## LISTA DE FIGURAS

Figura 2.1. Diagrama de la clasificación de los sistemas lidar de acuerdo a diferentes criterios.....	18
Figura 2.2. Ilustración del método de medición por tiempo de vuelo.....	19
Figura 2.3. Ilustración del método de medición por diferencia de fase.....	21
Figura 2.4. Ilustración de la onda de medición y la onda portadora del laser.....	22
Figura 2.5. Diagrama de los componentes principales de un sistema lidar aerotransportado.....	24
Figura 2.6. Ilustración del patrón de escaneo dentado.....	26
Figura 2.7. Ilustración del patrón de escaneo de líneas paralelas transversales a la dirección de vuelo.....	27
Figura 2.8. Ilustración del patrón de escaneo elíptico.....	27
Figura 2.9. Ilustración del patrón de escaneo en líneas paralelas en la misma dirección de vuelo.....	28
Figura 2.10. Ilustración de la suma de vectores necesaria para lograr el registro de una nube de puntos.....	31
Figura 3.1. Ilustración del diseño de los conglomerados utilizados para el muestreo.....	44
Figura 3.2. Imagen de la nube de puntos del conglomerado 575, todos los puntos.....	47
Figura 3.3. Imagen de la nube de puntos del terreno del conglomerado 575.....	48
Figura 3.4. Imagen de la nube de puntos de vegetación del conglomerado 575.....	49
Figura 3.5. Gráfica de dispersión de altura promedio contra diámetro de copa promedio de todos los sitios.....	52
Figura 3.6. Gráfica de dispersión del diámetro de copa promedio calculado con la ecuación del trabajo de Malleux (1970) contra sus diferencias con respecto al diámetro de copa promedio calculado con la ecuación de Gill et al. (2000), para todos los sitios.....	53
Figura 4.1. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación pastizal.....	68
Figura 4.2. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación oyamel.....	69
Figura 4.3. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación pino.....	69
Figura 4.4. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación bosque inducido.....	70



Figura 4.5. Gráfica de dispersión de árboles censados contra porcentaje de detección en los sitios con tipo de vegetación oyamel.....	70
Figura 4.6. Gráfica de dispersión de árboles censados contra porcentaje de detección en los sitios con tipo de vegetación pino.....	71
Figura 4.7. Gráfica de dispersión de árboles censados contra porcentaje de detección en los sitios con tipo de vegetación bosque inducido.....	71
Figura 4.8. Gráfica de dispersión de altura promedio censada contra porcentaje de detección en los sitios con tipo de vegetación pino.....	72
Figura 4.9. Gráfica de dispersión de porcentaje de detección contra el error en la estimación de la altura media en los sitios con tipo de vegetación oyamel.....	73
Figura 4.10. Gráfica de dispersión de porcentaje de detección contra el error en la estimación de la altura media en los sitios con tipo de vegetación pino.....	73
Figura 4.11. Gráfica de dispersión de diámetro normal promedio censado contra porcentaje de detección en los sitios con tipo de vegetación pino.....	74
Figura 4.12. Gráfica de dispersión de la variable carbono contra la variable predictora H1.....	81
Figura 4.13. Gráfica de dispersión de la variable carbono contra la variable predictora H90.....	81
Figura 4.14. Gráfica de dispersión de la variable carbono contra la variable predictora SIDrel.....	82
Figura 4.15. Gráfica de dispersión de la variable carbono contra la variable predictora SIDrelcorreg.....	82
Figura 4.16. Gráfica de dispersión de la variable carbono contra la variable predictora Ps1.....	83

## 1. INTRODUCCIÓN

El cambio climático es un tópico que ha cobrado mucha importancia en los últimos años a nivel mundial, debido a que es uno de los componentes que alteran la estructura y el funcionamiento de la Tierra como sistema y tiene efectos directos sobre todas las especies de seres vivos del planeta (Vitousek, 1994: 1861-1862).

Existen ecosistemas naturales que pueden presentar afectaciones importantes debido al cambio climático, como la tundra, bosques boreales, manglares, marismas, arrecifes de coral, etc., y existen también sectores relacionados con las actividades humanas, incluso en países desarrollados, que ya han demostrado ser vulnerables a eventos como olas de calor, ciclones, inundaciones y sequías, que aumentarían su frecuencia e intensidad en la medida en la que el calentamiento global aumentara (IPCC 2007:72). Por otro lado, la vida económica y social de la humanidad está basada en los beneficios obtenidos de los ecosistemas naturales y limitada por sus recursos; ya que el cambio climático ha tenido una influencia en estos últimos, también ha afectado a tales procesos económicos y sociales (Simard y Austin, 2010:139).

El Panel Intergubernamental en Cambio Climático (IPCC, del inglés Intergovernmental Panel on Climate Change) fue fundado en 1988 por el Programa del Medio Ambiente de las Naciones Unidas (UNEP, del inglés United Nations Environment Programme) y la Organización Meteorológica Mundial (WMO, del inglés World Meteorological Organization) para proveer al mundo de un punto de vista científico en lo que se refiere al estado actual de conocimiento del cambio climático y su potencial impacto socio-económico y sobre el medio ambiente. El IPCC define el cambio climático como el...

“...cambio en el estado del clima que puede ser identificado (por ejemplo, utilizando pruebas estadísticas) por medio de los cambios en la media y/o la variabilidad de sus propiedades, y que persiste por un periodo de tiempo extenso, típicamente décadas o mayor. Esto se refiere a cualquier cambio del clima en el tiempo, ya sea debido a la variabilidad natural o como resultado de las actividades humanas” (IPCC 2007:30).

Una de las componentes más conocidas del cambio climático es el aumento de la temperatura media global o calentamiento global, el cual es inequívoco, como sostiene la IPCC, pues es evidente en las observaciones del incremento de la temperatura media global del aire y del océano, la fundición de nieves y hielos y el aumento del nivel medio del mar (IPCC 2007:30).

Respecto a la causa del calentamiento global, la IPCC señala que la mayor parte del incremento de la temperatura media global, durante la segunda mitad del siglo XX, es debido al aumento de los gases de efecto invernadero derivados de las actividades humanas (IPCC 2007:39). Uno de estos es el dióxido de carbono, que representa el 3.62 % de la cantidad total de gases de efecto invernadero presentes en la atmósfera (el vapor de agua representa el 95 %). Aunque el dióxido de carbono es minoría, un incremento en su ritmo de emisión puede conducir a una gran acumulación durante un periodo de tiempo largo, pues sus moléculas pueden permanecer en la atmósfera por más de cien años (NCPA, 2007:6-7).

Una alternativa para mitigar el calentamiento global es reducir la cantidad de dióxido de carbono presente en la atmósfera. Prácticas señaladas para la mitigación de este fenómeno son evitar la deforestación, llevar a cabo acciones de reforestación, incrementar la cantidad de carbono en los depósitos de biomasa, reducir el uso directo o indirecto de combustibles fósiles (Cowie et al., 2007:344) y la sustitución y reuso de la madera en diversos productos. Incluso, el IPCC señala como tecnologías a ser aplicadas el mejoramiento de las especies de árboles para aumentar el secuestro de carbono y el desarrollo de técnicas de percepción remota para el análisis del potencial de secuestro de carbono de la vegetación (IPCC 2007:60).

Por esto, la administración de grandes superficies con cobertura vegetal es reconocida como una de las actividades claves que puede ser utilizada para la mitigación de las emisiones de carbono. México se ubica en el lugar número 11 de los países emisores de dióxido de carbono por deforestación y contribuye en un 1.6 % a las emisiones globales (Simard y Austin, 2010:196). Parte de las acciones de esta administración es conocer la cantidad de carbono retenido en estas áreas. Así mismo, es de utilidad conocer la cantidad contenida en una determinada área y la influencia de la especie de los árboles en determinadas circunstancias, que en cierto momento proporcionen una idea del potencial de secuestro de carbono de otras zonas.

Con esa información se pueden diseñar planes de manejo y administración de las zonas boscosas existentes –procurando con más criterio su conservación y evitando en lo posible la deforestación– y de forestación y reforestación, que puedan ser insertados con juicio en los proyectos de mitigación de calentamiento global existentes.

En 2010 la Procuraduría Ambiental y de Ordenamiento Territorial del Distrito Federal (PAOT) en conjunción con otras instituciones desarrolló el trabajo *Estimación de captura de carbono como*

*indicador del estatus del derecho de los habitantes del Distrito Federal a gozar de áreas verdes urbanas adecuadas para su desarrollo, salud y bienestar (PAOT, 2010)*, en el cual se determina la cantidad de carbono contenida en el Suelo de Conservación del Distrito Federal mediante una serie de muestreos en campo. El objetivo de la presente tesis es llevar a cabo la misma estimación –incluyendo también algunos otras variables– pero utilizando datos de lidar aerotransportado.

Actualmente existen dos grandes propuestas: estimación derivada de la identificación de árboles individuales o a partir de estadísticas de la distribución de las alturas del dosel; la idea es probar ambas, con la intención de poder ofrecer elementos que puedan llegar a ser empleados en la confección de una metodología de estimación en forma.

En concreto, los objetivos son los siguientes:

-General:

Evaluar la factibilidad de la estimación de parámetros forestales –número de árboles, diámetro normal promedio, altura media de los árboles y especialmente captura de carbono–, por área, utilizando datos lidar aerotransportado en el Suelo de Conservación del Distrito Federal.

-Específicos:

- a) Evaluar la exactitud de la identificación de árboles individuales utilizando datos lidar aerotransportado, con objeto de ser empleados en estimaciones de captura de carbono.
- b) Evaluar el desempeño de la estimación de parámetros forestales –número de árboles, altura media de los árboles, diámetro normal promedio y captura de carbono–, por área, utilizando el método de estadísticas de la distribución de las alturas del dosel.

La factibilidad se refiere al grado en el que existen los recursos necesarios y suficientes para desempeñar una cierta tarea. En este caso, la intención es determinar si los datos en conjunción con los procesos de transformación y estimación entregan buenos resultados en la determinación de las variables forestales de interés, considerados como buenos si igualan o mejoran los alcanzados por otros trabajos en el ámbito. En el caso de la identificación de árboles, existen estudios que presentan radios de individuos correctamente identificados de casi el 60 % (Morsdorf et al., 2003; e Hirata et al., 2009), y en cuanto a biomasa, se tienen coeficientes de determinación de los modelos de estimación de entre 0.70 y 0.75 (Hall et al., 2005; y Jochem et al., 2010). Es más común encontrar trabajos de estimación de biomasa que de cantidad de carbono y, debido a la similitud entre estas dos magnitudes, se consideran como límites

de buen desempeño de los modelos de estimación de carbono los coeficientes de determinación de algunos trabajos de estimación de biomasa. Para el caso de los modelos de estimación de número de árboles, altura media y diámetro normal medio, también se consideran de buenos resultados si sus coeficientes de determinación son mayores o iguales a 0.70.

Aunque todo el trabajo de estimación con datos lidar se compone de varias etapas: obtención y registro de datos, segmentación suelo–no suelo, obtención de alturas reales, extracción y generación de información de la nube, construcción y aplicación de los modelos de predicción y la presentación de los resultados, el presente trabajo se enfoca sólo en las etapas de la construcción de los modelos de predicción y extracción de información.

La obtención de los datos es una tarea que queda completamente fuera de este trabajo, pues de entrada no se está en la posibilidad de experimentar con diferentes métodos de colecta; la elección de los métodos de segmentación y de obtención de alturas se hacen sin demasiados preliminares, por supuesto, guardando el esfuerzo de que sean útil para los fines; la elección de los métodos de estimación es la que se hace con consideraciones más profundas.

Para la evaluación de la confianza de cada método y para la validación en general se realiza una comparación entre los resultados obtenidos utilizando lidar y los obtenidos por muestreo de campo. Las conclusiones obtenidas permiten entregar apreciaciones acerca de la posibilidad de incorporar estos métodos en la estimación forestal mexicana.

En el capítulo 2 se proporciona una breve reseña de los trabajos forestales hechos a lo largo de los años utilizando lidar, así como los componentes del lidar aerotransportado, la forma en cómo se procesan los datos y sus aplicaciones. El capítulo 3 contiene una corta descripción del área de estudio y la explicación de cómo se llevaron a cabo los métodos de estimación. En el capítulo 4 se muestran los resultados y las observaciones hechas a los mismos, y finalmente en el capítulo 5 se exponen las conclusiones.

## 2. ESTIMACIÓN DE VARIABLES FORESTALES

### 2.1 INTRODUCCIÓN

Un inventario forestal es el proceso de evaluación y cuantificación de uno o varios elementos de los recursos forestales, considerando también las condiciones en las que se desarrollan. Una definición de las más frecuentes es la estimación de la cantidad y calidad de los árboles presentes en un área.

Para llevar a cabo un inventario es necesario determinar una serie de etapas y procesos. En primer lugar debe conocerse la finalidad, por ejemplo, para reconocimiento, determinación de recursos maderables, evaluación de valor recreativo, etc., para después determinar la información exacta que será indispensable recabar que, a saber, la más común es la altura de los árboles, diámetro a la altura de pecho, altura a la base de la copa, especie, edad, ubicación, área basal, volumen, biomasa y ritmo de crecimiento, entre otra más especializada. También es necesario definir el diseño de muestreo, esto es, la densidad, tamaño, forma y demás características de la unidad de muestreo; no es práctico obtener mediciones de todos los árboles, por lo tanto se levantan datos de sólo unos de ellos, procurando sean representativos de toda la población. Así mismo, se fijan las técnicas de medición y se deciden los instrumentos a utilizar: medidores de altura y diámetro, cintas métricas, brújula, teodolito, medidores laser, mochila, etc., y se capacita al personal, para después propiamente ejecutar el levantamiento. Finalmente, se revisan y procesan los datos para elaborar el producto final.

La forma típica de hacer los inventarios es mediante mediciones en campo de las principales variables, tales como altura y diámetro a la altura de pecho, por ejemplo. Esta última es especialmente importante pues existen ecuaciones ya desarrolladas que relacionan el diámetro a la altura de pecho con otras variables, especialmente biomasa (Acosta et al., 2002; Avendaño et al., 2009). En muchas ocasiones es necesario derribar especímenes para determinar con más precisión los volúmenes y la biomasa de determinadas partes, como las hojas y el tronco, o para construir ecuaciones alométricas. Por otro lado, un elemento que implica mayores recursos es la estimación del crecimiento, pues es necesario realizar varias visitas a los mismos sitios y una medición de la altura más precisa, a menudo tomando lecturas de cambios de altura utilizando marcas permanentes en el tronco de los árboles; aunque en el caso de las coníferas esto se puede hacer extrayendo una muestra del tronco y evaluando los anillos anuales de crecimiento.

En contraparte, también existe la posibilidad de utilizar mediciones remotas para realizar inventarios forestales, cuya ventaja más importante es la reducción del trabajo de campo, lo que conlleva un ahorro de recursos en lo que respecta al transporte, el levantado de datos, el vaciado en una base de datos, etc. Por otro lado, y suponiendo ya una metodología en forma de estimación, la alternativa remota pudiera llegar a ser considerablemente más práctica, rápida y automatizada. En el caso de lidar, esto aunado al desarrollo creciente de sus tecnologías asociadas y la reducción paulatina de los costos de adquisición de los datos con el paso de los años.

Otra razón de peso en apoyo del uso de datos remotos, es que podrían reducirse las mediciones destructivas, disminuyendo la posibilidad de afectar el ambiente por mínimo que fuera, considerando como caso ideal que las metodologías pudieran llegar a ser lo suficientemente sofisticadas y los datos lo suficientemente densos como para poder caracterizar satisfactoriamente cada árbol y poder obtener la variable de interés con la precisión requerida. Por supuesto, el camino para llegar a este caso ideal es largo, pero el trabajar con lidar para medir cantidades de carbono bien podría contribuir a algún punto de partida.

Alrededor del mundo, existen regiones en donde ya se utiliza el lidar operacionalmente, especialmente en EU, Canadá y Europa existen compañías que se dedican a la adquisición, procesamiento y generación de cartografía (BC-CARMS, 2006:25-31), sin embargo, aún en esos lugares todavía existe un grueso flujo de trabajos dirigidos a mejorar y extender sus aplicaciones y abrir nuevas líneas de investigación, como lo demuestra el hecho de que la mayoría de los trabajos encontrados en la literatura se realicen en países de esas regiones (Henzel y Koch, 2011 [Alemania]; Rahman et al., 2009 [Holanda]; Hollaus et al., 2009 [Austria]; Hyyppä et al., 2006 [Finlandia]; Brandtberg, 2007 [Suecia]; Orka et al., 2009 [Noruega]; Hall et al., 2005 [Estados Unidos]; y Bortolot y Wynne, 2005 [Estados Unidos], por ejemplo), lo cual demuestra que todavía está en desarrollo. No se diga entonces en aquellos lugares donde su uso está poco extendido, incluyendo México, en donde la necesidad de explorarlo es mayor.

Por otra parte, además de esta falta de conocimiento de lidar en México, también existe una carencia de personal capacitado tanto para fines operacionales, académicos y de investigación, e incluso el desconocimiento de posibles consumidores, lo cual hace conveniente el desarrollo de trabajos que apunten a la expansión y comprensión de esta tecnología.

En adelante se hace una revisión de un extracto de los trabajos de estimación de variables forestales encontrados en la literatura, con objeto de que sirvan de orientación en la elección de los algoritmos que

se consideren oportunos para los fines de esta tesis, además de proporcionar un contexto en este ámbito. Además, se presenta una descripción del funcionamiento del lidar aerotransportado.

## 2.2 ESTIMACIÓN DE VARIABLES FORESTALES MEDIANTE LIDAR AEROTRANSPORTADO

Los primeros estudios de estimación de variables forestales llevados a cabo con mediciones laser se realizaron mediante la generación de perfiles de áreas boscosas, en los ochentas y noventas. La idea fue relacionar las variables que se pueden obtener de un perfil, alturas y área por ejemplo, con otras variables forestales, como altura media por parcela, densidad de copas, volumen y biomasa (Hyypä et al., 2008:1340).

Con el desarrollo de los equipos de medición laser aerotransportado, y la posibilidad de producir puntos georeferidos del dosel y la superficie terrestre, se abrió toda una gama de posibilidades para la estimación forestal, tanto para la determinación de alturas reales y construcción de modelos del terreno, como para la estimación de variables para los inventarios. En estas últimas las propuestas de trabajo, de la manera más general, se pueden dividir en dos grandes grupos: estimación de variables por árbol, y estimación por parcela o área forestal. En el primer caso se segmenta la nube para obtener los puntos pertenecientes a cada árbol, o generar modelos de las copas por árbol, de los cuales se pueden determinar tanto la altura como los parámetros de la copa (diámetro, área), usados en ecuaciones alométricas, de biomasa y regresiones, que posteriormente servirán para obtener datos agregados. En el segundo caso, la tendencia general es generar variables estadísticas de la nube, para mediante regresiones, y utilizando áreas ya censadas en campo, estimar las variables forestales de interés por área (Shan y Toth, 2009:347).

### 2.2.1 Estimación de variables forestales en general

Las variables de medición más directas al realizar estudios forestales identificando árboles individuales son la altura y el radio o diámetro de la copa, las cuales son base para calcular otros parámetros. De manera general la precisión de la estimación de alturas se encuentra en un rango de entre 0.5 m y 1.5 m. Schardt et al. (2002) reportan una subestimación promedio de 0.64 m, mientras que Roberts et al. (2005) obtienen un valor de RMSE (Raíz del Error Medio Cuadrático, del inglés Root Mean Square Error) de 0.48 m. De las mejores precisiones encontradas son los estudios de Heurich (2008), con subestimaciones medias de 0.54 m para la altura y 0.35 m para el radio de copa, y de Persson et al. (2002) con un RMSE de 0.63 m para la altura y 0.61 m para el diámetro de la copa. Roberts et al. (2005)



sobreestiman el área de la copa por 5 m<sup>2</sup>. La precisión de la estimación de estos valores está ligada con el proceso de segmentación y la densidad de puntos. La mayoría de los trabajos que se han realizado han sido con densidades de alrededor de 5 pts/m<sup>2</sup> o mayores; Naesset y Okland (2002) trabajando con aproximadamente 1 pt/m<sup>2</sup> obtienen una desviación estándar de la altura de 3.15 m, representando el 17.6 % de la media. Varios investigadores han enfocado sus esfuerzos en probar diferentes métodos de segmentación sin presentar mejoras significativas. Morsdorf et al. (2003) obtienen un RMSE de 0.60 m de la altura utilizando un algoritmo de agrupación K-means para definir los puntos por árbol. Hirata et al. (2009) se enfocan en probar la segmentación basada en cuencas en áreas con diferentes grados de densidad de árboles, obteniendo el mejor RMSE de 0.65 m. Pitkanen et al. (2004) proponen cinco formas diferentes de segmentación, obteniendo la mejor desviación promedio de altura de 0.46 m. Pysalo y Hyypä (2002) obtienen una desviación media de 0.29 m y error estándar de 1.2 m. Solberg et al. (2006) introducen un módulo de control de forma en su algoritmo de construcción de región obteniendo un RMSE del diámetro de la copa de 1.1 m.

Como se ha mencionado, a partir de la altura y parámetros de la copa por árbol se pueden estimar otras variables, las más comunes son el diámetro a la altura de pecho (dbh, del inglés diameter at breast height), volumen y biomasa, de esta última se hablará con más detalle más adelante. El estudio de Persson et al. (2002) muestra de las mejores precisiones: un RMSE de 0.38 m (10% de la media) para el diámetro del tronco y 0.21 m<sup>2</sup> (22% de la media) para el volumen del tronco. Hyypä et al. (2005), utilizando ecuaciones alométricas, obtienen una desviación estándar del volumen del tronco de 0.36 m<sup>2</sup> (30% de la media). Popescu (2007) y Yu et al. (2011), trabajando con una densidad de 2.6 pts/m<sup>2</sup>, obtienen RMSE del diámetro a la altura de pecho del 18% (0.049 m) y del 21.35 % de la media, éste último obteniendo un RMSE del volumen del tronco del 45.77 %. Heurich (2008) reporta RMSE del diámetro a la altura de pecho del 0.057 m para árboles caducifolios y 0.053 m para coníferas.

El paso siguiente después de obtener atributos por árbol, es agregarlos para tenerlos por área. Aunque la mayoría de los estudios de estimación basados en árboles individuales se quedan hasta este nivel, se pueden encontrar en la literatura ejemplos de quienes sí los agregan. Hyypä e Inkinen (1999) obtienen el área basal, la altura media y el número de troncos por área, Naesset y Okland (2002) calculan la altura a la base de la copa promedio y la longitud relativa de la misma (longitud vertical de la copa con respecto a la altura total del árbol) promedio por parcela, Schardt et al. (2002) estiman el volumen de madera por área, y Bortolot y Wynne (2005) obtienen la densidad, número de árboles, por hectárea.

Así mismo, se pueden encontrar toda una serie de mediciones de otros atributos que no son tan comunes en los inventarios forestales, pero que pueden servir de apoyo para, de alguna manera, cuantificar y caracterizar la estructura de cada árbol o área forestal. Algunos son la altura de la rama más baja (Kato et al., 2009), el área foliar por árbol (Roberts et al., 2005), la longitud relativa de la copa (Naesset y Okland, 2002), anteriormente citada, y la altura media del 20% de los árboles más altos (Schardt et al., 2002). Hirata et al. (2009) generan tres áreas con diferentes grados de adelgazamiento del dosel para estimar el radio de penetración de los pulsos lidar por zona. El adelgazamiento fue generado artificialmente mediante una tala controlada, entregando áreas muy adelgazadas, medianamente adelgazadas y no adelgazadas. El radio de penetración fue definido como el porcentaje de pulsos que alcanzaron el suelo con respecto al número total de pulsos en una determinada zona. Utilizando una densidad de puntos de 40.5 pts/m<sup>2</sup> los radios de penetración obtenidos fueron del 50.6 % para al área muy adelgazada, 43.1 % para la medianamente adelgazada y de 9.2 % para la no adelgazada. Los RMSE de la altura media en cada tipo de zona fueron de 0.95 m para la muy adelgazada, 0.65 m para la medianamente adelgazada y de 0.68 m para la no adelgazada. Era de esperarse que la estimación de altura para el área muy adelgazada hubiera sido más exacta, pues por haber más espacios entre árboles, estos se pueden delinear y caracterizar mejor por la nube de puntos, sin embargo, el RMSE es mayor para esta área. Hirata et al. no proporcionan conclusión o explicación alguna para este comportamiento.

Hasta aquí todos los estudios citados han sido de estimaciones basadas en árboles individuales, sin embargo, vale la pena hacer mención también de algunos trabajos llevados a cabo con la aproximación del cálculo de variables forestales por área. La mayoría de estos estudios utilizan la nube de puntos como datos de entrada, sin embargo los atributos por área también se pueden generar de datos raster o de onda continua. En Lefsky et al. (1999) se obtienen modelos de regresión para calcular biomasa y área basal utilizando un perfil del dosel de un área ya censada, para después aplicarlos a otra zona con el perfil del dosel generado mediante lidar. Las regresiones fueron hechas utilizando como variables independientes índices de altura (máximo, mediana, media y media cuadrática de la altura del dosel) calculados a partir del perfil. Por otro lado, Nilsson (1996), utilizando datos lidar de onda completa, estima el volumen y la altura media por parcela, basándose en las diferencias entre picos de la onda recibida y el área de la misma. Las diferencias reportadas en la altura son de entre 2.1 m y 3.7 m.

Con respecto a los trabajos utilizando datos discretos para estimar variables por área, para la mayoría de ellos no es necesario una gran densidad de puntos, encontrándose valores de alrededor de 1 pt/m<sup>2</sup>. Los estudios de Maltamo et al. (2005, 2006) son una excepción, trabajando con densidades de 12 pts/m<sup>2</sup> y 10 pts/m<sup>2</sup>. En Maltamo et al. (2006) se calcula el volumen del tronco por parcela, obteniendo un RMSE

del 10 % al 14 %. Bortolot (2006) estima la densidad de árboles obteniendo un RMSE del 8.4 % del valor de la media. Uno de los estudios más completos es el llevado a cabo por Hall et al. (2005), en donde trabajando con 1.2 pts/m<sup>2</sup> y más de 5 retornos por pulso con datos de intensidad y número de retorno por pulso se genera una gran cantidad de variables basadas en las estadísticas de la altura, los valores de la intensidad, la energía y el tipo de punto (suelo, vegetación) de la nube, y se buscan los modelos de regresión multivariable de mejor desempeño para la estimación de algunos de los parámetros forestales más importantes, a saber: la altura media (coeficiente de determinación del 57 %), altura media de Lorey (coef. determinación 87 %), biomasa total (coef. determinación 74 %), biomasa del follaje (coef. determinación 79 %), área basal (coef. determinación 79 %), densidad (coef. determinación 67 %), altura a la base del dosel promedio (coef. determinación 80 %) y densidad del combustible de la copa (coef. determinación 83 %).

Menos común es la caracterización de la estructura de los niveles de vegetación del bosque. Zimble et al. (2003) mediante un análisis de variabilidad de alturas del dosel en células de 30 m x 30 m generan una clasificación del bosque en zonas uni-nivel o multi-nivel, obteniendo una precisión de clasificación del 97 %. Maltamo et al. (2005) estiman el número y altura media de los árboles suprimidos en el área de estudio, obteniendo errores estándar de 0.31 para el número de árboles suprimidos y 0.13 para la altura media.

Como se ha mencionado, la forma típica de calcular parámetros forestales por área es mediante la implementación de modelos de regresión, en donde el parámetro a estimar se ha obtenido de muestreos de campo y las variables independientes son estadísticas de las alturas de la nube, variables obtenidas de la intensidad, proporción de cada tipo de retorno (vegetación o suelo, primer retorno, "n" retorno o último retorno, etc.) y densidad del dosel; sin embargo también se han trabajado con otras variables. Bortolot (2006) identifica grupos de árboles en el raster generado a partir de los datos lidar y de ahí obtiene el porcentaje de píxeles que pertenecen a algún grupo con respecto al total de píxeles de la imagen y el porcentaje de píxeles interiores en los grupos, que utiliza como variables de entrada en las regresiones. En Jochem et al. (2010) se utiliza el volumen del dosel como variable independiente para calcular biomasa, tomado como el volumen contenido entre la superficie del dosel y la superficie del terreno.

## 2.2.2 Estimación de biomasa

Se realiza una revisión de algunos estudios de estimación de biomasa por separado, por la similitud entre el cálculo de ésta y el de almacenamiento de carbono. La biomasa es la cantidad de materia orgánica (en peso) presente en una determinada área, que en el caso específico de los inventarios forestales, se limita únicamente a aquella representada por la vegetación. La cantidad de carbono secuestrada se puede considerar como un segmento de la biomasa, cuya proporción está en función del tipo de especies, pudiendo derivarse de ahí su cálculo; o incluso directamente de las variables que pueden ser extraídas de los datos de percepción remota.

En lo que respecta a las propuestas de estimación por árbol, Popescu (2007) presenta un método que muestra dos fases: primero, después de obtener la altura y el diámetro de copa por árbol, mediante regresiones estima el diámetro a la altura de pecho y la biomasa, y después introduce un segundo método de cálculo de biomasa e incluso la proporción de ésta que pertenece a cada parte del árbol (biomasa del follaje, de las raíces, de la corteza y del tronco) mediante ecuaciones alométricas, presentando un RMSE de 169 kg por árbol, lo cual representa el 47 % de la media.

En general, la forma de trabajo es un poco diferente: se segmentan los datos por árbol y se agregan para, mediante regresiones, estimar la biomasa por área. En Popescu et al. (2003) se utilizan variables derivadas de las alturas y de la copa de los árboles por parcela, obteniendo un RMSE de la biomasa de 31.28 ton/ha. En Bortolot y Wynne (2005) las variables independientes son estadísticas tomadas de los percentiles de alturas y número de árboles detectados por área, para obtener un RMSE de la biomasa de 13.9 ton/ha, representando el 14.9 % de la media.

Ya que en muchas ocasiones la biomasa se utiliza expresada en unidades de peso por hectárea, no es necesario segmentar los datos por árbol. Se han realizado varios estudios de este modo: Lefsky et al. (1999) obtienen índices de altura obtenidas del perfil del dosel del área de estudio y reporta una subestimación de la biomasa de 16 ton/ha; Jochem et al. (2010), ya mencionados anteriormente, presentan una ecuación de biomasa en función del volumen del dosel, calculado tanto a partir de la nube de puntos como del raster generado por la misma nube, y obtiene desviaciones estándar de 87.60 ton/ha utilizando los puntos, y 84.11 ton/ha utilizando raster; Hall et al. (2005) calculan la biomasa total y la del follaje utilizando estadísticas de altura, intensidad, energía y tipo de puntos de la nube; y Bortolot (2006), con su propuesta de identificar grupos de árboles en la imagen generada de la nube, obtiene un RMSE del 13.3 % del valor de la media.

### 2.2.3 Segmentación de copas

Durante los últimos años y con el avance de los equipos lidar y la implementación de metodologías para la extracción de información de las nubes de puntos, se han desarrollado una variedad de algoritmos y métodos de segmentación, muchos de ellos para identificar árboles. Estos procesos de segmentación se pueden dividir en dos fases: la identificación propia del árbol, y la delineación y/u obtención de atributos de la copa, que aunque en muchos casos se realizan conjuntamente, en otros casos puede observarse una división clara entre estas dos operaciones.

En lo referente a la identificación de árboles la tendencia general y el método lógico es establecer a aquellos puntos de mayor altura como los caracterizadores de la ubicación de cada individuo. Se genera una superficie o una imagen raster del dosel, cuya altura sobre el nivel de suelo se obtiene al sustraerle la altura del modelo digital de terreno en cada punto, se le aplica un filtro para suavizarla y sobre ésta se buscan los puntos más altos localmente (Hyypä et al., 2005, Schardt et al., 2002, y Vega y Durrieu, 2011, por citar algunos ejemplos). Para reducir el riesgo de árboles mal identificados, las ventanas de búsqueda se hacen dinámicas: circulares de radio variable en función del tamaño esperado de los árboles en la zona (Roberts et al., 2005), o de dimensiones que cambian en función de la altura local (Popescu, 2007). En Popescu et al. (2003) se experimenta con dos formas de ventana, circular y cuadrada, de tamaño que varía en función de la altura y especie local, éstas últimas obtenidas de una clasificación llevada a cabo en una imagen ATLAS (Airborne Terrestrial Land Applications Scanner, de la NASA).

Sin embargo, también se han dado otros esfuerzos, más sofisticados, en dirección de una búsqueda de mejora en el reconocimiento de árboles. En el estudio de Pitkanen et al. (2004) se implementan cinco formas diferentes de identificación: a) por máximos locales en el modelo digital de alturas del dosel con suavizado en función de la altura, b) abundante selección de posibles ubicaciones para después reducirlas en base a la pendiente del área circundante de cada una y a la distancia y profundidad del valle encontrado entre dos ubicaciones candidatas, c) método basado en la identificación de cuerpos en una representación de escala-espacio, d) por máximos locales en el modelo digital de alturas del dosel crudo (sin filtrado) y, e) por máximos locales en el modelo digital de alturas del dosel con un filtro gaussiano. Los porcentajes de individuos correctamente identificados por método, respectivamente, fueron de 37.0 %, 41.6 %, 41.5 %, 49.2 % y 36.7 %. Es de notar que el mejor porcentaje haya sido del método en el cual el modelo digital de alturas del dosel se utiliza sin filtrado, el método más simple. En Pyysalo y Hyypä (2002), primeramente se generan las nubes de puntos por cada árbol apoyándose de un algoritmo basado en cuencas, para después obtener la ubicación precisa del tronco promediando las

coordenadas de todos los puntos, obteniéndose una error estándar de 0.87 m en el sentido del eje de las x y de 0.79 m en el eje de las y.

Para la fase de la segmentación de copas, dos de los algoritmos de mayor uso son el basado en cuencas, en el cual el modelo digital del dosel se invierte, quedando los puntos identificados como ubicaciones de árboles como los más bajos, y a cada uno de ellos se les asocian todos los píxeles cuya pendiente confluya hacia ellos, construyendo así la “cuencas” que representan las copas (ejemplos en Hyyppä e Inkinen, 1999, y Yu et al., 2011); y el de crecimiento de región, en el cual se parte una serie de puntos semilla (los puntos que representan la ubicación de cada individuo), a los cuales se van agregando píxeles en base a un criterio de similitud para ir construyendo las regiones que finalmente representarán las copas. Un ejemplo de este último se puede encontrar en Solberg et al. (2006), en donde también se incorpora un algoritmo de control de forma de la región.

Propuestas alternativas a las anteriores también han sido desarrolladas. En Bortolot y Wynne (2005) primeramente se calcula el radio de copa más probable para cada píxel que esté por encima de un umbral de altura y que no sea un mínimo local, para después elegir como una ubicación correcta de árbol a todo píxel que dentro de su circunferencia asociada no se encuentre otro píxel con una altura mayor. En Persson et al. (2002) la localización de cada individuo es dada por máximos locales en el modelo digital del dosel con diferentes grados de suavizado, a cada punto máximo se le ajusta una superficie parabólica utilizando sólo el 30 % más alto de los datos, y mediante un análisis de las diferencias de cada superficie con respecto al modelo digital del dosel en las zonas en donde se traslapan, se descartan árboles incorrectamente identificados. Este método también es usado en los estudios de Heurich (2008) y Holmgren y Persson (2004). En Morsdorf et al. (2003) cada máximo encontrado es un punto semilla para la implementación de un algoritmo de agrupación (K-means) que minimiza la suma de las distancias de los puntos que componen cada clase con respecto al centroide de la clase. En Rahman et al. (2009), trabajando con una densidad de 70 pts/m<sup>2</sup> se consigue clasificar la nube de puntos, además de por árbol, incluso en grupos pertenecientes a la copa, al tronco y a la vegetación subyacente. La delineación de la copa es llevada a cabo mediante un algoritmo de segmentación basado en cuencas invertido, la ubicación del tronco es el punto donde se encuentre la mayor densidad de puntos dentro de cada segmento, los puntos pertenecientes a la vegetación subyacente se determinan mediante un umbral de altura, y finalmente la clasificación en puntos del tronco y de la copa se realiza con un algoritmo de crecimiento de región basado en distancias y densidad de puntos.

Existen otros métodos que no son propiamente algoritmos de segmentación de copas, pero cuyo objetivo también está enfocado en la obtención del diámetro de éstas. En Roberts et al. (2005), después de identificar los puntos máximos, se calculan los radios en cuatro direcciones perpendiculares, sencillamente tomando como límites de la copa aquellos píxeles con menor altura en el perfil del dosel en cada dirección, dentro de una ventana de 3.4 m a partir del punto máximo. En Popescu et al. (2003) y Popescu (2007) se ajustan ecuaciones polinómicas de cuarto grado a los dos perfiles perpendiculares del dosel que se intersecan en cada máximo local, para mediante la identificación de los puntos críticos de las funciones definir los límites de la copa, a partir de los cuales se calcula el diámetro.

#### 2.2.4 Identificación de especies

Existen parámetros forestales para los cuales es necesario conocer la especie para poder calcularlos o para mejorar la precisión de la estimación, como es el caso de la biomasa y la captura de carbono, para las cuales existen ecuaciones ya desarrolladas que hacen discriminación entre especies.

Muchos de los trabajos realizados en clasificación utilizan Análisis de Discriminantes Lineales como método de diferenciación entre especies, y también la mayoría de ellos a nivel de árbol individual. En Heinzl y Koch (2011) se definen cuatro niveles de variables a utilizar para la discriminación: a) características de los datos (intensidad, amplitud, ancho de onda, retornos por pulso), b) medidas de tendencia central y de la variabilidad, c) tipo de retorno (primero, último, todos, alguno en especial), y d) tipo de filtro aplicado a los puntos, y se prueban todas las posibles combinaciones tomando una variable de cada nivel para la clasificación, obteniendo como mejores precisiones el 57 % tomando en cuenta todas las especies involucradas en el estudio (6 en total), 78 % para las cuatro especies principales y 91 % para la diferenciación entre coníferas y árboles de hoja ancha. En Hollaus et al. (2009) las variables involucradas son la densidad del dosel, el ancho promedio de ecos, la desviación estándar de los anchos de los ecos y la sección transversal de la retrodispersión, todas ellas para los datos por encima del percentil 50 de altura, las cuales mediante una clasificación basada en funciones lineales de probabilidad entregan una precisión del 83% entre caducifolios y coníferas, y del 75 % considerando las tres especies involucradas en el estudio. Ambos trabajos utilizan datos lidar de onda completa.

Partiendo de datos lidar discretos, Holmgren y Persson (2004) obtienen una de las mejores precisiones, 95 % en la clasificación, utilizando funciones de discriminación lineales a partir de variables basadas en la proporción de los diferentes tipos de retornos, distribución de alturas, geometría e intensidad. El estudio de Orka et al. (2009) sale del criterio de los discriminantes lineales y propone modelos de regresión lineal

para la clasificación, teniendo como variables independientes la especie y la altura y como variable dependiente cada uno de los parámetros obtenidos de los datos lidar, estos últimos basados en la distribución de alturas, profundidad de penetración de los pulsos al dosel y densidad de la copa (variables estructurales), así como de la intensidad. Las precisiones obtenidas fueron del 77 % usando variables estructurales, 73 % variables de intensidad y 88 % combinadas. En Brantberg et al. (2003) se utilizan el valor máximo, número e índices de la distribución de los retornos del suelo y de la vegetación para la clasificación mediante el Análisis de Discriminantes Lineales, obteniendo una precisión del 60 %.

También existen casos en los que se busca mejorar la clasificación valiéndose de la fusión entre datos lidar con imágenes. En Puttonen et al. (2011) se utilizan datos hiperespectrales para caracterizar cada punto lidar, aparte de las variables típicas obtenidas de la nube, en este caso estadísticas de las alturas y de la proporción de retornos por alturas. Las precisiones obtenidas fueron del 90.5 % (caducifolios-coníferas) y 65.4 % (todas las especies) utilizando sólo variables lidar, del 90.5 % (caducifolios-coníferas) y 62.4 % (todas las especies) utilizando sólo variables hiperespectrales y del 95.8 % (caducifolios-coníferas) y 83.5 % (todas las especies) utilizando ambos tipos de variables conjuntamente.

Hasta aquí todos los estudios mencionados han sido de clasificación por individuos, sin embargo también se han llevado a cabo trabajos haciendo la discriminación por área. En Moffiet et al. (2005) se exploran diferentes variables potenciales para la clasificación de especies por parcela: proporción de retornos singulares, permeabilidad de la vegetación, media y desviación estándar de las alturas de primer retorno, variables de intensidad y Cobertura Proyectiva de Follaje, las cuales mediante un Análisis de Discriminantes Lineales entregan una precisión del 77 %. En Suratno et al. (2009) se clasifican células de 400 m x 400 m en especies utilizando las medias y desviaciones estándar de las intensidades y alturas del dosel para tres tipos de retornos: todos, los primeros de múltiples retornos y los retornos sencillos, mediante un Algoritmo de Clasificación de Máxima Probabilidad supervisado, obteniendo una precisión del 75 %.

### 2.3 EL SISTEMA LIDAR AEROTRANSPORTADO

El láser (amplificación de luz por emisión estimulada de radiación, del inglés light amplification by stimulated emission of radiation) es un haz de luz monocromático, coherente, direccional y que puede ser fuertemente enfocado (Lohani, 2008:2), y es una tecnología que hoy es utilizada en una amplia gama de aplicaciones, incluyendo sistemas de medición de distancias. Para que un rayo laser pueda ser utilizado en sistemas lidar debe poder ser producido en pulsos cortos de alta potencia y con poca



divergencia, abarcando un ancho de banda angosto dentro del espectro electromagnético (en el caso de la medición por tiempo de vuelo), o con la capacidad de poder modularse su frecuencia (que es el caso de la medición por diferencia de fase) (Lohani, 2008:2; Wehr y Lohr, 1999:74).

LIDAR (del inglés Light Detection and Ranging o Laser Imaging Detection and Ranging) es un sensor remoto activo que utiliza señales laser para la detección de distancias entre él y el objeto de interacción, mediante las cuales se genera una nube de puntos georeferidos en un espacio tridimensional cuya forma se amolda a la superficie que se está percibiendo y a los elementos que están sobre ésta; esto implica necesariamente el conocimiento de la ubicación del sensor en el momento de cada medición, lo cual se logra equipándolo con un receptor GPS (Sistema de Posicionamiento Global, del inglés Global Positioning System).

Existen dos formas de medir la distancia entre el sensor y el objeto de interacción: medición del tiempo de vuelo del pulso y mediante diferencia de fase. En el primer caso el sensor emite un pulso laser corto y de gran intensidad y mide el tiempo que transcurre entre la emisión y la detección del mismo, después de ser reflejado por el objeto. Conociendo la velocidad de la luz se puede determinar la distancia. En el segundo caso el sensor emite un haz de energía continuo y modulado (frecuencia constante), que al momento de originarse presenta una intensidad inicial, y al momento de ser captado por el sensor después de ser reflejada llega con una intensidad diferente. Conociendo esta diferencia entre las fases de salida y llegada y la longitud de onda se determina la distancia. Comercialmente, el método de medición más utilizado es el de tiempo de vuelo del pulso.

Los primeros esfuerzos que se hicieron para obtener mediciones de la superficie terrestre utilizando laser fueron llevados a cabo en los sesentas y setentas del siglo pasado. Se montaba el dispositivo de medición sobre una aeronave y se obtenían distancias entre ésta y la superficie únicamente sobre la ruta de vuelo, lo cual permitía obtener un perfil de los elementos avistados. Esto contribuía a limitar su utilización, por la poca capacidad de cobertura.

La adopción y uso más extenso de los escáneres laser aerotransportados empezó a mediados de los noventa; hubo varios factores que contribuyeron a la explotación de estos sistemas hasta esa década a pesar de que el laser apareció a mediados de ese siglo: el principal, el desarrollo y puesta en funcionamiento del sistema GPS que permitió dotar de coordenadas a la aeronave, necesarias para dar coordenadas también a los puntos detectados; el desarrollo de las Unidades de Medición Inercial, IMU (del inglés Inertial Measurement Unit); las mejoras en el diseño y capacidades de los sensores laser en sí

mismos; el incremento en las capacidades de procesamiento y almacenamiento de los sistemas computacionales y; el mayor conocimiento de los usuarios y potenciales usuarios de los escáneres laser (Wagner et al., 2004:1)

Actualmente existe una gran variedad de sistemas lidar, usados en diferentes ámbitos comerciales y de investigación (figura 2.1). Desde el punto de vista más amplio, se suelen dividir en dos grupos: aquellos sistemas usados para fines topográficos, utilizados sobre la superficie, y que operan en la porción infrarroja del espectro electromagnético; y los usados para batimetría, utilizados para obtener mediciones de los fondos de los cuerpos de agua, que operan en la porción azul y verde del espectro electromagnético (Burtch, 2002:2), aunque en los últimos años también se pueden agregar aquellos usados para mediciones atmosféricas como temperatura, condiciones de las nubes, detección de contaminantes, etc., que operan en la porción ultravioleta visible e infrarrojo cercano del espectro (Real, 2011:9). En este trabajo únicamente se tratarán los del primer grupo (para fines topográficos).

Respecto al diseño del sistema, se pueden dividir en tres categorías: aquellos manufacturados en masa para fines comerciales; los personalizados, ajustados a necesidades específicas de los usuarios y; los utilizados para experimentación y demostración, destinados a propósitos científicos. El patrón de la distribución de los puntos sobre la superficie terrestre también puede adoptar varias formas en función del tipo de dispositivo de escaneo utilizado: en forma de sierra dentada o sinusoidal, en forma de líneas paralelas dispuestas de forma transversal a la dirección de vuelo, de forma elíptica, o de líneas paralelas dispuestas en la misma dirección del vuelo. Más adelante se proporcionará una explicación más a detalle de estos. De acuerdo a la distancia máxima de medición se pueden considerar equipos de dos tipos: escáneres de corto alcance, operados a relativa baja altura y velocidad, montados típicamente sobre helicópteros y utilizados para obtener las características de elementos lineales como líneas de transmisión eléctrica, redes viales, etc.; y escáneres de mayor alcance, montados sobre aviones, utilizados para la trabajos de gran cobertura (Shan y Toth, 2009:43-47).

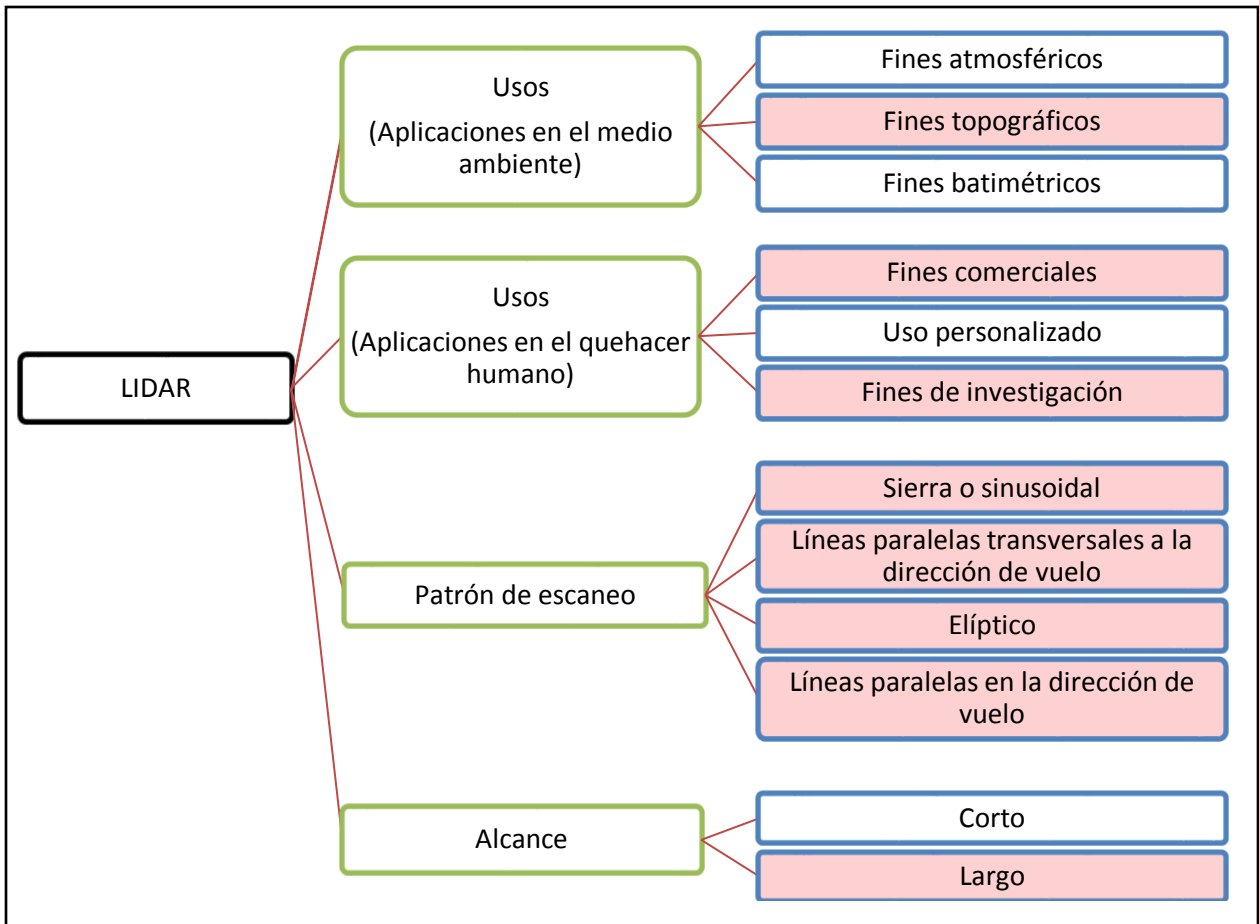


Figura 2.1. Diagrama de la clasificación de los sistemas lidar de acuerdo a diferentes criterios. En rosa se destacan los utilizados para fines forestales.

### 2.3.1 Medición de distancias por tiempo de vuelo del pulso

La medición de una distancia por este método básicamente consiste en la emisión de un pulso corto laser que impacta en el objetivo y es reflejado, para regresar al dispositivo, donde es detectado (figura 2.2). Mediante la medición del tiempo que transcurre entre la emisión y la detección, y considerando la velocidad del haz, se puede calcular la distancia que recorrió éste en toda su trayectoria, que será el doble de la distancia del dispositivo al objetivo, pues realizó un trecho de ida y otro de vuelta. Esto es:

$$D = \frac{v * t}{2} \quad (2.1)$$

Donde:

$D$  es la distancia entre el dispositivo y el objetivo,

$v$  es la velocidad del haz, y

$t$  es el tiempo transcurrido entre el momento de la emisión y la detección (Shan y Toth, 2009:4).

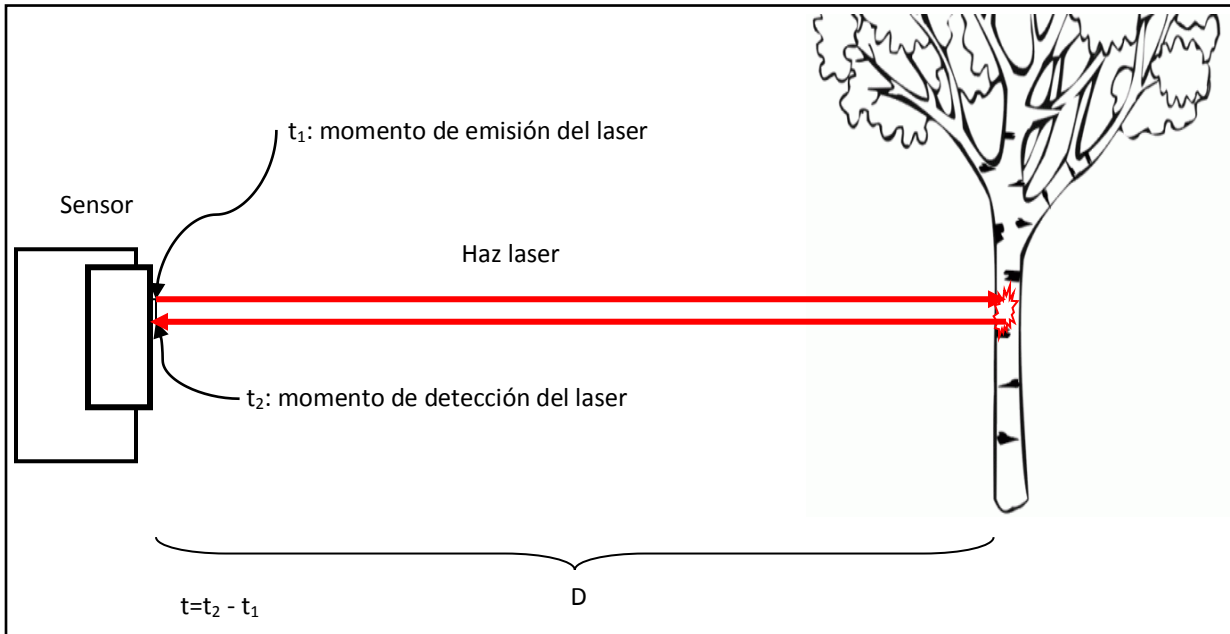


Figura 2.2. Ilustración del método de medición por tiempo de vuelo.

La resolución de las mediciones de distancia está en función de la resolución del dispositivo de medición del tiempo, únicamente, puesto que la velocidad de la luz se conoce con gran precisión:

$$\Delta D = \frac{v * \Delta t}{2} \quad (2.2)$$

Donde:

$\Delta D$  es la resolución en la medición de distancias, y

$\Delta t$  es la resolución del dispositivo de medición del tiempo (Lohani, 2008:10).

Para poder hacer mediciones con resolución de 1 cm, es necesario que la medición del tiempo se haga con una precisión de 0.3 ps (0.3 picosegundos, esto es,  $3/10^{13}$  segundos) (Shan y Toth, 2009:17).

La máxima distancia que puede ser medida está en función de la potencia del pulso laser emitido, pues debe ser suficientemente potente para originar un retorno que pueda ser medible y distinguible de la

energía recibida del sol u otras fuentes. En este proceso también intervienen otros factores: el grado de divergencia del rayo, la atmósfera, la reflectividad del objetivo y la sensibilidad del sensor (Lohani, 2008:11).

En los dispositivos con este principio de medición típicamente el rayo laser es generado en un medio activo de estado sólido o semiconductor, después una pequeña parte de éste es desviado y canalizado al dispositivo de medición de tiempo para inicializarlo, y una serie de lentes alinean el rayo procurando que tenga la menor divergencia posible. Al incidir sobre el objetivo, el pulso es dispersado y sólo una parte regresa al sensor. Esta energía recibida es concentrada mediante otra serie de lentes y necesita ser amplificada, para después ser filtrada, eliminando de la señal los efectos de la luz del sol u otras fuentes que pudieran introducir ruido. Finalmente, se detiene el contador de tiempo, y se obtiene el cálculo de la distancia (Shan y Toth, 2009:15-16).

Si varios retornos son detectados para un solo pulso, entonces el número de retorno también se registra. La longitud del haz liberado por el dispositivo determina la separación mínima que debe haber entre dos elementos propensos a generar retornos para poder efectivamente detectarlos. Si el pulso dura 10 ns (10 nanosegundos, esto es,  $1/10^{10}$  segundos), por ejemplo, tendrá una longitud de 3 m; entonces no podrán generarse dos retornos de elementos separados por menos de esta distancia (Shan y Toth, 2009:16).

### 2.3.2 Medición de distancias por diferencia de fase

A diferencia de la anterior, en este caso el sensor emite un haz laser continuo, con una longitud de onda controlada y bien definida. La señal rebota en algún elemento del medio que se esté escaneando y regresa al sensor, donde es captada y amplificada. Al momento de recibirse, la fase de la señal muy probablemente será diferente a la de salida; esta diferencia de fase es medida y transformada a distancia. La medición final está dada por el número de longitudes de onda completas ocurridas a lo largo de toda la trayectoria del rayo desde su salida hasta su llegada al sensor, más la fracción de longitud de onda necesaria para cubrir la distancia (la diferencia de fase transformada a distancia), todo dividido entre dos (figura 2.3):

$$D = \frac{n\lambda + \Delta\lambda}{2} \quad (2.3)$$

Donde:

$D$  es la distancia del sensor al objetivo,

$\lambda$  es la longitud de onda de la señal,

$n$  es el número de longitudes de onda completas en todo el recorrido de la señal, y

$\Delta\lambda$  es la fracción de longitud de onda (Shan y Toth, 2009:5-6).

Cabe mencionar que la longitud de onda de las señales laser es muy pequeña (señal portadora), lo que se hace es que su amplitud se modula en función de una señal con longitud de onda mayor (onda de medición) que es más práctica para obtener las diferencia de fase (figura 2.4) (Shan y Toth, 2009:5).

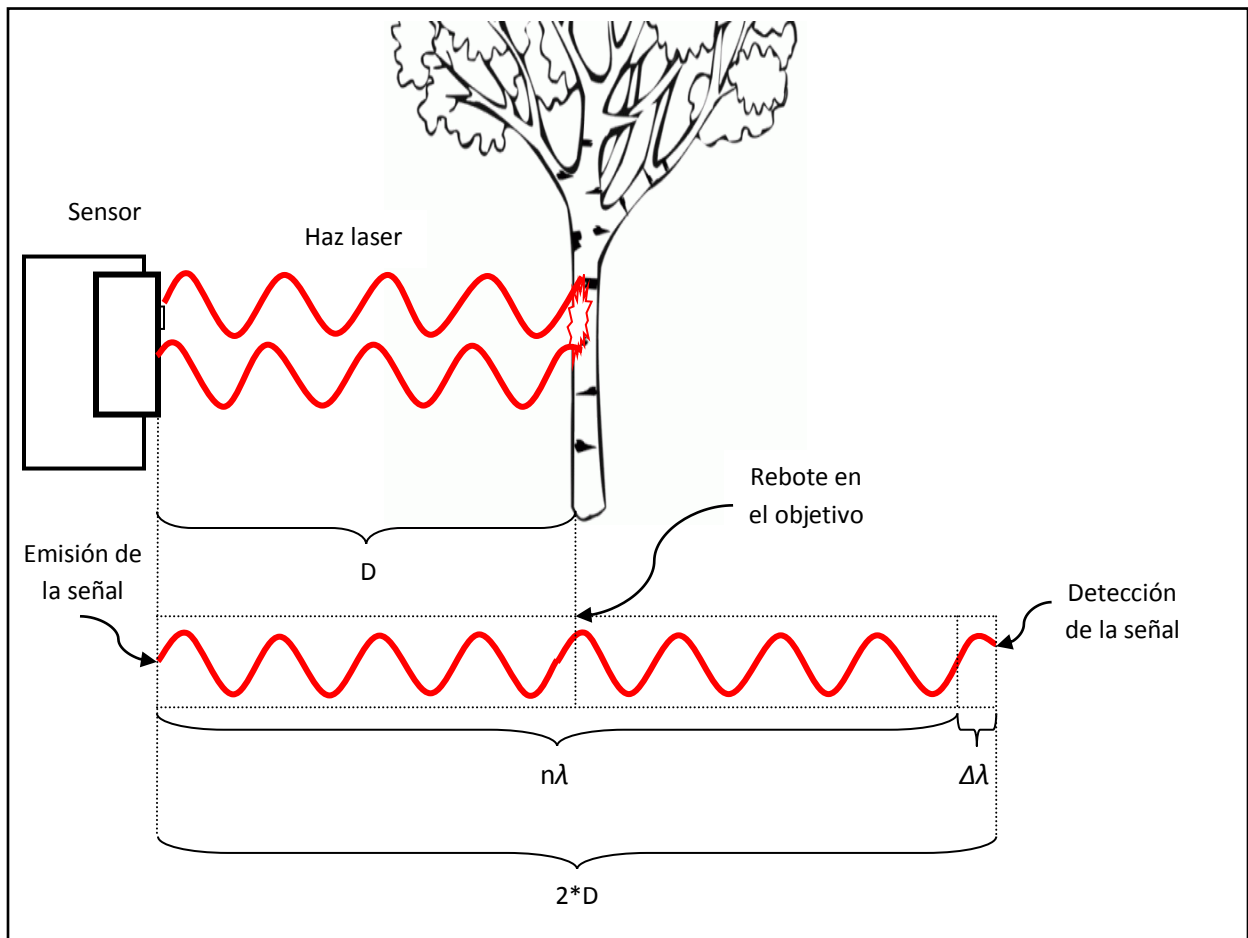


Figura 2.3. Ilustración del método de medición por diferencia de fase.

$\Delta\lambda$  es calculado a partir de la diferencia de fase medida en radianes ( $\varphi$ ):

$$\Delta\lambda = (\varphi/2\pi) \cdot \lambda \quad (2.4)$$

El número de longitudes de onda completas ( $n$ ) no es posible conocerlo con una sola medición. Determinar  $n$  es lo que se conoce como resolución de la ambigüedad; la forma de hacerlo es realizar varias mediciones al mismo objetivo con diferentes longitudes de onda, esto nos dará una serie de ecuaciones de la siguiente forma:

$$D = \frac{n_1 \lambda_1 + (\varphi_1 / 2\pi) \cdot \lambda_1}{2}$$

$$D = \frac{n_2 \lambda_2 + (\varphi_2 / 2\pi) \cdot \lambda_2}{2}$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$D = \frac{n_m \lambda_m + (\varphi_m / 2\pi) \cdot \lambda_m}{2}$$

En donde los parámetros desconocidos son  $D$  y  $n_m$ , que se encuentran resolviendo el sistema (Shan y Toth, 2009:19).

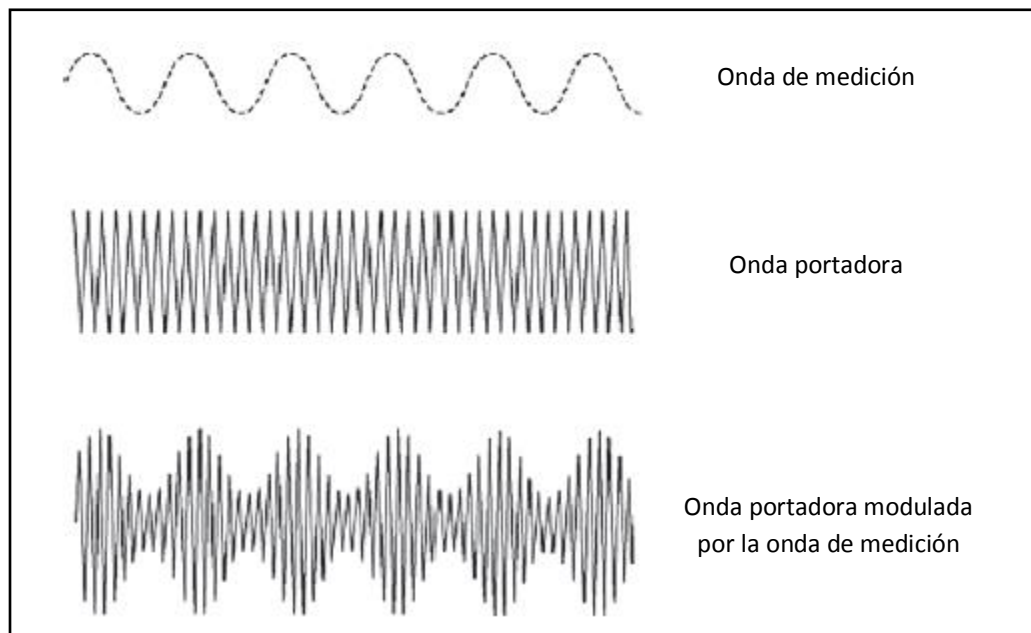


Figura 2.4. Ilustración de la onda de medición y la onda portadora del laser. La longitud de onda del laser es muy pequeña, para poder hacer las mediciones es necesario modularla, para tener una mayor.

La resolución de las mediciones está en función del intervalo de fase mínimo que sea capaz de leer el dispositivo y de la mínima longitud de onda utilizada. A mayor resolución de la medición de fase y menor longitud de onda, se obtiene mayor precisión en la distancia:

$$\Delta D = \frac{\lambda_{menor}}{4\pi} \cdot \Delta\varphi \quad (2.5)$$

Donde:

$\Delta D$  es la resolución en la medición de la distancia,

$\lambda_{menor}$  es la longitud de onda menor utilizada por el sensor, y

$\Delta\varphi$  es la resolución en la medición del ángulo de fase (Wehr y Lohr, 1999:71-72).

Las consideraciones hechas con respecto a la máxima distancia que puede ser medida por el método de vuelo del pulso también pueden ser apreciadas para este método, la mayor distancia estará en función de la potencia del rayo principalmente, en conjunción con otros factores.

La ventaja de la medición por diferencia de fase es que permite obtener valores altamente precisos, por supuesto, en función de la longitud de onda mínima utilizada; sin embargo, su desventaja reside en que es complicado mantener la generación continua de ondas de alta energía (Lohani, 2008:10). Por ejemplo, utilizando una longitud de onda de 30 cm y una resolución en la medición de la fase de  $0.4^\circ$ , se pueden obtener mediciones de las distancias con una precisión de 0.2 mm, la cual requeriría de un medidor de tiempo muy demandante en su diseño en un sistema basado en tiempo de vuelo. Ello aunado a la necesidad de varias frecuencias para resolver el problema de la ambigüedad y de que no se requieren grandes precisiones (mayores a 5 cm) en la medición de las distancias en escaneos aéreos relega a este método a ser usualmente sólo utilizado en aplicaciones de corto alcance (Shan y Toth, 2009:134-135).

### 2.3.3 Partes que componen el sistema lidar aerotransportado

En la figura 2.5 se muestra, de forma esquemática, los componentes principales del lidar aerotransportado, siendo la gran mayoría de estos equipados para obtener mediciones por tiempo de vuelo del pulso. Estrictamente hablando sólo se compone de la unidad generadora y receptora de la señal, de la de distribución de los pulsos, el sistema de posicionamiento (GPS/Unidad de Mediciones Inerciales) y la unidad de almacenamiento y control, sin embargo, también debe considerarse el control terrestre necesario para conseguir precisiones aceptables en el posicionamiento de la aeronave y, en este caso, también se repara en los elementos necesarios para sincronización y registro de las nubes de puntos, tareas necesarias para poder utilizarlas. En muchos casos, igualmente se considera algún elemento para obtener imágenes (cámara) como parte integral del sistema.



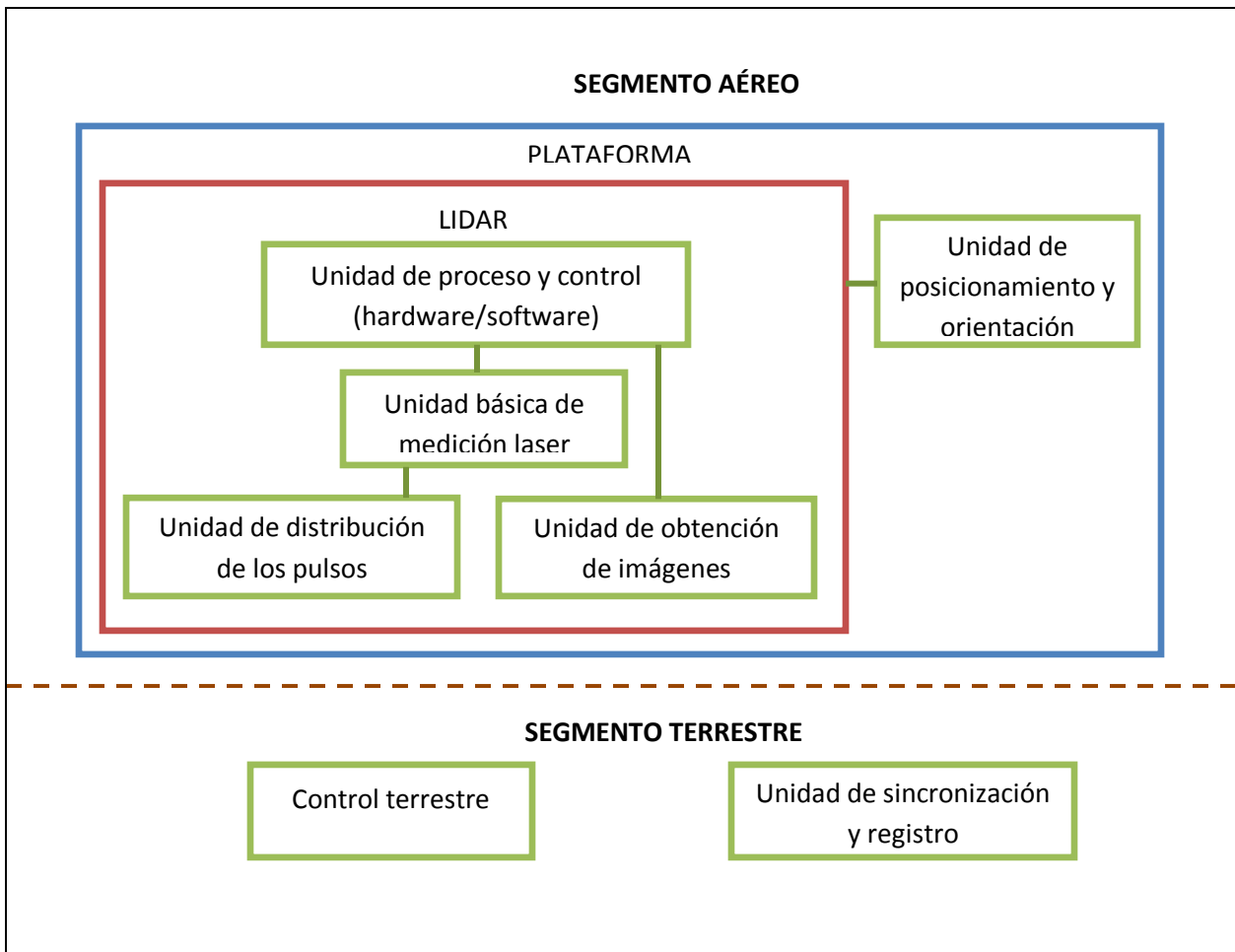


Figura 2.5. Diagrama de los componentes principales de un sistema lidar aerotransportado.

### 2.3.3.1 Unidad básica de medición laser

Está compuesta por tres elementos principales: el generador laser, el receptor y las piezas ópticas de transmisión y recepción.

-*Generador laser.* En el escaneo aéreo normalmente son utilizados láseres de estado sólido o semiconductor, los cuales son diseñados para producir pulsos cortos de alta potencia y con capacidad de repetición de emisión muy alto. El haz debe ser colimado (la colimación es la homogeneización de las trayectorias de los rayos, esto es, hacerlos paralelos) de forma que tenga la menor huella posible sobre el terreno, esta depende de la magnitud de las distancias a medir y de la divergencia del rayo, que a su vez es función de su longitud de onda. A altitudes de vuelo de 1000 m la huella típicamente es de 30 cm a 1 m de diámetro. Las longitudes de onda comerciales son de 900, 1064 y 1550 nm (Shan y Toth, 2009:136).

-*Receptor*. Compuesto por un receptor óptico, el detector y un amplificador. El detector utiliza un fotodiodo que convierte la señal óptica recibida en corriente eléctrica (Shan y Toth, 2009:136). Ya que esta señal no regresa con la misma forma que la de salida, es necesario establecer criterios de detección que apunten a disminuir los errores en las distancias (pues un pulso detectado al inicio del mismo o al final entregará diferentes mediciones). Al respecto existen varios métodos: por *umbral*, *centro de gravedad*, *máximo*, *fracción constante*, etc. Los más común y básico es detener el medidor del tiempo en el detector cuando la señal exceda cierto umbral de energía (*umbral*) (más detalles en Wagner et al. 2004:3).

-*Piezas ópticas de transmisión y recepción*. Son todos los lentes y espejos utilizados para conducir, dividir y filtrar el rayo laser en su camino de emisión y recepción dentro del sensor. El rayo es colimado por una serie de lentes que controlan su divergencia, el retorno disperso pasa a través de un filtro de interferencias producidas por la luz del sol y otras fuentes para eliminar el ruido de la señal que pudiera contribuir a distancias erróneas y después es enfocado por otros lentes al detector (Shan y Toth, 2009:138-139).

#### 2.3.3.2 Unidad de distribución de los pulsos

Dispositivo compuesto de elementos ópticos que distribuyen los pulsos laser en la dirección perpendicular de la de vuelo, generalmente espejos planos o poligonales, los cuales presentan movimientos oscilatorios o rotatorios de gran velocidad. En cada momento se conoce la posición de giro o ángulo del espejo con precisión, con objeto de determinar la dirección hacia la cual se emitió cada pulso.

Existen cuatro sistemas principales de distribución de los pulsos, de acuerdo al patrón que los puntos adoptan sobre la superficie terrestre:

-*Dentado*. Un par de espejos planos o uno sólo reflejan los puntos en la dirección perpendicular a la línea de vuelo, presentando un movimiento de ida y vuelta. Un codificador angular registra la posición del espejo a cada momento. Este mecanismo entrega un patrón dentado, en forma de "Z" o sinusoidal sobre el terreno, como lo muestra la figura 2.6. Para poder tener mayores densidades de puntos o poder hacer el vuelo a mayor velocidad el movimiento del espejo debe ser más rápido, esto implica disminuir la inercia (frenado del espejo para emprender el movimiento de regreso) con mayor rapidez (mayor potencia del sistema de escaneo). Naturalmente que entre menor tamaño del espejo menores esfuerzos inerciales, sin embargo, sus medidas están relacionadas con la dimensión de la apertura del sensor, la cual entre mayor sea mejorará la precisión de las mediciones laser para una longitud de onda constante.

Entonces el diseño del escáner debe optimizar por un lado la velocidad del movimiento del espejo y la precisión de las mediciones por el otro (Shan y Toth, 2009:44,145-147). Presenta la desventaja de que la distribución de los pulsos en el terreno es menos homogénea que en otros sistemas (Pfeiffer y Briese, 2007:312).

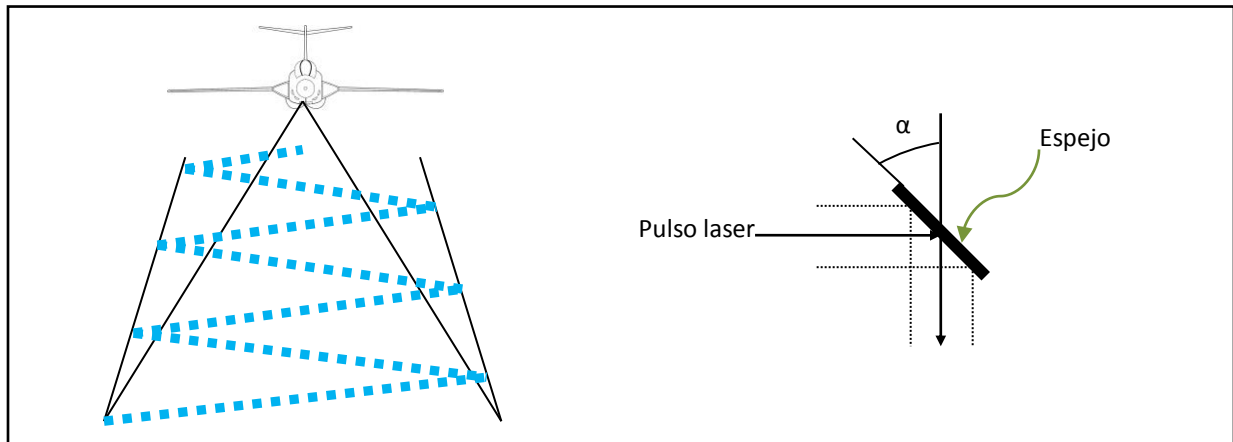


Figura 2.6. Ilustración del patrón de escaneo dentado. A la derecha un esquema del funcionamiento del espejo, el ángulo  $\alpha$  es el que se hace variar.

*-Líneas paralelas.* Un espejo poligonal en constante giro distribuye los pulsos en la dirección perpendicular a la línea de vuelo formando líneas paralelas sobre el terreno, como se ve en la figura 2.7. Presenta la ventaja de ofrecer un mejor control en el espaciamiento de los puntos, pues éste depende de la velocidad del espejo y de la de disparo de los pulsos; y mayor velocidad de escaneo, pues el espejo no experimenta desaceleración en su movimiento mientras trabaja. La oblicuidad de las líneas de escaneo con respecto a la línea de vuelo y el espaciamiento entre las mismas depende también de la velocidad del espejo, y la del avión (Shan y Toth, 2009:44-45,147).

*-Elíptico.* Conocido también como escáner de Palmer, consiste en un espejo colocado en la cara circular de un cilindro rotatorio; la normal a la superficie del espejo tiene una ligera desviación con respecto al eje del cilindro (en la figura 2.8 se puede ver un esquema de esta configuración, para mayor claridad), de tal forma que un haz electromagnético que incida sobre el espejo en movimiento se proyectará dibujando un elipse sobre la superficie del terreno. Esto produce una serie de elipses superpuestas como patrón de escaneo –como se observan en la figura 2.8– cuando se está sobre una plataforma móvil (avión) (Shan y Toth, 2009:45-46,147). Debido a esto, varios puntos en el terreno son escaneados dos veces, lo cual es favorable para la calibración del escáner y del sistema de posicionamiento y orientación (Wehr y Lohr, 1999:78).

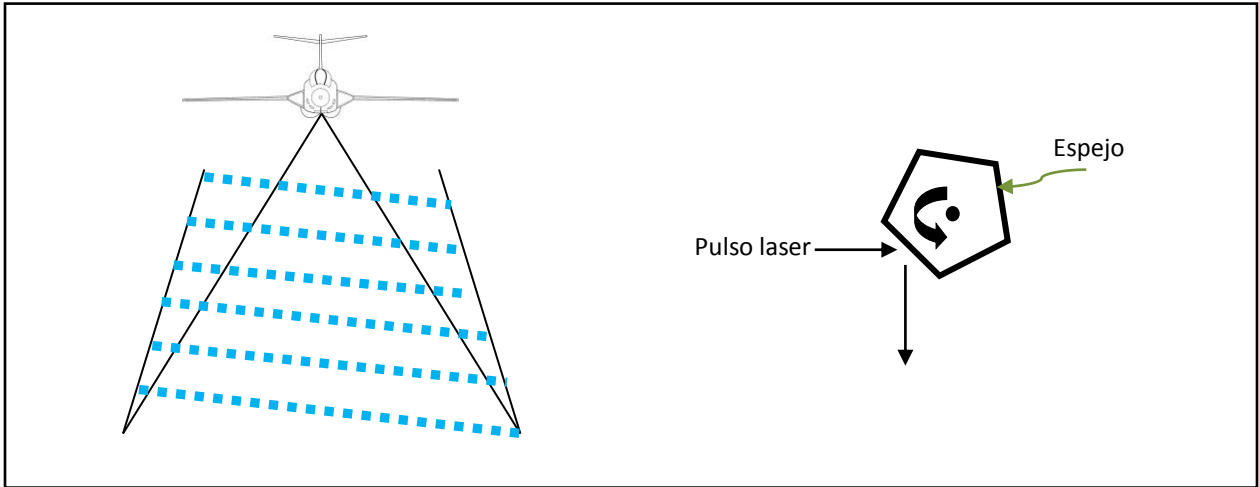


Figura 2.7. Ilustración del patrón de escaneo de líneas paralelas transversales a la dirección de vuelo. A la derecha un esquema del funcionamiento del espejo.

-*Líneas paralelas a la línea de vuelo.* En algunos sistemas se utiliza un par de arreglos de fibra óptica conjuntamente con un espejo giratorio para distribuir los pulsos. Esto genera un patrón de líneas de escaneo que corren paralelas a la dirección de vuelo, tal como se muestra en la figura 2.9. Presenta las ventajas de entregar una alta velocidad de escaneo (mayor que la de los otros sistemas) y es más estable y robusto en sus mediciones (Shan y Toth, 2009:45-47,148), esto debido en gran parte a que no es necesario medir ninguna posición angular de ningún espejo.

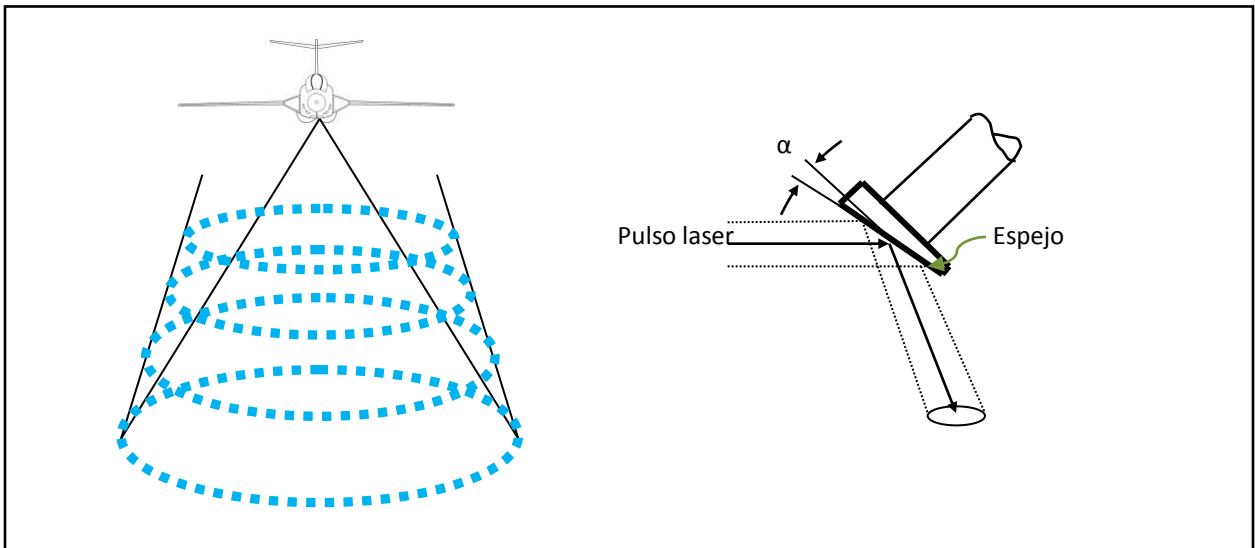


Figura 2.8. Ilustración del patrón de escaneo elíptico. A la derecha un esquema del espejo. El ángulo  $\alpha$  es la desviación de la normal al espejo con respecto al eje del cilindro.

Detalles acerca de las ventajas y desventajas de cada sistema son mencionadas por Jenkins (2006) en *Key drivers in determining lidar sensor selection* (Principales factores que determinan la selección del sensor lidar).

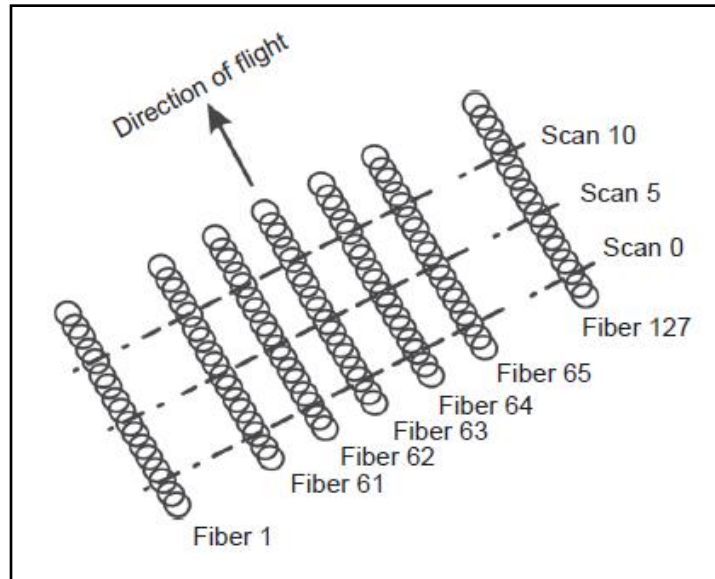


Figura 2.9. Ilustración del patrón de escaneo en líneas paralelas en la misma dirección de vuelo. Fuente: Shan y Toth, 2009:150.

Entre mayor sea la densidad de puntos mejor se caracteriza la superficie terrestre y los elementos sobre ella, esto requiere más recursos económicos y de tiempo, mejor equipo y mayores demandas en cuanto a procesamiento; por esto debe determinarse en función de la aplicación en la cual se vaya a utilizar la nube. La densidad varía en función de un gran número de factores: altura de vuelo, velocidad y variación de la aceleración de la plataforma; ángulo de escaneo, ritmo de disparo de los pulsos y patrón de escaneo del sensor; y de la geometría y reflectividad del terreno o de los elementos sobre él (Lohani, 2008:23).

### 2.3.3.3 Unidad de proceso y control (hardware/software)

Unidad electrónica-computacional, equipada con una pantalla y la interface necesaria para manipular todo el sistema lidar. Básicamente se encarga de la sincronización de la unidad de medición con la de distribución de los pulsos, disparando los haces laser en concordancia con el movimiento de los espejos de tal forma que se distribuyan los pulsos como se desea; y por otro lado, recolecta y guarda los datos en una unidad de almacenamiento, generalmente un disco duro, tales como la forma de la onda de retorno, intensidad, ángulos de la posición del espejo, tiempos de retorno de los pulsos, mediciones de distancia

y la información de posición del GPS y la IMU. Cabe mencionar que se requiere de un dispositivo de manipulación de los datos muy poderoso, pues es necesario procesar y hacer circular una gran cantidad de ellos en el sistema en lapsos muy cortos de tiempo (Shan y Toth, 2009:41,150).

Respecto al software utilizado, este tiene varios propósitos a lo largo de todo el proceso de obtención de datos lidar. Primeramente se requiere un módulo para el diseño del plan de vuelo y la implementación de los recorridos correctos al llevarse a cabo, que permita al piloto realizar los ajustes necesarios de dirección en el momento oportuno en conjunción con la información recibida por el GPS. Se requiere otro módulo para configurar los parámetros del sensor, tales como los ritmos de escaneo y de disparo de los pulsos, y el ángulo de movimiento de los espejos. Uno más para el proceso de recolección y almacenamiento. Y finalmente otro más para visualizar los datos y realizar procesamientos preliminares en vuelo, y para llevar a cabo el postprocesamiento en tierra (Shan y Toth, 2009:42).

#### 2.3.3.4 Unidad de posicionamiento y orientación

Conformada por un receptor GPS y una Unidad de Mediciones Inerciales (IMU, del inglés *Inertial Measurement Unit*); entre ambos recogen observaciones continuamente durante el vuelo, el primero las posiciones y la segunda las aceleraciones lineales y angulares del avión, así como los ángulos de sus desviaciones con respecto al plano horizontal y la dirección. El receptor GPS necesita de apoyo terrestre (otro receptor) con objeto de alcanzar las precisiones requeridas; el IMU debe colocarse convenientemente lo más cerca posible del escáner. Es necesario determinar también las posiciones relativas de la unidad de medición, el escáner, la IMU y el GPS entre sí. Toda esta información es procesada posteriormente para obtener la ubicación y altitud del escáner, y poder derivar las coordenadas de todos los puntos de la nube lidar (Shan y Toth, 2009:41,151-152).

#### 2.3.3.5 Unidad de obtención de imágenes

Generalmente se considera algún dispositivo de obtención de imágenes como parte de un sistema lidar aerotransportado integral, en muchos casos una cámara fotográfica digital o algunas veces una video cámara. Va empotrada conjuntamente con la unidad de medición y su funcionamiento está sincronizado con el mecanismo de escaneo (Shan y Toth, 2009:43).

### 2.3.3.6 Control terrestre

Conociendo el área que se cubrirá en el levantamiento y los requerimientos del proyecto se definen las estaciones base de apoyo y los campos de control, con el objeto de poder obtener la ubicación con la precisión requerida de la aeronave y el sensor, de controlar la trayectoria y hacer las correcciones necesarias de los puntos en el posprocesamiento.

Las estaciones base son puntos elegidos estratégicamente en las zonas del levantamiento o cercanas a él, los cuales se dotan de coordenadas de acuerdo al marco geodésico de referencia del lugar. En estos puntos se coloca un receptor GPS cuya labor es proveer de correcciones al GPS de la aeronave, en vuelo, y por añadidura a los puntos, en el posprocesamiento. Para obtener posiciones con precisiones de centímetros, y utilizando los equipos actuales y los procesos de posprocesamiento comerciales, las estaciones base no deben estar más alejadas de 25 km del lugar de vuelo (Shan y Toth, 2009:151).

### 2.3.3.7 Unidad de sincronización y registro

El procesamiento que es preciso hacerle a los datos después del vuelo para dotar a los puntos de coordenadas geodésicas también se considera parte del sistema lidar. Después del vuelo se tienen datos de posición del GPS, desviaciones del avión en todo su recorrido (alabeo, cabeceo y deriva) y las variables recolectadas de los puntos (tiempos, distancias, intensidades, etc), que son necesarios trabajar en gabinete para calcular tales coordenadas. Esta unidad se refiere al hardware y software imprescindible para llevar a cabo las tareas de sincronización y registro.

La sincronización se refiere al acoplamiento en tiempo que debe hacerse de los datos recogidos del GPS, la IMU y el sensor lidar. Estos elementos son construidos por separado por diferentes compañías, y trabajan de forma independiente; aunque el GPS y la IMU entregan sus observaciones de forma sincronizada, esto no ocurre con la unidad de medición y el escáner, que tienen un contador de tiempo propio. El acoplamiento se logra mediante la generación en vuelo de un archivo de protocolos que registra los tiempos del reloj GPS y del sensor a cada cierto lapso, que luego será utilizado en posproceso para la sincronización. Normalmente, las observaciones del sensor son hechas en períodos de tiempo más cortos que las del GPS/IMU, pues se asienta un registro por cada recepción de cada pulso; esto requiere la interpolación de las posiciones y movimientos angulares obtenidas por el GPS/IMU (Shan y Toth, 2009:152-153).

El registro se refiere al proceso que hay que llevar a cabo para propagar las coordenadas geodésicas obtenidas por el GPS a cada uno de los puntos obtenidos por el sensor. Básicamente se representa como (figura 2.10):

$$\vec{t} = \vec{s} + \vec{m} \quad (2.6)$$

Donde:

$\vec{t}$  es el vector que va del origen de coordenadas geodésicas (centro de la Tierra) al punto lidar sobre la superficie,

$\vec{s}$  es el vector que va del origen de coordenadas geodésicas al punto origen de medición de la distancia (punto de reflexión en el espejo del escáner), y

$\vec{m}$  es el vector de medición de la distancia del sensor al punto sobre la superficie.

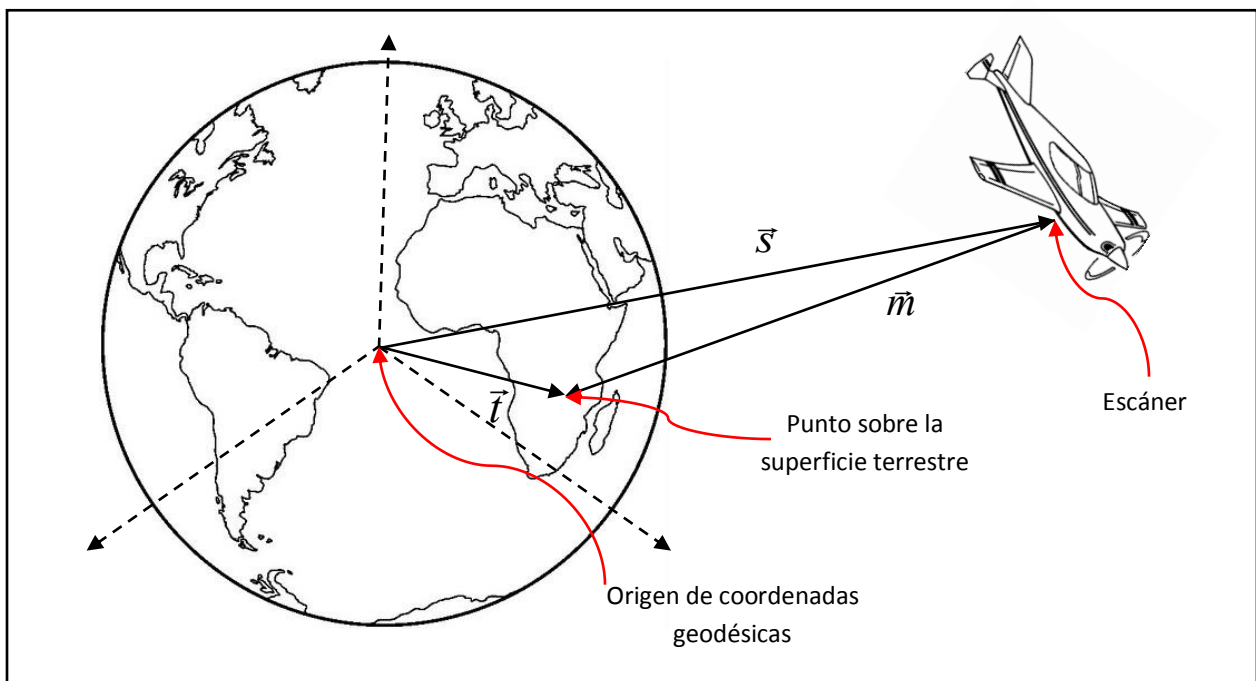


Figura 2.10. Ilustración de la suma de vectores necesaria para lograr el registro de una nube de puntos.

El origen del vector  $\vec{m}$  es el punto donde el pulso laser colisiona con el espejo, y es necesario conocer la distancia de este punto al GPS para derivar sus coordenadas. Ya que el vector  $\vec{s}$  está en el sistema WGS84 y por conveniencia se entregan las coordenadas de los puntos en este sistema, es necesario llevar a cabo una serie de transformaciones del vector  $\vec{m}$  que se recoge en el sistema de coordenadas del avión: el eje x es la dirección de vuelo, el y es a la derecha del avión en la dirección de la línea que une las puntas de las alas, y el z es hacia el nadir. Primeramente debe corregirse la posición del escáner con



respecto a la de la IMU, pues es esta la que registra las desviaciones del avión; una segunda transformación es la del IMU/escáner con respecto al sistema horizontal, esto es, las correcciones de las desviaciones en alabeo, cabeceo y deriva; y por último, la transformación de este sistema horizontal al sistema geodésico WGS84 (Shan y Toth, 2009:154-156).

#### 2.3.4 Corrección del error de altura sistemático de la nube de puntos

Para tener un control de calidad de las alturas obtenidas de la nube de puntos, sobre el terreno se ubican una serie de campos de control, de los cuales se levantan sus alturas para compararlas con las alturas de los puntos en esa zona y determinar si es que existe alguna diferencia sistemática y poder corregirla. Los campos de control son áreas planas, despejadas y de forma regular, que se delimitan sobre el terreno o, en otros casos, superficies planas construidas artificialmente que se fijan en el terreno; de estos campos se determina su ubicación y su altura.

La forma de corregir el error en altura que se realiza en los trabajos lidar del INEGI (Instituto Nacional de Estadística, Geografía e Informática) es de la siguiente forma: ya que se realiza un escaneo transversal en toda la zona de trabajo, en cada extremo de esta franja se ubica un campo de control, lo que permite hacer la corrección en esa zona, y a partir de estas se interpola la corrección en toda la franja; para posteriormente obtener la diferencia en altura de cada franja restante con respecto a la transversal en su punto de intersección y aplicar esta corrección a toda esa franja completa (Arellano et al., 2005).

#### 2.3.5 Divergencia del rayo y huella

La divergencia es una medida de la magnitud del incremento de la sección del haz laser al viajar a través del espacio, observada en dimensiones angulares. La huella es el área cubierta del objetivo (superficie terrestre, por ejemplo) por el haz, siempre circular o elíptica sobre una superficie lisa, en función del grado de incidencia del rayo. En la práctica casi siempre las superficies son irregulares o con cierto grado de rugosidad, entonces la señal de retorno es la suma de todas las reflexiones ocurridas dentro de la huella.

El tamaño de la huella es función del ángulo de divergencia del rayo, del diámetro de apertura del emisor laser y de la distancia del sensor al objetivo; mientras mayores sean estos parámetros, mayor será la huella. En un vuelo a una altura de 1000 m y con divergencias de 0.3 y 0.5 mrad se generan huellas de 30 y 50 cm respectivamente (Shan y Toth, 2009:22).

Existen equipos con la capacidad de regular la divergencia de los pulsos emitidos. Aquellos con alta divergencia (0.8 mrad, por ejemplo) tienen aplicaciones para trabajos de baja altitud, pues proporcionan mayores retornos en superficies pequeñas, como por ejemplo el levantamiento de líneas de transmisión; algunos de estos sistemas también tienen la opción de reducir la potencia de los haces, con objeto de evitar posibles daños a las personas en tierra. Los equipos de poca divergencia (0.2 mrad, por ejemplo) proveen mejor precisión en la posición, y tienen aplicaciones en tareas donde se requiere mayor penetración, como por ejemplo mediciones forestales (Jenkins, 2006:13-14).

### 2.3.6 Reflectividad

La reflectividad es también un factor a considerar para el desempeño de un sistema lidar, pues las máximas distancias que se pueden medir también están en función del grado en que las superficies reflejan la energía del laser. Ésta es definida como el radio de la radiación incidente que es reflejada por una superficie particular (Shan y Toth, 2009:22), y depende de dos factores: el tipo de superficie y el ángulo de incidencia de la radiación.

La gran mayoría de las superficies encontradas sobre el terreno tienen algún grado de rugosidad y por lo tanto presentan un patrón de reflexión difusa, idealizado como una esfera (reflexión en todas direcciones) con centro en el punto de incidencia del haz, aunque en realidad la reflexión no es uniforme y puede ser que en alguna dirección no se presente en absoluto, y aún más, las propiedades reflectivas también varían en función de la longitud de onda. El porcentaje de reflectividad de los árboles caducifolios y de las coníferas es típicamente del 60 y 30 % respectivamente para longitudes de onda de 900 nm (Wehr y Lohr, 1999:74).

Con respecto al ángulo de incidencia, si este no es cercano a los  $0^\circ$ , y especialmente hablando de superficies poco rugosas, casi toda la señal es reflejada a alguna dirección diferente a la de la posición de la aeronave y muy difícilmente se registrará un retorno. En el caso del dosel forestal por ejemplo, es bastante conveniente que los pulsos penetren hasta la superficie del suelo, y esto sólo ocurre cuando los haces presentan bajos ángulos de incidencia, mientras se incrementa éste la habilidad de penetración decaerá.

### 2.3.7 Procesamiento de los datos

Inmediatamente después de que ya se tienen los datos lidar georeferidos y corregidos se empiezan a procesar para diferentes aplicaciones y productos finales. Dentro de estos procesos existen una serie de manipulaciones genéricas, no vinculadas a aplicaciones específicas, que se utilizan para explorar los datos (visualizaciones), extraer información de la nube (segmentaciones, filtrados, agrupamientos) o prepararlos para ello (transformaciones).

#### 2.3.7.1 Visualización

El deseo más primitivo que se tiene al iniciar a ocuparse de la nube es mirar cómo luce, la visualización es el tratamiento más básico que se puede hacer de ella. Su función es poder tener una idea de la calidad del levantamiento, mostrar los datos en diferentes etapas del procesamiento general para planearlo y controlarlo y poder, finalmente, ver el producto terminado (Fernández et al., 2007:5).

Fundamentalmente la visualización va desde poder presentar los puntos en un solo color y tamaño, pasando por las operaciones elementales de acercamiento, rotaciones y navegación sobre la nube y hasta, incluso, poder colorear los puntos en función de su intensidad, elevación, clase, etc. Comandos más avanzados incluyen la selección y manejo de grupos de puntos o individuales, la medición de distancias y ángulos entre líneas de puntos, el ajuste de primitivas y la generación de vistas de perfil (Fernández et al., 2007:6-7).

#### 2.3.7.2 Transformaciones

Son operaciones que implican un cambio en la ubicación de los datos dentro de su marco de referencia. Las rotaciones y traslaciones básicamente son el cambio de coordenadas de un grupo de puntos sin que pierdan la posición relativa entre ellos. En una traslación a todos los puntos se les aplica el mismo cambio, mientras que en una rotación sufren desplazamientos desiguales, lo cual implica un cambio de orientación de la nube. Un recorte es la separación de un grupo de puntos de otros, con objeto de eliminar unos, crear grupos de una determinada zona de interés o moverlos. La fusión es la integración y correcta ubicación relativa de un grupo de puntos con otro, que posiblemente hayan sido recolectados en diferente tiempo o diferentes condiciones.

### 2.3.7.3 Segmentación y filtrado

La segmentación es un proceso de organización de puntos alcanzado por medio de la delimitación de grupos que presenten características homogéneas. Existen diversos métodos para ello, dos de los más conocidos y utilizados son el de crecimiento de región y el de agrupación. En el primero se parte de un punto modelo llamado semilla y a partir de la similitud entre este y el candidato a ser agregado de entre sus vecinos se crea el segmento, al cual se van agregando sucesivamente más puntos mientras sea posible. El criterio de similitud puede estar basado en la diferencia de alturas, el parecido del vector normal, la distancia a un plano, etc; si se compara siempre la semilla con cada punto candidato o si se compara el último punto agregado con el candidato, se pueden construir segmentos curvados o no (Pfeiffer y Briese, 2007:316). En los métodos de agrupación se conforman colecciones de puntos de acuerdo a sus características (por ejemplo intensidad, densidad de su vecindario, etc.) por medio de la maximización del contraste en las diferencias entre ellos.

La clasificación implica la dotación de un significado a cada punto, que puede ser lo que representa en la realidad: suelo desnudo, vegetación baja, dosel, edificación, etc., o con alguna otra estructura: retorno sencillo, primer retorno de varios, último retorno de varios, segundo retorno, etc.; ó punto sin vecinos cercanos, con pocos vecinos, muchos vecinos, etc., por ejemplo.

El filtrado es la remoción de un conjunto de puntos de la nube en base a algún esquema de segmentación o clasificación (Fernández et al., 2007:8), y tiene su aplicación más común en la generación de datos de sólo suelo o sólo no-suelo, aunque también podría pensarse en la eliminación de puntos por debajo de cierta altura, vegetación de una escena citadina, o incluso puntos de calles en la preservación de sólo puntos de edificios. El adelgazamiento no es más que un proceso de filtrado, en este caso dirigido a disminuir la cantidad de puntos con la menor pérdida de información posible, con objeto de poder hacer a los datos más fáciles de manipular y procesar.

### 2.3.7.4 Gridding

La transformación de los datos lidar a malla o cuadrícula, también conocido como gridding, consiste en el proceso de convertir un conjunto de datos (puntos) irregularmente distribuido en una malla rectangular regular e igualmente espaciada llamada ráster. Por lo general el parámetro que se utiliza para construirla es la elevación, aunque también puede utilizarse otro, como la intensidad. La magnitud

del espaciamiento entre celdas contiguas es lo que se conoce como resolución, normalmente es de 1 m si se utilizan datos lidar de alta densidad (Starek y Slatton, 2005:2).

Para determinar qué valor debe tomar cada celda –pues en muchos casos caen varios datos dentro de una misma celda o no cae ninguno, en cuyos casos es necesario decidir qué valor asignarle– se utilizan procedimientos de interpolación espacial, como son el lineal, del inverso de la distancia, vecino más cercano, etc.

La naturaleza regular de un grid permite llevar a cabo análisis que son de gran utilidad, como el cálculo de áreas y volúmenes, comparación y álgebra de superficies, gradientes de diferenciación y todos los algoritmos propios del procesamiento digital de imágenes (Fernández et al., 2007:10).

Las ventajas que presenta el operar los datos en grid es una mayor facilidad de visualización, la relativa simplicidad en la aplicación de procesamientos y la posibilidad de hacer análisis de frecuencias. En contraparte, presenta las desventajas de tener el riesgo de una posible pérdida de puntos críticos en las zonas en donde haya sido necesario reducir datos, y un desperdicio en el espacio de almacenamiento al representar grandes extensiones de superficie plana –si es que existen– mediante celdas (Starek y Slatton, 2005:3).

### 2.3.8 Aplicaciones

Durante los últimos años, con el mejoramiento de los sistemas computacionales, se han podido desarrollar un sin número de programas de cómputo y nuevos esquemas de procesamiento de los datos lidar, lo que ha desembocado en la implementación de una gran variedad de aplicaciones en diferentes ámbitos del quehacer humano. Sin embargo, antes de poder especializar la manipulación de los datos para un objetivo específico, existen un par de productos básicos utilizados como insumos en muchas ocasiones: el Modelo Digital de Superficie (DSM, del inglés Digital Surface Model) y la imagen de intensidad.

Un DSM es una cuadrícula (grid) cuyos valores de altura son obtenidos a partir de la nube de puntos. Las ventajas de trabajar con superficies ya se han mencionado anteriormente. Aquí se distinguen dos casos de construcción de DSM de gran uso:

-*Modelo Digital de Terreno (DTM, del inglés Digital Terrain Model)*. Superficie construida a partir de sólo aquellos puntos que se suponen reflejados por el terreno, dejando fuera retornos de infraestructura, vegetación y objetos aéreos.

-*Modelo Digital de Dosel (DCM, del inglés Digital Canopy Model)*. Superficie construida a partir de los puntos reflejados por la vegetación, especialmente la copa de los árboles en sus partes altas y más exteriores. El DCM es socorrido en aplicaciones forestales para estimar biomasa, volumen y segmentar copas.

La imagen de intensidad es una imagen raster monocromática obtenida de la nube lidar, en donde el valor de gris de cada pixel está en función de la intensidad de todos los puntos incluidos en el área que cubre; en aquellos en donde no existen puntos, el valor es interpolado a partir de sus vecinos. Este atributo, la intensidad, puede ser usado para mejorar los algoritmos de filtrado, remoción, clasificación y separación de objetos en combinación con la información geométrica (Wehr y Lohr, 1999:81).

En específico existen un sinnúmero de aplicaciones que utilizan lidar, muchas de ellas estrechamente relacionadas con la topografía. A continuación se mencionan las más comunes, con objeto de tener idea de las tareas en las que puede ser aplicado y ofrecer un contexto que nos permita advertir que, aunque esta tecnología se ha utilizado ampliamente en trabajos forestales, no es la única aplicación y también se han investigado con otras.

#### 2.3.8.1 Aplicaciones en la administración de recursos naturales y desastres

En cuanto a la administración de recursos naturales, y al ser lidar un sensor remoto, es aprovechado en tareas de mapeo y obtención de datos de áreas de difícil acceso y/o para hacerlo con mayor rapidez, por ejemplo en la obtención de variables forestales, en el mapeo de minas (estimación de volúmenes de tierras y grandes concentraciones de materiales, monitoreo de subsidencias, planeación de operaciones de explotación, etc.) y adquisición de datos de glaciares (conocimiento de su topografía, monitoreo del movimiento y velocidad de los hielos y de las acumulaciones de nieve). Respecto a las aplicaciones en desastres se utiliza en la obtención de las estructuras tridimensionales presentes en la superficie terrestre para la generación y mejoramiento de modelos de inundaciones, el monitoreo de deslizamientos y cambios de pendientes y en la producción de mapas de riesgos de las propias inundaciones, de incendios y de avalanchas, así como para mapear los daños después de un evento. También se aprovecha en batimetría para el mapeo de ríos, canales y costas, la construcción de modelos de flujos de las aguas, la obtención de la topografía de los fondos, el conocimiento de la dinámica de los

sedimentos y el monitoreo de la erosión (Lohani, 2008:27-31; BC-CARMS, 2006:19-22; Wehr y Lohr, 1999:81).

#### 2.3.8.2 Aplicaciones en áreas urbanas

Lidar se utiliza en la generación de mapas de áreas metropolitanas en general, en muchos casos para la planeación de proyectos de nuevas urbanizaciones e infraestructura. Con la densidad de puntos suficiente se puede conseguir, además de la localización, la delimitación de los edificios y e incluso su geometría, pudiéndose emplear en el mapeo de construcciones, estimación de población e ingresos, estudios de detección de cambio, planeación de rutas de emergencia, etc. La habilidad de lidar para recolectar datos en corredores estrechos, sombríos y calles muy transitadas le proporciona ventajas sobre otros métodos de levantamientos y permite el trabajo de estas aplicaciones de una forma más rápida (Lohani, 2008:27-31; BC-CARMS, 2006:19-22; Wehr y Lohr, 1999:81).

#### 2.3.8.3 Aplicaciones en ingeniería civil no urbana

Los datos lidar son auxiliares en el diseño y construcción de centrales hidroeléctricas y presas. Así mismo, en la extracción de información de las redes viales existentes para alimentar las bases de datos y planear los trabajos de mantenimiento y reparaciones, todo esto sin interrupción de los servicios. También se utilizan en las líneas de transmisión eléctrica, tuberías, ferrocarriles y vías rápidas de autotransporte, tanto en el diseño del proyecto y planeación de su construcción como en el monitoreo de deformaciones y planeación de mantenimiento, reparaciones y modificaciones (Lohani, 2008:27-31; BC-CARMS, 2006:19-22; Wehr y Lohr, 1999:81).

#### 2.3.9 Ventajas y desventajas de utilizar lidar

Como todas las tecnologías, lidar presenta una serie de atributos que son más provechosos de utilizar que empleando otras alternativas de percepción remota:

-Puede ofrecer datos de elevación de mejor precisión que otras opciones, a las alturas en las cuales se obtienen las imágenes para fotogrametría (alrededor de 700 m sobre el terreno o mayores). Esta precisión, tanto en las imágenes aéreas como en lidar, disminuye a medida que aumenta la altura de adquisición, sin embargo, se degrada en menor medida en este último (BC-CARMS, 2006:22).

-Para poder obtener la altura de un punto en una imagen es necesario tener dos puntos de vista de éste, es decir, verlo desde dos direcciones diferentes; en el caso de lidar es suficiente que el punto pueda verse desde una sola para poder conseguir su elevación (BC-CARMS, 2006:22).

-Pueden hacerse escaneos también en la noche, pues lidar no depende de la iluminación solar para coleccionar mediciones (BC-CARMS, 2006:23).

-Es una tecnología muy versátil que ofrece aplicaciones en diversas áreas: batimetría, topografía, estudios atmosféricos, investigaciones glaciales, etc (Burtch, 2002:11).

-Ofrece una alta precisión en la generación de Modelos Digitales de Terreno y una buena densidad de mediciones (Burtch, 2002:11). Típicamente la precisión horizontal es de 30 a 50 cm y la vertical de 5 a 15 cm, y se pueden conseguir más de 24 puntos por metro cuadrado (Lohani, 2008:31-32).

-Especialmente en trabajos forestales lidar ofrece la gran ventaja de poder penetrar a través de la copa de los árboles y así poder contar con datos de los cuales poder derivar la superficie terrestre, lo cual es extremadamente difícil de conseguir con otras tecnologías de percepción remota (Burtch, 2002:11).

Por el contrario, las principales desventajas de lidar son las siguientes:

-Comúnmente los pulsos del laser utilizado en levantamientos de la superficie terrestre son absorbidos por el agua, lo que hace a los retornos de estas áreas poco confiables. Esto hace difícil determinar los límites de cuerpos de agua en los datos lidar (BC-CARMS, 2006:23).

-Prestando atención específica a nubes de poca densidad, lidar en muchos casos ofrece una pobre detección de los cambios bruscos de pendientes, difícilmente se tendrán puntos exactamente o muy cercanos a los puntos de quiebre (BC-CARMS, 2006:23).

-Debido al relativo poco tiempo de haberse desarrollado el lidar, aún no existen una serie de procedimientos de levantamiento estándar aplicables que permitan prever las precisiones que se obtendrán en determinadas situaciones y con equipos específicos (BC-CARMS, 2006:23).

-Adquirir un equipo lidar tiene un alto costo, pues además del precio inicial que hay que pagar, es probable que el equipo y sus complementos sufran innovaciones tecnológica en el corto plazo, lo cual los dejaría, en cierta forma, obsoletos en poco tiempo (Burtch, 2002:11).



### 3. METODOLOGÍA

#### 3.1 INTRODUCCIÓN

En la presente tesis se pretende construir un modelo de estimación de la cantidad de carbono presente en el Suelo de Conservación del Distrito Federal utilizando un método alternativo al tradicional de muestreo de campo: la utilización de datos lidar aerotransportado; y, ya que los datos lo permiten, también ensayar modelos de estimación de número de árboles, altura media de los mismos y diámetro normal.

En 2010 la Procuraduría Ambiental y de Ordenamiento Territorial del Distrito Federal (PAOT) en coordinación con otras instituciones realizaron el estudio *Estimación de captura de carbono como indicador del estatus del derecho de los habitantes del Distrito Federal a gozar de áreas verdes urbanas adecuadas para su desarrollo, salud y bienestar* (PAOT, 2010), en el cual calcularon la cantidad de carbono en el Suelo de Conservación levantando datos en campo de los árboles de un conjunto de sitios o muestras. El hecho de contar con estos datos y con un levantamiento lidar de estas zonas representa una oportunidad para probar métodos de predicción, así como para cuantificar su exactitud y validarlos.

En lo referente a la estimación de parámetros forestales mediante datos lidar existen dos grandes propuestas, los métodos basados en la identificación de árboles individuales y los basados en la distribución de la nube de puntos. En el presente trabajo se ensayan ambas, cada una tiene sus propias peculiaridades y se observa el comportamiento con cada una de ellas con los mismos datos lidar. Con respecto a los métodos por identificación de árboles individuales, estos requieren nubes de buenas densidades de puntos para obtener resultados satisfactorios, lo cual aumenta el costo de obtención de los datos, y se requieren procesos más especializados y complejos en su tratamiento, aunque con las ventajas de que se necesitan relativamente pocas muestras de campo para la calibración de los modelos y que puede llegar a entregar una mayor cantidad de información forestal con más precisión y resolución, al obtenerla por individuo y no por área. En cuanto a los métodos de estimación por distribución de las alturas del dosel, no requieren nubes tan densas, lo cual aminora el costo de adquisición; los procesos de tratamiento de los datos tienen un alto contenido de conocimientos estadísticos ya validados, generalmente no es necesaria mucha innovación en este sentido, sólo aplicar la teoría; y son más fáciles de integrar con los inventarios forestales, pues se pueden trabajar utilizando como unidad de área la misma que la del inventario. La desventaja más importante que presentan es

que para obtener resultados confiables se requieren datos de campo relativamente extensos, precisos, representativos y caros para la creación y calibración de los modelos (Shan y Toth, 2009:352).

Una de las desventajas al transformar los datos lidar a superficie es la degradación de la información. Una nube de puntos contiene información tridimensional de los objetos sobre la superficie terrestre que, en función de la precisión del sensor y los parámetros de vuelo, se ajusta lo más posible a las posiciones y formas reales de estos objetos, información *ad hoc* para la derivación de información estructural. En cambio, para generar el modelo de superficie, los datos de alturas se ajustan a una malla regular cuadrangular (píxeles), en donde, si se tiene varios puntos dentro de un píxel, sólo se toma un valor de altura (el mayor, o el promedio), y si se tiene que no hay puntos en un píxel, el valor que toma éste se obtiene de interpolar a partir de los valores adyacentes. En una superficie, los objetos que se pueden apreciar están en función de la resolución espacial (tamaño de píxel), en los datos lidar la única restricción es la precisión en la obtención de la posición de los puntos. Por esto, se opta por efectuar la segmentación sobre la nube de puntos, y no alterar los datos originales.

### 3.2 DESCRIPCIÓN DEL ÁREA DE ESTUDIO

En el Distrito Federal, para su administración el suelo se divide en dos tipos: suelo urbano y suelo de conservación, siendo este último el que contiene las áreas forestales y objeto de estudio del presente documento.

El Suelo de Conservación del Distrito Federal tiene un área de 87 294.36 ha y en él habitan 2 200 000 personas, de las cuales alrededor de 700 000 tienen una relación directa con él. Cabe mencionar que muchos de ellos poseen aún conocimientos acerca del aprovechamiento y manejo de los recursos naturales del lugar. Presenta un clima templado sub húmedo con lluvias en verano y temperatura media anual de 16 °C en las partes de baja altitud, hasta climas semifríos con alto contenido de humedad y abundantes lluvias en verano, y temperaturas de hasta -8 °C en las partes altas. Los principales tipos de vegetación presentes son: pastizal o zacatonal, bosque de coníferas, bosque mixto, bosque de encino, bosque mesófilo<sup>1</sup>, matorral xerófilo<sup>2</sup> y vegetación acuática y subacuática (zona lacustre en Xochimilco y

---

<sup>1</sup> Caracterizado por una densa cubierta de árboles con numerosos géneros, frecuentemente incluyendo pinos y encinos. Exuberante, con plantas que crecen sobre los árboles, helechos y lianas. Se desarrolla en altitudes donde se forman bancos de niebla. (SEMARNAT, 2005).

<sup>2</sup> Vegetación dominada por arbustos, típica de zonas áridas y semiáridas. Es el tipo de vegetación más extenso en México. . (SEMARNAT, 2005).

Tláhuac). En general es una región de gran valor para la conservación de la biodiversidad, pues en ella viven más de 2 500 especies de flora y fauna y para la sobrevivencia de la población, tanto de la que habita en ella, como la de la zona metropolitana. (SMA, 2007:19-21).

Con respecto a los servicios ambientales que presenta y su importancia, la *Agenda Ambiental de la Ciudad de México 2007-2012*, elaborada por la Secretaría del Medio Ambiente del Distrito Federal (SMA, 2007:21), señala:

“Entre éstos sobresalen, en primer término, la recarga del acuífero (en el SC se capta la mayor parte del agua que se recarga a los mantos acuíferos del Valle de México), la captación de partículas suspendidas, la producción de oxígeno y la captura de carbono (indispensable para contener y mitigar el impacto del cambio climático), la regulación climática y microclimática, la mitigación de la contaminación auditiva, la retención de suelo fértil, así como su función como espacios de esparcimiento turísticos y culturales”.

Para la estimación de carbono en el Suelo de Conservación, se consideraron 50 áreas, de las cuales se levantaron datos de campo de la vegetación, como también de las que se cuenta con la nube de puntos lidar.

### 3.3 DATOS

#### 3.3.1 Datos de campo

En el año 2010 la Procuraduría Ambiental y de Ordenamiento Territorial del Distrito Federal (PAOT), en coordinación con la Secretaría del Medio Ambiente del Distrito Federal (SMA), la Comisión de Recursos Naturales del Distrito Federal (CORENA) y el Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) realizaron el estudio *Estimación de captura de carbono como indicador del estatus del derecho de los habitantes del Distrito Federal a gozar de áreas verdes urbanas adecuadas para su desarrollo, salud y bienestar* (PAOT, 2010), con el objeto de estimar la cantidad de carbono secuestrada en el Suelo de Conservación del Distrito Federal. Para ello se utilizaron una serie de muestras tomadas en campo, llamadas conglomerados, de las cuales se calculó la cantidad de carbono capturado, para posteriormente extrapolar y generar mapas regionales de esta variable.

Dentro del estudio se utilizaron 299 conglomerados, de los cuales 50 fueron levantados en los meses de enero a marzo de 2010, como complemento a los datos que se tenían originalmente (249 conglomerados). La información de campo de estos últimos 50 es la que se utiliza para el desarrollo del

presente trabajo, y fue levantada por personal de INIFAP, PAOT, SMA y CORENA, coordinados por la INIFAP.

El muestreo por conglomerados se hizo de acuerdo al diseño del Inventario Nacional Forestal y de Suelos de la Comisión Nacional Forestal (CONAFOR), en el cual se define cada conglomerado por cuatro áreas circulares llamadas sitios, dispuestos en forma de “Y” invertida (figura 3.1), uno central y los otros tres alrededor en los azimuts 0°, 120° y 240°, e identificando a cada uno de ellos por un número (sitio 1, ... .. ,sitio 4). Cada sitio tiene un área de 400 m<sup>2</sup>, radio de 11.28 m y los sitios 2 al 4 están a una distancia con respecto al 1 de 45.14 m. (CONAFOR, S/F:14-16).

Dentro de cada sitio se levantaron los datos de todos los árboles cuyo diámetro a la altura de pecho fue mayor a 7.5 cm. La información recabada fue la siguiente:

--Por cada árbol:

- Nombre científico y común.
- Diámetro normal (a la altura de pecho) (cm).
- Altura total (m).

--Por cada sitio:

- Fecha.
- Número de conglomerado.
- Número de sitio.
- Tipo de cobertura vegetal.
- Coordenadas geográficas (En la proyección UTM Zona 14, Datum WGS 84).

Cabe aclarar que en adelante se mencionan sitios que presentan tipo de cobertura vegetal pastizal y matorral; cuando se hace referencia a las estimaciones hechas en esos sitios, estas son sólo de los árboles presentes en el área, y no precisamente de la cobertura no arbórea a la que hace referencia su nombre. En varios casos también se recopiló otra información complementaria, como edad y condición de cada árbol, y pendiente y altura sobre el nivel mar por cada sitio, la cual no fue utilizada por el presente trabajo.

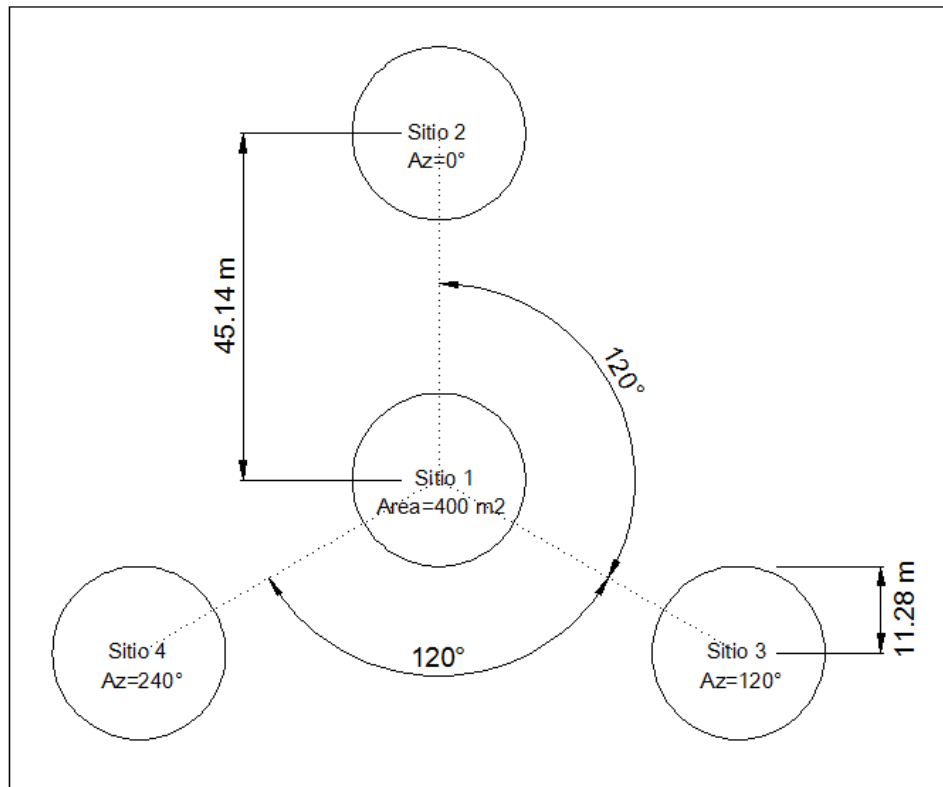


Figura 3.1. Ilustración del diseño de los conglomerados utilizados para el muestreo.

### 3.3.2 Datos lidar

Se cuenta con las nubes de puntos de cada uno de los 50 conglomerados utilizados, que incluyen sus coordenadas X, Y y Z en la proyección Universal Transversa de Mercator (UTM) zona 14, número de retorno (hasta 3 retornos) e intensidad. El vuelo lidar se llevó a cabo en los meses de noviembre y diciembre de 2007 por Fugro Chance de México S. A. de C. V. y fueron adquiridos vía el Instituto Nacional de Estadística y Geografía (INEGI), obtenidos a una altura de vuelo de 2000 a 3600 m, velocidad promedio de vuelo de 165 nudos y un ángulo de captura de 25 y 45 grados, dentro del Sistema Geodésico de Referencia 1980 (GRS80, del inglés Geodetic Reference System 1980). La densidad de la nube es de aproximadamente 0.5 pts/m<sup>2</sup>.

El filtrado de los datos en puntos del suelo y de no-suelo se llevó a cabo utilizando la propuesta multi-resolución basada en la Transformación Hermite Multiescala (MHT, del inglés Multiscale Hermite Transform) (Silván-Cárdenas y Wang, 2006), directamente por el Dr. José Luis Silván, actualmente investigador del Centro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo” A. C. (CentroGeo) en el área de procesamiento digital de imágenes. Para mayor detalle consultar Silván-Cárdenas y Wang: *A multi-resolution approach for filtering LIDAR altimetry data* (2006).

### 3.4 OBTENCIÓN DE ALTURAS SOBRE EL TERRENO

En general, la forma de obtener la altura del dosel de la vegetación sobre la superficie del terreno consiste en generar un modelo de la parte alta del dosel y otro del terreno para después realizar la diferencia de alturas entre estos. Propiamente se genera un Modelo Digital de Superficie (DSM, del inglés Digital Surface Model) para el dosel y un Modelo Digital de Terreno (DTM, del inglés Digital Terrain Model) para la superficie del suelo, y entonces se obtiene un Modelo de Alturas del Dosel (CHM, del inglés Canopy Height Model) al hacer la diferencia. Para esto es necesario determinar qué puntos pertenecen al suelo y cuáles son de la vegetación; una forma típica es considerar los primeros retornos como vegetación para construir el DSM y los últimos retornos como puntos del terreno para construir el DTM, sin embargo, esto sólo es posible cuando los datos lidar se recolectan conteniendo estos dos tipos de retornos y haciendo distinción entre ellos. Otra forma de diferenciarlos es simplemente considerar como vegetación aquellos puntos más altos dentro de cierto vecindario y como terreno los más bajos. (Shan y Toth, 2009:343).

También se pueden calcular las alturas directamente en los puntos de la nube lidar, al restarle a la elevación de cada punto la elevación de la superficie del terreno. Se realiza un filtrado de los puntos removiendo de la nube original todos aquellos que se consideren vegetación, mediante algoritmos con criterios estadísticos o morfológicos de la distribución. La superficie del terreno se construye con los puntos no removidos e interpolados. (Lim et al, 2003:96).

En este caso, y considerando que tanto la identificación de árboles como el cálculo de carbono por área se pretenden hacer sobre la nube, se optó por calcular la altura de cada punto de manera individual, esto es, sin la construcción de algún modelo digital de superficie. Ya teniendo los puntos clasificados en vegetación y no vegetación (cabe aclarar que ninguno de los sitios contiene elementos urbanos o construcciones), a la altura de cada punto de vegetación se le restó el promedio de alturas de todos los puntos del terreno que se encontraran alrededor de él dentro de un radio determinado, en este caso de 8 m.

A manera ilustrativa, se presenta la nube del conglomerado 575, con todos los puntos (figura 3.2), sólo los puntos del terreno (después de la segmentación en vegetación-no vegetación) (figura 3.3), y sólo los puntos de la vegetación después de calcularseles la altura sobre el suelo (figura 3.4), utilizando el programa Fusion para procesamiento de datos lidar. Cada conglomerado está identificado por un número, asignado por la PAOT. En este trabajo se utilizaron 50 conglomerados, cuya numeración no es

precisamente de 1 a 50, por eso se tiene el conglomerado 575. La superficie del terreno en este lugar es escarpada, como se puede ver en las figuras 3.2 y 3.3, aunque en la figura 3.4 pareciera como si los puntos de vegetación estuvieran sobre un terreno plano. Esto es debido a que en el proceso de cálculo de alturas sobre el suelo, a la elevación de cada punto vegetación se le resta la elevación del terreno en su misma ubicación, y la figura 3.4 sólo grafica esta diferencia, “suponiendo” a la superficie del terreno completamente horizontal.

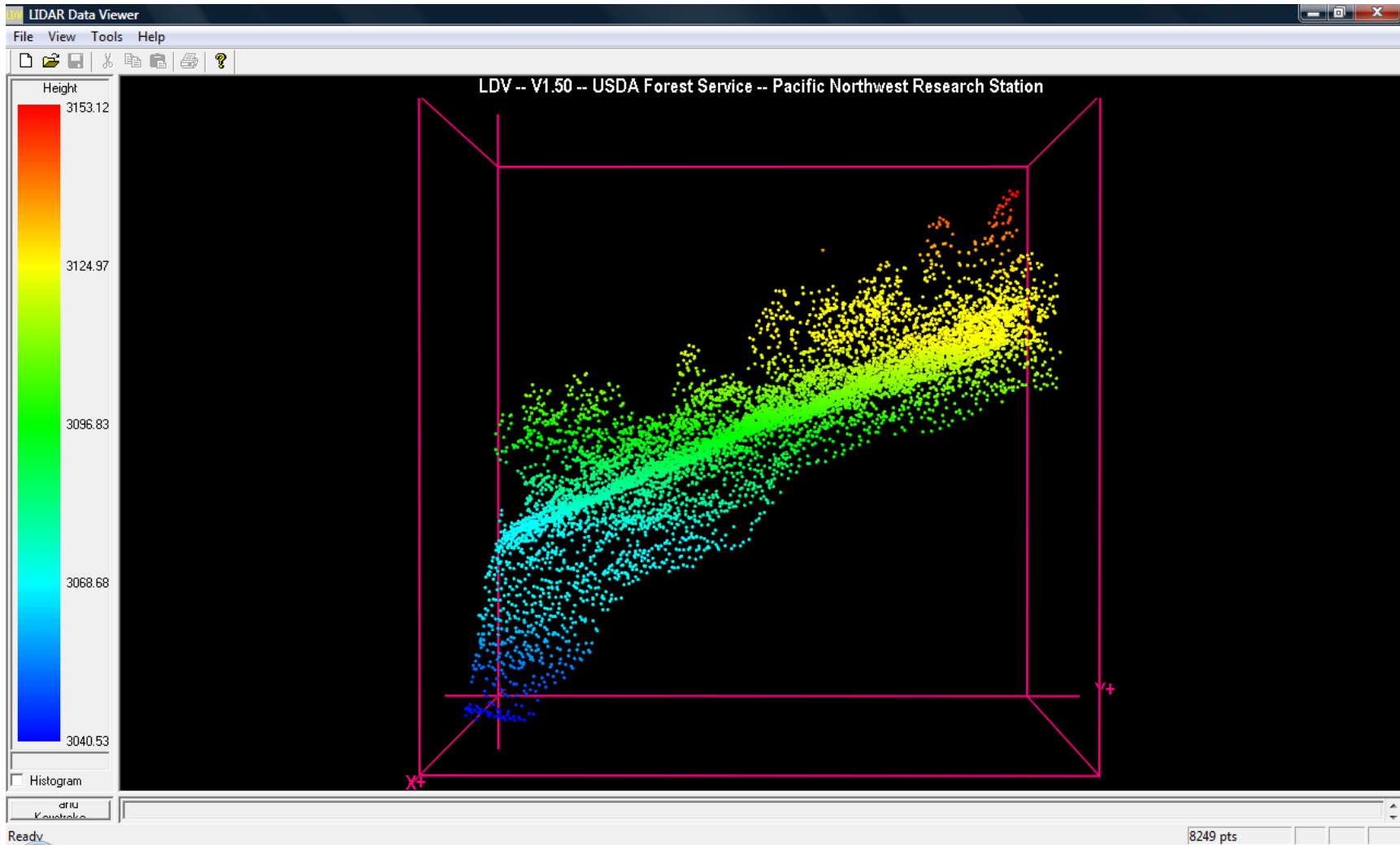


Figura 3.2. Imagen de la nube de puntos del conglomerado 575, todos los puntos. El color de cada punto está en función de su coordenada z (elevación), de acuerdo con la escala de la izquierda.



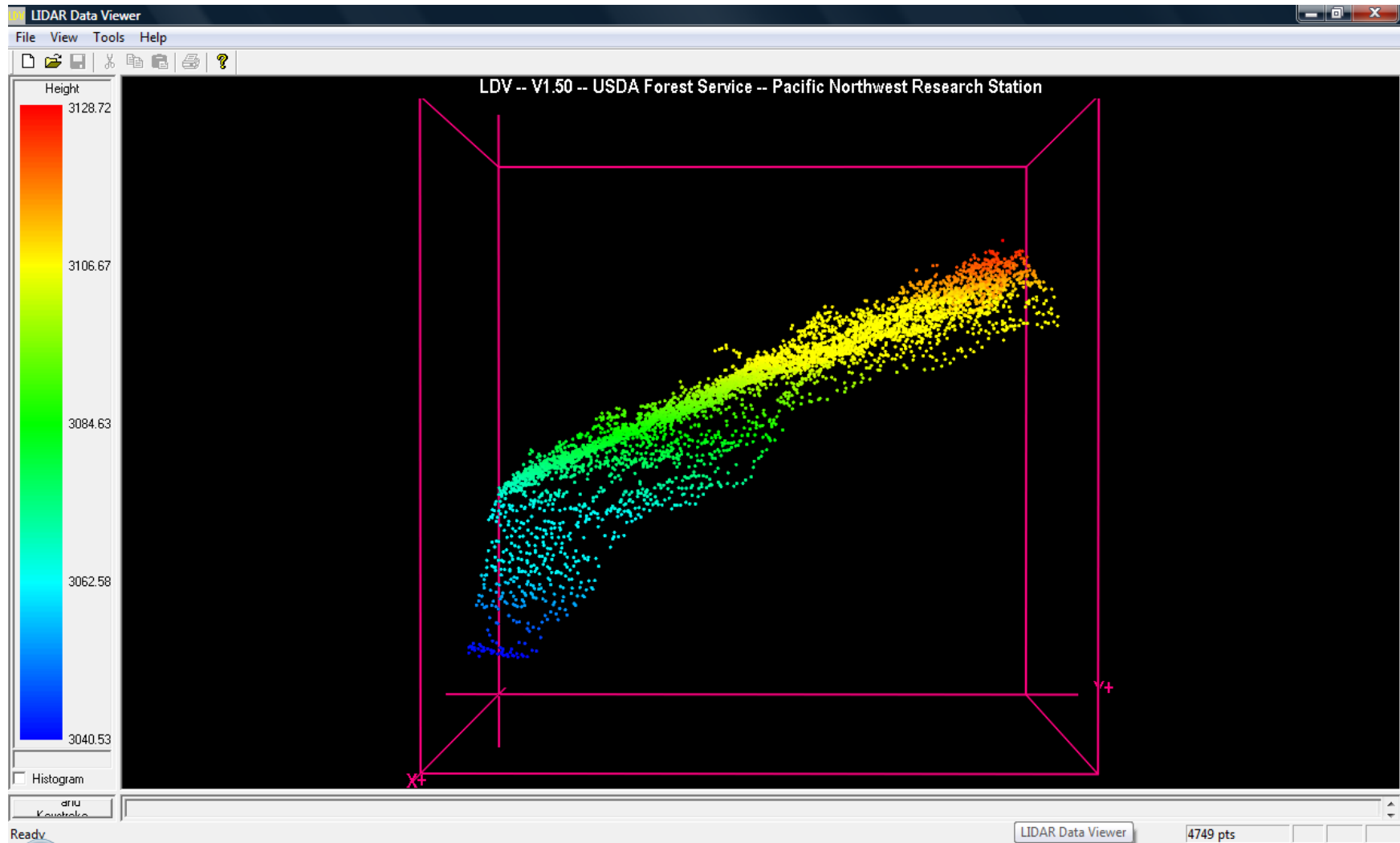


Figura 3.3. Imagen de la nube de puntos del terreno del conglomerado 575. El color de cada punto está en función de su coordenada z (elevación), de acuerdo con la escala de la izquierda.

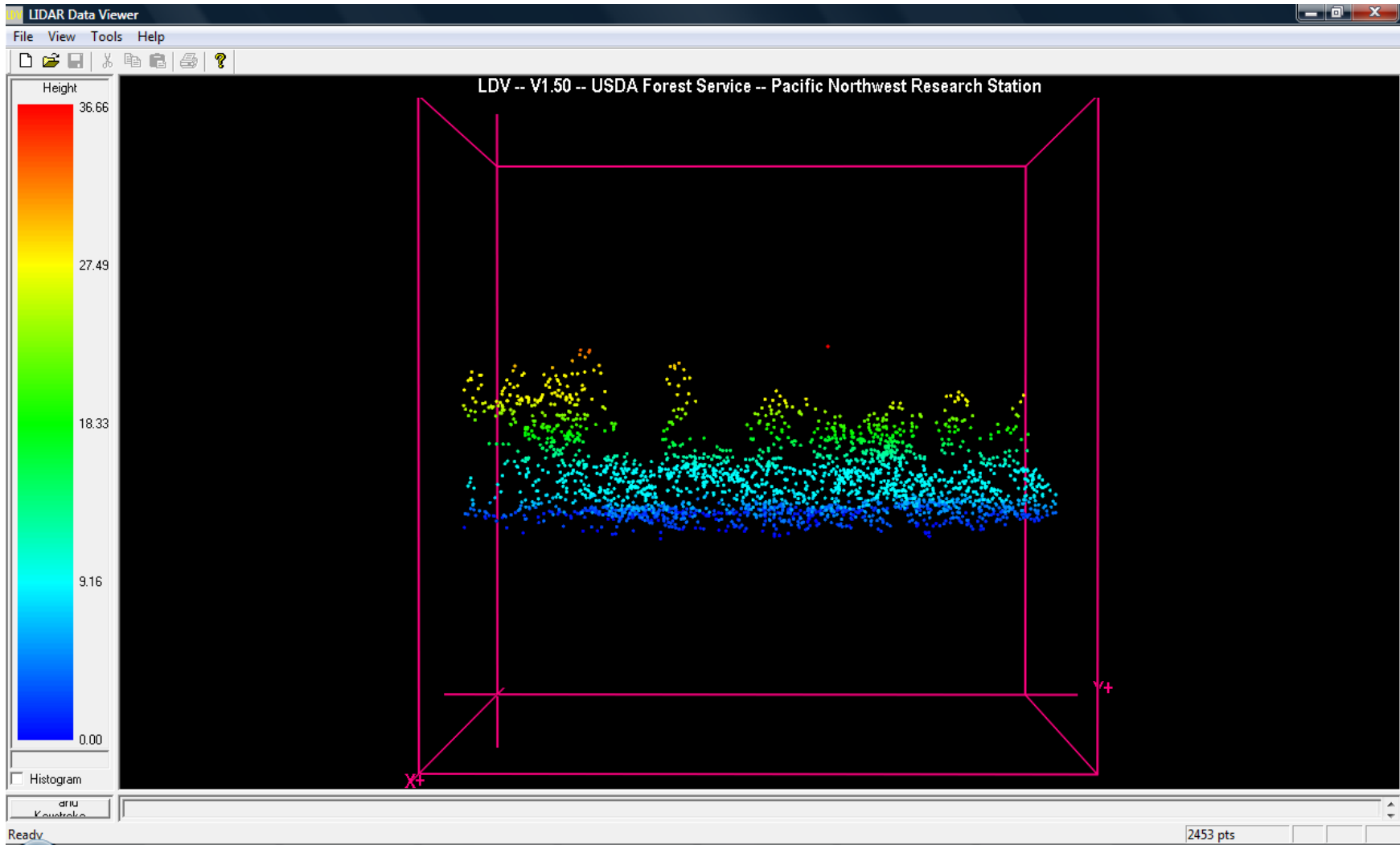


Figura 3.4. Imagen de la nube de puntos de vegetación del conglomerado 575. El color de cada punto está en función de su altura sobre el terreno, de acuerdo con la escala de la izquierda.

### 3.5 IDENTIFICACIÓN Y SEGMENTACIÓN DE ÁRBOLES INDIVIDUALES

Las dos grandes fases dentro del proceso de segmentación son la localización del árbol y la delineación de su copa. La delineación es la determinación de los límites de la copa en todas direcciones, lo que muestra su forma, y sirve para determinar sus dimensiones. Lo más común es encontrar metodologías implementadas sobre imágenes raster o grids obtenidos de la nube de puntos (Hyypä et al., 2005; Popescu, 2007; y Yu et al., 2011, por citar algunos ejemplos), aunque también existen propuestas que combinan el uso de una superficie o imagen conjuntamente con la nube tridimensional, ya sea que se localicen los árboles en la superficie y se segmente sobre la nube (Morsdorf et al., 2003) o se segmente sobre el raster y se refine la localización del árbol sobre la nube (Pyysalo y Hyypä, 2002). Es complicado encontrar propuestas que realicen todo el proceso utilizando sólo la nube de puntos lidar.

En cuanto a la localización el método más usual es la identificación de máximos locales (Popescu et al., 2003; Schardt et al., 2002; y Vega y Durrieu, 2011, por ejemplo), no obstante, también se encuentran otros (Pitkanen et al., 2004; Pyysalo y Hyypä, 2002; y Rahman et al., 2009). Se optó por utilizar el reconocimiento por máximos locales porque es el método más usual, empleado en la mayoría de los trabajos de identificación de árboles, es práctico, pues por su sencillez, es rápido en su desempeño e implementación, y es lógico e intuitivo, pues es fácil entrever y comprender que lo más probable es que los puntos más altos en la nube sean los puntos más altos de los árboles y los más cercanos a la ubicación de su tronco, y que difícilmente dos árboles estarán lo suficientemente juntos para ser captados sus puntos más altos como uno solo y que difícilmente un árbol tendrá una sección de su copa lo suficientemente separada como parecer otro árbol independiente.

Los métodos de delineación más conocidos son el basado en delimitación de cuencas (Hyypä e Inkinen, 2009; Yu et al., 2011) y el de crecimiento de región (Solberg et al., 2006), si bien se han dado otros (Persson et al., 2002, por ejemplo), propuestas implementadas sobre la nube son pocas y por lo general se apoyan de la construcción de algún modelo de superficie en algún paso del proceso. Se decidió trabajar el algoritmo K-means como segmentación porque es una propuesta aplicada sobre la nube de puntos, y por ser más fácil de implementar, además de ser un proceso intuitivo, relativamente sencillo de comprender. Aunque la mayoría de los trabajos en segmentación utilizan buenas densidades de puntos, empero una alta densidad tampoco es garantía de precisión (además de que también deben considerarse los métodos empleados). Morsdorf et al. (2003) trabajando con 10 pts/m<sup>2</sup> obtienen resultados similares que Rahman et al. (2009) con 70 pts/m<sup>2</sup> (ambos obtienen una precisión de detección

del 60 %), siendo éste último más reciente y sofisticado en sus métodos; esto también es un punto de apoyo para la utilización del algoritmo de agrupación.

El objetivo de los algoritmos de agrupación es la clasificación de los objetos de acuerdo a un criterio de similitud -en alguna propiedad- entre ellos. En el caso de K-means se trata de minimizar iterativamente la suma de distancias de todos los puntos que componen un grupo con respecto al centroide del mismo, siendo un proceso muy sensible a la ubicación de los puntos iniciales de agrupación, o puntos semilla (Morsdorf et al., 2003:28). En este caso, se eligen como puntos semilla los máximos locales.

La implementación, en este trabajo, se llevó a cabo en Matlab versión 7.8 y consta de dos fases:

*-Identificación de puntos semilla.* La búsqueda de máximos locales se hace analizando punto por punto; para cada uno de ellos se define un vecindario circular de radio igual al diámetro de copa esperado (DCE), que está en función de la altura del punto, y si todos los puntos contenidos en este vecindario están por debajo del punto trabajado, entonces éste se define como un máximo local. El DCE se determina utilizando la ecuación 3.1, y se puede interpretar como el diámetro de copa más probable de un árbol que tuviera una altura igual a la altura del punto. La ecuación 3.1 es la recta de regresión entre los diámetros de copa promedio por sitio y la altura promedio de los árboles por sitio (figura 3.5), dividida entre dos. En campo no se midió el diámetro de copa de los árboles, por lo tanto fue necesario calcularlo, esto se hizo sirviéndose de la ecuación 3.2, obtenida del trabajo de Jorge Malleux *Estudio de la Relación D.A.P. con el Diámetro de Copa en un Bosque Húmedo Sub Tropical* (Malleux, 1970)<sup>3</sup>, utilizándose estos valores para determinar la regresión diámetro de copa–altura. Sin embargo, en el Suelo de Conservación predominan las coníferas, las cuales presentan diferencias respecto a los árboles del bosque húmedo sub tropical. Por esto, posteriormente el diámetro de copa también se calculó utilizando una ecuación diseñada para Abeto Douglas (que es una conífera), extraída del trabajo de Samantha J. Gill et al. (2000) *Modeling conifer tree crown radius and estimating canopy cover* (ecuación 3.3). Ya que el diámetro calculado a partir de esta última ecuación es menor que el calculado con la ecuación 3.2 (como se muestra en la figura 3.6), y considerando también el hecho de que al correr el algoritmo utilizando sólo la regresión diámetro de copa–altura (sin multiplicar por 0.5, en la ecuación

---

<sup>3</sup> Normalmente las ecuaciones que relacionan dimensiones de las diferentes partes de los árboles son dependientes de las especies, pues cada especie tiene sus peculiaridades estructurales, sin embargo en el estudio de Malleux (1970) no se hace distinción entre ellas, únicamente se menciona que los árboles utilizados pertenecen al bosque húmedo subtropical.

3.1) para calcular el DCE pasaba por alto máximos locales que visualmente sí lo eran, se dividió el DCE entre dos (agregándose el factor 0.5 en la ecuación 3.1), entregando mejores resultados.

$$DCE = (0.36 * A + 4.43) * 0.5 \quad (3.1)$$

$$DC = 3.45 + 0.1314 * DN + 0.001169 * (DN)^2 \quad (3.2)$$

$$DC = \frac{2(1.6654 + 0.0355 * DN)}{\sqrt{2}} \quad (3.3)$$

Donde:

DCE es el diámetro de copa esperado (m),

A es la altura sobre el terreno, del punto (m),

DC es el diámetro de copa (m), y

DN es el diámetro normal (cm).

Para este proceso, no se toman en cuenta todos los puntos por debajo de los 2 m de altura sobre el terreno, aunque este umbral puede ser modificado según las necesidades.

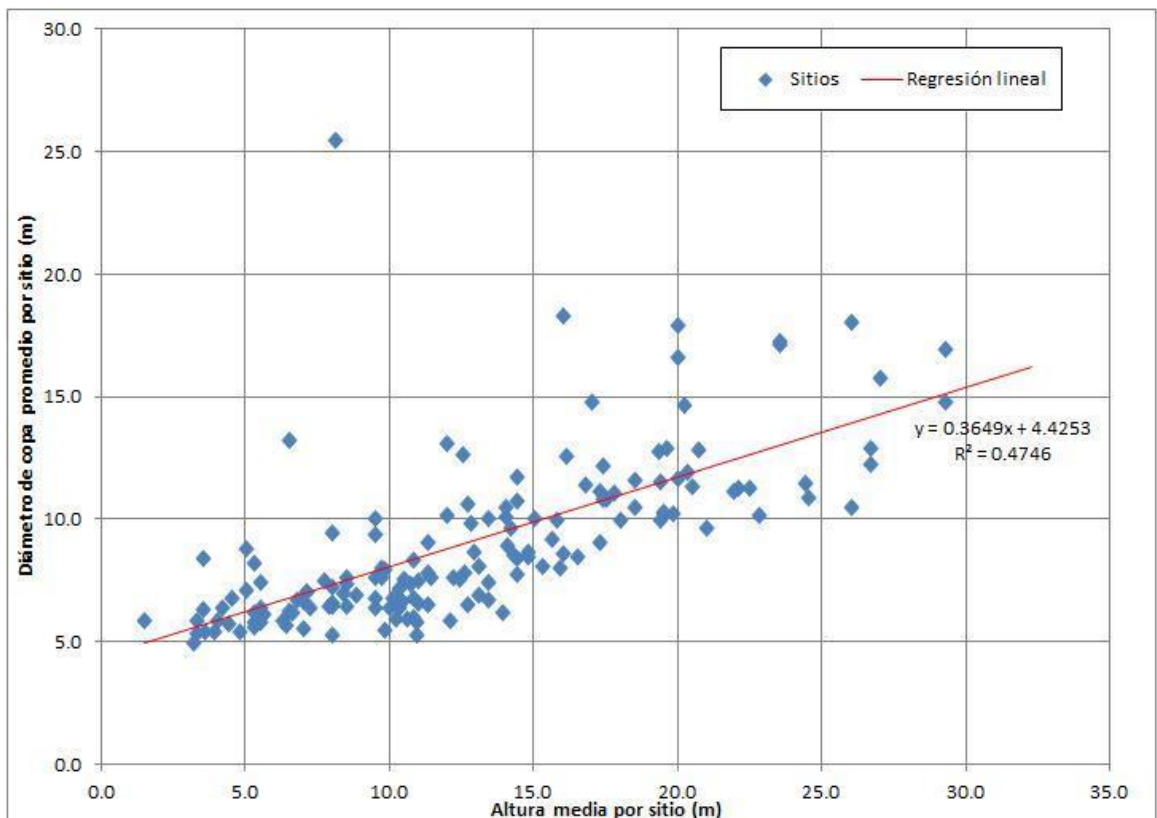


Figura 3.5. Gráfica de dispersión de altura promedio contra diámetro de copa promedio de todos los sitios.

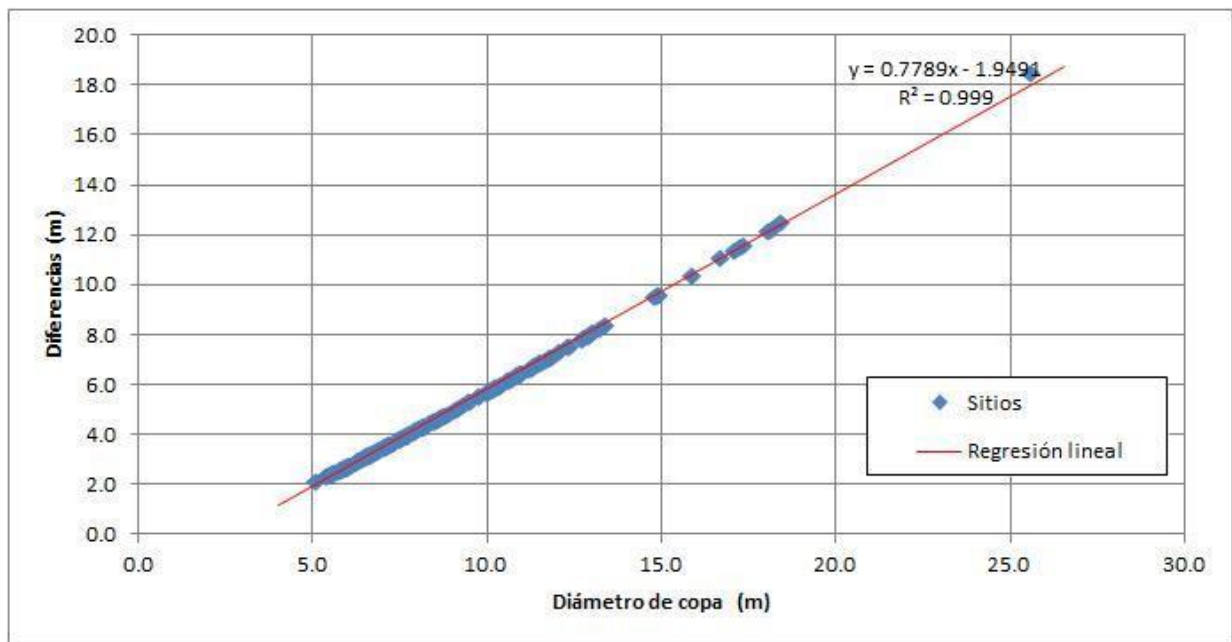


Figura 3.6. Gráfica de dispersión del diámetro de copa promedio calculado con la ecuación del trabajo de Malleux (1970) contra sus diferencias con respecto al diámetro de copa promedio calculado con la ecuación de Gill et al. (2000), para todos los sitios. La ecuación del trabajo de Malleux (1970) es la 3.2, la ecuación del trabajo de Gill et al. (2000) es la número 3.3. En los valores de diámetro de copa donde está la mayor concentración de los puntos (entre 5.0 m y 13.0 m, aproximadamente, de acuerdo a la gráfica), las diferencias son cercanas a la mitad del valor de diámetro de copa.

-*Agrupación basada en K-means*. Se realiza sobre la nube de puntos tridimensional iterativamente en dos pasos:

- Asignación de cada punto a un grupo. La ubicación inicial de los centroides está dada por los máximos locales encontrados en fase anterior; cada punto se asigna al grupo cuya distancia euclidiana con respecto a su centroide sea la menor.
- Recálculo de los centroides. De cada grupo formado se calcula nuevamente su centroide para volver a reasignar los puntos

Este proceso termina cuando la diferencia entre la última y la penúltima posición de alguno de los centroides no varíe, o no varíe más de una distancia especificada por el usuario en el algoritmo. Es importante mencionar que en muchas especies arbóreas la altura de la copa es mayor que su diámetro; Morsdorf et al. (2003) trabajaron con pino como especie dominante y menciona que la altura de la copa de éstos es de 3 a 6 veces mayor que su diámetro (Morsdorf et al., 2003:29). Aunque las muestras obtenidas del Suelo de Conservación del DF incluyen diferentes especies, las dominantes (oyamel y pino) también presentan una copa más alta que ancha, por lo tanto, al calcular las distancias euclidianas, se re-escaló el eje z en un factor de 1/3 para favorecer la construcción de grupos de puntos más alargados en el sentido de este eje, pues K-means tiende a formar grupos de forma esférica.

### 3.6 ESTIMACIÓN DE CARBONO POR DISTRIBUCIÓN DE ALTURAS DEL DOSEL

Básicamente la estimación de carbono –y en general de variables forestales– por los métodos basados en la distribución de las alturas del dosel consiste en obtener en campo una serie de parámetros forestales que se quieran estimar de un área o un conjunto de muestras y una serie de variables estadísticas de los datos lidar de esas mismas áreas, para después generar el modelo o modelos de predicción. Los datos pueden trabajarse en formato de nube de puntos o convertidos a rasters de alturas e intensidades; en el caso particular del presente se utiliza la nube, con objeto de no degradar la información.

El método utilizado para encontrar el modelo de predicción es la regresión lineal múltiple; una regresión es un conjunto de técnicas usadas para establecer una relación entre una variable cuantitativa llamada *variable dependiente* y una o más variables independientes llamadas *variables predictoras*. La ecuación que representa la relación se llama *modelo de regresión* (Acuña, 2011:1). En este caso las variables dependientes son los parámetros forestales a estimar: altura media, diámetro normal medio, cantidad de árboles y cantidad de carbono, por muestra; y las variables predictoras son obtenidas de los datos lidar.

Después de la ubicación horizontal de los puntos, las otras características que presentan y que son utilizadas como base para generar las variables predictoras son la altura sobre el suelo, la intensidad y el número de retorno. Variables de altura pueden ser generadas un sinnúmero: promedios de alturas del dosel, promedio por percentiles, datos de variabilidad (desviación estándar, por ejemplo), altura máxima por área, alturas a la base de copa, etc. Tanto la intensidad como el tipo de puntos (número de retorno y clasificación vegetación-no vegetación) son utilizadas para caracterizar la densidad del follaje y la cantidad de cobertura vegetal en el área. Variables de intensidad como sumatoria de intensidades por área, sumatorias normalizadas, etc., revelan un índice de la densidad del follaje especialmente, mientras que las variables de tipo de punto como número de primeros retornos vegetación por área y proporción de retornos vegetación por retornos totales, por ejemplo, revelan un índice de la cantidad de cobertura vegetal en el área, especialmente.

El estudio de Hall et al. (2005) es uno de los más completos en lo que respecta a la utilización de variables predictoras. Ellos proponen 39 variables procurando sean útiles para cualquier sistema lidar, tomadas y/o modificadas de la literatura de estimación forestal, esforzándose por que sea un conjunto lo más completo posible (Hall et al., 2005:193). Los parámetros que estiman son la altura media, biomasa

total aérea, biomasa del follaje, área basal, densidad de árboles, altura a la base de la copa y la densidad del combustible de la copa, algunas de ellas utilizadas como indicadores de almacenamiento de carbono y otras como argumentos de entrada en modelos de comportamiento de incendios (Hall et al., 2005:190). En general se obtiene buena correlación entre los valores estimados y los observados (coeficientes de determinación entre 57% y 87%), presentando resultados útiles para altura media, biomasa y área basal.

Las variables predictoras que se utilizan en el presente están basadas en el trabajo de Hall et al. (2005), muchas de ellas son las mismas y algunas de ellas modificadas, y se dejan de lado algunas otras, pues en este caso, aunque se tiene el número de retorno de cada punto, no se cuenta con la información de qué retornos pertenecen a cada pulso, lo cual es necesario para calcular algunas. Se eligieron las variables de este estudio por ser completo, enfocado a variables de almacenamiento de combustible, con resultados satisfactorios y con una baja densidad de puntos (1.2 pts/m<sup>2</sup>), tal y como es el caso de este trabajo.

Las 25 variables utilizadas se describen brevemente a continuación y en la tabla 3.1 se muestra qué es lo que se pretende caracterizar de la vegetación con cada una de ellas.

H1.-Altura media de los primeros retornos vegetación (m).

H2.-Altura media de los primeros retornos vegetación mayores a 2 m (m).

Hmax.-Altura del retorno más alto (m).

H5.-Altura media del 5% más alto de los retornos vegetación (m).

H10.-Altura media del 10% más alto de los retornos vegetación (m).

H25.-Altura media del 25% más alto de los retornos vegetación (m).

H50.-Altura media del 50% más alto de los retornos vegetación (m).

H75.-Altura media del 75% más alto de los retornos vegetación (m).

H90.-Altura media del 90% más alto de los retornos vegetación (m).

H95.-Altura media del 95% más alto de los retornos vegetación (m).

Hc.-Altura media de la última clase de 0.5 m con retornos vegetación, de arriba hacia abajo. La altura mayor (Hmax) se redondea al medio entero más alto, se construyen de esta forma clases a cada 0.5 m y se cuenta el número de retornos que caen en cada una de ellas. Para calcular Hc se elige la última clase con algún retorno vegetación antes de encontrar la primera con cero retornos, sin considerar la clase más alta ni las que tengan cero retornos que se encuentren inmediatamente después de la más alta (m).

DE2.-Desviación estándar de la altura de los primeros retornos vegetación mayores a 2 m (m).

DEr.-Desviación estándar relativa: DE2/H2 (sin unidades).



Rango.-Rango de las alturas de los primeros retornos vegetación mayores a 2 m (m).

Rangorel.-Rango relativo:  $\text{Rango}/H2$  (sin unidades).

SIDtodos.-Suma de intensidades de todos los retornos vegetación entre el área (suma de intensidades/m<sup>2</sup>).

SID2.-Suma de intensidades de todos los retornos vegetación mayores a 2 m de altura, entre el área (suma de intensidades/m<sup>2</sup>).

SIDrel.- Suma de intensidades de todos los retornos vegetación entre la suma de intensidades de todos los retornos, entre el área (suma de intensidades/m<sup>2</sup>).

SIDrelcorreg.- SIDrel corregida por reflectancia del suelo. Todas las intensidades de los retornos del suelo se multiplican por dos (suma de intensidades/m<sup>2</sup>).

SIDpesada.- Suma de intensidades de todos los retornos vegetación mayores a 2 m de altura, en donde cada intensidad se multiplica por el inverso de su altura; todo dividido entre el área (suma de intensidades/m<sup>2</sup>).

VxP.-Número de retornos vegetación entre número total de pulsos (sin unidades).

VxA.-Número de retornos vegetación por área (retornos/m<sup>2</sup>).

V2xA.-Número de retornos vegetación mayores a 2 m de altura por área (retornos/m<sup>2</sup>).

Ps1.-Proporción de primeros retornos del suelo respecto a todos los primeros retornos (sin unidades).

P1s.-Proporción de primeros retornos del suelo respecto a todos los retornos suelo (sin unidades).

Tanto las variables predictoras como las variables independientes son calculadas por sitio, cada sitio representa una muestra, lo cual significa que se cuenta con una cantidad de 200 muestras (pues cada conglomerado tiene 4 sitios, siendo 50 conglomerados). Sin embargo, existen sitios en los cuales no se presenta ningún árbol, o son arbustos pequeños que no fueron censados, tales sitios se descartaron del proceso pues son fuentes de desviaciones (la variable dependiente valdría cero y las variables predictoras sí tendrían un valor), quedando un total de 170 muestras.

Para cada muestra se recortó la nube de puntos contenida en el área y se calcularon todas las variables predictoras de acuerdo a la descripción antes hecha de ellas, utilizando Matlab. En cuanto a las variables dependientes, se contó el número de árboles censados en campo en cada sitio (esta es la variable número de árboles), se calculó el promedio de alturas de todos ellos, por sitio, para obtener la altura media; el promedio de los diámetros normales de todos para obtener el diámetro normal medio; y se calculó el carbono por árbol utilizando las ecuaciones presentes en el estudio *Estimación de captura de carbono como indicador del estatus del derecho de los habitantes del Distrito Federal a gozar de áreas verdes urbanas adecuadas para su desarrollo, salud y bienestar* (PAOT, 2010), (págs. 16-19), para

después obtener la cantidad total de carbono por sitio, al sumar el de cada árbol. Comúnmente el carbono se expresa por área (Ton/Ha), en este caso se hizo la conversión de unidades necesaria para expresarlo así e introducirlo en el proceso de búsqueda del modelo.

<b>VARIABLES PREDICTORAS OBTENIDAS DE LA NUBE DE PUNTOS</b>	<b>PARÁMETRO FORESTAL QUE CARACTERIZAN</b>
H1 H2	Altura media del dosel
Hmax	Altura máxima del dosel
H5 H10 H25 H50 H75 H90 H95	Altura media del dosel fraccionada (Altura media del 5% más alto del dosel, del 10% más alto del dosel, etc.)
Hc	Altura media de la base de la copa
DE2 Der Rango Rangorel	Variabilidad de la altura del dosel
SIDtodos SID2 SIDrel SIDrelcorreg SIDpesada VxP VxA V2xA	Densidad del follaje
Ps1 P1s	Cantidad de cobertura vegetal en el área

Tabla 3.1. Parámetros forestales que caracteriza cada variable predictora utilizada.

Antes de iniciar el proceso de regresión se realizó un análisis exploratorio de la relación entre cada una de las variables predictoras con cada una de las variables dependientes. Acuña en su *Análisis de Regresión* (2011), sostiene que existen tres alternativas para mejorar el poder predictivo de un modelo de regresión univariado: a) transformar la variable predictora, la variable dependiente, o ambas; b) utilizar un modelo de regresión polinómico, o; c) aumentar el número de variables predictoras para pasar al campo de la regresión múltiple; y demuestra de forma interactiva con un ejemplo como un modelo múltiple con una variable transformada presenta mejores resultados que un modelo con un mayor número de variables predictoras sin alguna transformación (págs 43-48). Por esta razón se opta por determinar cuáles son los pares de variables que presentan alguna relación no lineal y



Que en forma matricial se muestra como:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (3.5)$$

En donde  $\mathbf{Y}$  es el vector de valores de la variable dependiente,  $\mathbf{X}$  es el vector de valores de las predictoras,  $\boldsymbol{\beta}$  es el vector de coeficientes y  $\mathbf{e}$  es un vector de valores aleatorios, cuya solución es:

$$[\mathbf{X}'\mathbf{X}]\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \quad (3.6)$$

La resolución de este sistema genera un modelo de predicción que incluye todas las variables predictoras, sin embargo, es necesario elegir sólo un grupo de ellas por varias razones: a) no todas las variables predictoras tiene igual importancia, siendo más eficiente trabajar sólo con las importantes, b) eliminar las redundantes que puedan afectar el desempeño del modelo, c) es más fácil trabajar con un grupo de variables pequeño, d) es más económico recolectar información para pocas variables, y e) con el objeto de conseguir un modelo que ajuste bien los datos con la menor cantidad de predictoras posible, haciéndolo más robusto (Acuña, 2011:141).

El método para seleccionar las variables predictoras es el *stepwise selection* (selección paso a paso) que consiste en elegir el mejor modelo de predicción en forma secuencial, añadiendo una variable a cada paso, la que en mayor medida mejore el modelo, y al mismo tiempo descartando aquellas que ya hayan estado incluidas en él, pero que con el ingreso de la nueva variable pueda llegar a ser más conveniente el sacarlas. Es considerado como uno de los mejores procedimientos de selección de variables y recomendado por Draper y Smith en *Applied Regression Analysis* (Análisis de Regresión Aplicado) (1998), ya que economiza el uso de las capacidades computacionales en su cálculo, evita trabajar con más predictoras de las necesarias mientras mejora la ecuación de regresión en cada etapa del proceso (Draper y Smith, 1998:338).

Tanto las gráficas de exploración de la relación entre variables como el proceso de regresión stepwise fueron llevados a cabo en Matlab. El algoritmo de regresión se invoca con el comando `stepwisefit`, y empieza con un modelo inicial simple, para posteriormente ir añadiendo variables y descartando otras en función del incremento del poder explicativo del nuevo modelo, esto se realiza mediante una prueba de contraste de hipótesis. Si el término no está en el modelo la hipótesis nula es que tendría un coeficiente de cero si se ingresara, si existe suficiente evidencia para rechazarla se agrega el término. Si el término ya está en el modelo, la hipótesis nula es que tiene un coeficiente de cero, si no existe suficiente evidencia para rechazarla, se saca el término. El valor utilizado en la prueba de hipótesis es el conocido como F-parcial, que se calcula de la siguiente forma:

$$Fp = \frac{\frac{SSR(C) - SSR(R)}{k}}{MSE(C)} \quad (3.7)$$

Donde:

SSR(C) es la sumatoria de los cuadrados de los residuos de la regresión, del modelo completo,

SSR(R) es la sumatoria de los cuadrados de los residuos de la regresión, del modelo reducido,

MSE(C) es el cuadrado medio del error del modelo completo, y

k es el número de términos excluidos del modelo reducido (en este caso es 1) (Acuña, 2011:56-57).

Llamando modelo completo al cual en el que se incluye el término candidato y modelo reducido al cual en el que no se incluye. Ya que se tiene el F-parcial, entonces se calcula la probabilidad de obtener un valor al menos tan extremo como este F-parcial (a este valor de probabilidad se le llama valor  $p$ ) suponiendo que la hipótesis nula fuera cierta y considerando una distribución de probabilidad F de Fischer-Snedecor. Si en estas condiciones, el valor  $p$  es muy pequeño (generalmente menor al 5 %) entonces hay evidencia para rechazar la hipótesis nula, o si no es tan pequeño (generalmente mayor al 10 %) entonces no se puede rechazar la hipótesis nula.

De manera que la secuencia del algoritmo es:

- a) Se ajusta el modelo inicial.
- b) Si algún término fuera del modelo presenta un valor  $p$  menor que la tolerancia de entrada (esto es, que es improbable que tuviera coeficiente cero si se agrega al modelo), se agrega el que tenga el menor valor  $p$ , y se repite este paso; de otra forma se pasa el paso c.

c) Si algún término en el modelo presenta un valor  $p$  mayor que la tolerancia de salida (esto es, que es improbable que la hipótesis de que tenga coeficiente cero sea rechazada), se remueve el que tenga el valor  $p$  mayor y se pasa al paso b; de otra forma, se finaliza el proceso.

Es importante mencionar que el modelo final de la regresión está en función del modelo inicial con el que se empieza y de la secuencia en la que las variables son agregadas o quitadas; no existe garantía de que con un modelo inicial diferente y/o secuencia diferente no se pueda llegar a un mejor ajuste. Para mayor información del método stepwise y pruebas de hipótesis se puede consultar *Applied Regression Analysis* (Draper y Smith, 1998) y *Análisis de Regresión* (Acuña, 2011).

De las 170 muestras se tomaron la mitad para la construcción de los modelos (segmento de entrenamiento) y la otra mitad para la validación y cuantificación de la bondad del ajuste (segmento de prueba). La forma de repartir las muestras se hizo procurando que quedaran equilibradas en cuanto a tipo de vegetación y conglomerado, es decir, de todos los sitios con el mismo tipo de vegetación la mitad se colocó en el segmento de entrenamiento y la mitad en el segmento de prueba, y de los sitios de un mismo conglomerado, la mitad (2) se colocó también en el segmento de entrenamiento y la otra mitad en el sitio de prueba (tabla 3.2).

Se construyó un modelo por cada variable a estimar: cantidad de carbono, número de árboles, altura media y diámetro normal medio, utilizando todas las muestras del segmento de entrenamiento. Estos modelos estimarían cada variable independientemente del tipo de vegetación. Por otro lado, y considerando que oyamel y pino tienen el número más alto de muestras, 60 para oyamel y 71 para pino, también se construyeron modelos de estimación de carbono, que es la variable que en este caso reviste mayor importancia, específicamente para estos tipos de vegetación

<b>SEGMENTO</b>	<b>Oyamel</b>	<b>Pino</b>	<b>Pastizal y Matorral</b>	<b>Bosque mixto e inducido</b>	<b>Total</b>
<b>Entrenamiento</b>	30	36	7	12	85
<b>Prueba</b>	30	35	9	11	85

Tabla 3.2. Distribución por tipo de vegetación de los sitios usados en los segmentos de entrenamiento y prueba.

### 3.7 EVALUACIÓN Y VALIDACIÓN DE LAS ESTIMACIONES

La forma de validar y conocer que tan bien se desempeñan los procesos elegidos para calcular las variables de interés es llevando a cabo una comparación entre los valores estimados y los realmente existentes, o computados con los datos recabados en campo.

La unidad básica de muestra para los procesos de implementación y predicción es el sitio. Como se ha mencionado antes, se cuenta con datos de campo de 50 conglomerados, cada uno conteniendo 4 sitios, por lo tanto se tienen 200 sitios, o 200 muestras. En el caso de la estimación de carbono por distribución de alturas del dosel, se descartaron 30 muestras por no contener árboles censados, utilizándose sólo 170.

#### 3.7.1 Identificación de árboles individuales

En el caso de la identificación de árboles individuales, las variables estimadas son el número de árboles y la altura media por sitio. De campo se levantaron el número de árboles y la altura de cada uno de ellos, por lo tanto es posible saber la altura media también. Entonces, para evaluar el método de identificación, se realiza una confrontación entre número de árboles detectados y número de árboles censados, obteniéndose así la exactitud, y la altura media censada y la altura media detectada, como otro indicador de la calidad del proceso. Es importante mencionar que la diferencia de tres años entre el momento del vuelo lidar y la toma de datos de campo implicará una diferencia entre las alturas de los árboles estimadas y censadas. Ya que el ritmo de crecimiento está en función de la especie, la edad, el clima y suelo, entre otros factores (Quinto y Moreno, 2010:13), no es sencillo estimar esta diferencia, sin embargo, se debe tener en cuenta que esto contribuirá a que la altura media calculada esté por debajo de la verdadera. La exactitud de la detección y los errores de altura se calculan como sigue:

$$\text{Porcentaje de detección promedio} = \frac{1}{N} \sum_{i=1}^N \left( \frac{Ad_i}{Ac_i} * 100\% \right) \quad (3.8)$$

$$\text{Error absoluto de la detección} = \frac{1}{N} \sum_{i=1}^N \left( \left| 1 - \frac{Ad_i}{Ac_i} \right| * 100\% \right) \quad (3.9)$$

$$\text{Error promedio de altura media} = \frac{1}{N} \sum_{i=1}^N (Hmd_i - Hmc_i) \quad (3.10)$$

$$\text{Error absoluto de altura media} = \frac{1}{N} \sum_{i=1}^N (|Hmd_i - Hmc_i|) \quad (3.11)$$

Donde:

$N$  es el número de sitios,

$Ad_i$  es el número de árboles detectados en el sitio  $i$ ,

$Ac_i$  es el número de árboles censados en el sitio  $i$ ,

$Hmd_i$  es la altura media detectada en el sitio  $i$ , y

$Hmc_i$  es la altura media censada en el sitio  $i$ .

El porcentaje de detección promedio es la media de las detecciones de todos los sitios. La detección de cada sitio es la proporción de árboles detectados con respecto a los censados. Esto es, si de 10 censados se detectaron 8, la detección es del 80 %; si de 10 censados se detectaron 12, la detección es del 120 %. El porcentaje de detección promedio no expresa el número total de detectados con respecto al número total de censados, de todos los sitios, como pudiera llegar a pensarse, por ser un promedio de proporciones. Si el porcentaje de detección promedio es del 100 %, por ejemplo, esto manifiesta que los errores de comisión y omisión (expresados en porcentaje) son de igual magnitud.

El error absoluto de la detección es el promedio de los errores de detección de todos los sitios. Por cada sitio se computa el error expresado en porcentaje, esto es, si de 10 censados se detectan 7, el error absoluto es del 30 %, y si de 10 censados se detectan 12, el error absoluto es del 20 %, para después promediarse.

El error promedio de altura media es la media de las diferencias entre la altura media detectada y la censada, para todos los sitios. Si éste vale 0, significa que los errores de sobreestimación y subestimación de la altura se compensan en la misma medida, si es mayor a 0, en promedio se está sobreestimando y si es menor a 0, se está subestimando la altura. Mientras que el error absoluto de altura media es el promedio de las desviaciones de la altura detectada con respecto a la censada. Este último no ofrece ninguna prueba de si se están obteniendo en el proceso alturas medias estimadas mayores a las reales o menores, sólo expresa la magnitud promedio del error.

Las muestras (sitios) se agruparon en siete clases: oyamel, pino (que son las especies predominantes en los conglomerados), bosque mixto, bosque inducido, pastizal, matorral y agrícola. Pastizal y matorral hacen referencia a tipos de vegetación no arbórea, sin embargo en esas zonas se presentan algunos árboles, que son los utilizados para el estudio. Ya que en las zonas agrícolas no existen árboles censados, se dejaron de lado en el análisis. Estas cuatro variables se calculan por separado para cada una de las clases, lo cual proporciona la pauta para evaluar el desempeño del algoritmo en cada tipo de vegetación.



### 3.7.2 Estimación de carbono por distribución de alturas del dosel

En este caso, las variables estimadas son la cantidad de carbono, número de árboles, altura media y diámetro a la altura de pecho promedio por sitio, contando con las últimas tres de estas de datos levantados de campo. La cantidad de carbono fue calculada primero por árbol utilizando las ecuaciones presentes en el estudio *Estimación de captura de carbono como indicador del estatus del derecho de los habitantes del Distrito Federal a gozar de áreas verdes urbanas adecuadas para su desarrollo, salud y bienestar* (PAOT, 2010), (págs. 16-19), para después agregarla por sitio y dividirla entre la superficie del mismo (400 m<sup>2</sup>), obteniendo así –después de la conversión de unidades– esta cantidad expresada en Ton/Ha.

Para evaluar el desempeño de los modelos se calculó el coeficiente de determinación ( $R^2$ ), el coeficiente de correlación de Pearson y la raíz del error medio cuadrático (RMSE, del inglés Root Mean Square Error) de cada uno, tanto para los datos de entrenamiento como para los de prueba, para determinar el grado de consistencia entre uno y otro, y también por cada tipo de vegetación. Esta se dividió en cuatro clases: oyamel, pino, matorral y pastizal, y bosque inducido y bosque mixto. La razón de separar los resultados de oyamel y de pino es que son las especies predominantes en la muestra y por lo tanto las que más pesan en las predicciones; matorral y pastizal son zonas que presentan pocos árboles, por eso se consideraron conjuntamente, para dejar en una última clase a bosque inducido y mixto. Con esto se busca revelar si existe tendencia en los modelos en función del tipo de vegetación.

Las fórmulas de las variables para estimar el desempeño de los modelos son las siguientes:

$$\text{Coeficiente de correlación de Pearson} = \frac{\sum(V_{ce} - \overline{V_{ce}})(V_{ca} - \overline{V_{ca}})}{\sqrt{\sum(V_{ce} - \overline{V_{ce}})^2 \sum(V_{ca} - \overline{V_{ca}})^2}} \quad (3.12)$$

$$R^2 = (\text{Coeficiente de correlación de Pearson})^2 \quad (3.13)$$

$$RMSE = \sqrt{\frac{\sum(V_{ce} - V_{ca})^2}{n-1}} \quad (3.14)$$

Donde:

$V_{ce}$  es el valor obtenido de los datos de campo de la variable de interés,

$\overline{V_{ce}}$  es la media de los valores obtenidos de campo de la variable de interés,

$V_{ca}$  es el valor calculado con el modelo de la variable de interés,

$\overline{V_{ca}}$  es la media de los valores calculados con el modelo de la variable de interés, y

n es el número de sitios.

El coeficiente de determinación ( $R^2$ ) es el cuadrado del coeficiente de correlación de Pearson, y refleja la proporción de la varianza de la variable que se está estimando que está explicada por el modelo de predicción, en una escala de 0 a 1; 0 significa que el modelo no explica nada de la variable, 1 significa que el modelo explica el 100 % del valor de la variable.

El coeficiente de correlación de Pearson es un índice que valora el grado de relación lineal entre dos variables, esto es, si el cambio o variación de una de ellas es función proporcional de la variación de la otra. En este caso, los valores a comparar son la cantidad de carbono computada con los datos de campo con respecto a la cantidad de carbono estimada con el modelo de predicción, y de la misma manera con las demás variables (número de árboles, altura media y diámetro a la altura de pecho promedio). El coeficiente de Pearson puede tomar valores de -1 a 1, cuando es 1 quiere decir que las variables guardan una relación lineal directa perfecta, cuando está entre 0 y 1 existe correlación positiva, si vale 0 no existe relación lineal en absoluto, si está entre 0 y -1 existe correlación negativa y, cuando es -1 significa que la relación entre las variables es lineal inversa perfecta.

La raíz del error medio cuadrático es un buen indicador de precisión, pues la base de su cálculo son los errores de predicción, esto es, las diferencias entre los valores obtenidos con los datos de campo y los valores obtenidos del modelo de predicción para la variable de interés. Tanto el RMSE como el  $R^2$  y el coeficiente de Pearson son de ayuda para la evaluación de la calidad de los modelos de predicción y proporcionan puntos de apoyo para la comparación con otros trabajos.

## 4. RESULTADOS Y DISCUSIÓN

### 4.1 IDENTIFICACIÓN DE ÁRBOLES INDIVIDUALES

Uno de los elementos básicos para obtener estimaciones exactas por árbol, quizá la más importante, es una buena identificación y correspondencia entre los individuos presentes en el campo y los detectados en gabinete. En la tabla 4.1 se muestran los resultados divididos por tipo de vegetación; ya que el error absoluto promedio de la detección es del 58.1 %, la exactitud global fue de alrededor del 41.9 %.

Tipo de vegetación	Porcentaje de detección promedio	Error absoluto de la detección	Error promedio de altura media	Error absoluto de altura media
Oyamel	68.1 %	54.5 %	2.1 m.	5.1 m.
Pino	91.0 %	49.0 %	4.3 m.	6.8 m.
Bosque mixto	105.4 %	62.6 %	2.8 m.	3.1 m.
Bosque inducido	90.3 %	46.8 %	-0.4 m.	2.6 m.
Pastizal	109.4 %	28.1 %	0.3 m.	1.1 m.
Matorral	168.0 %	107.6 %	-0.5 m.	1.6 m.
<b>Promedio</b>	105.4 %	58.1 %	1.4 m.	3.4 m.

Tabla 4.1. Resultados de la identificación de árboles, por tipo de vegetación. El porcentaje de detección es el promedio de los porcentajes de detección de los sitios de ese tipo de vegetación, el error absoluto de la detección engloba los errores de omisión y comisión conjuntamente expresados en porcentaje, el error promedio de altura es el promedio de las diferencias entre la altura media detectada y la altura media censada por sitio, y el error absoluto de altura es similar al anterior, pero utilizando el valor absoluto de las diferencias.

El porcentaje de detección es la media de las detecciones de todos los sitios, y muestra el tipo de error (omisión o comisión) que se cometió con más frecuencia, si es mayor al 100 %, la sobredetección es más pronunciada, si es menor al 100 %, la subdetección es más pronunciada. En general, los errores de omisión son más numerosos que los de comisión y están distribuidos de manera más uniforme a lo largo del espectro del número de árboles censados en los sitios, esto se nota en las gráficas de dispersión de individuos censados contra detectados (figuras 4.1 a 4.4) y en la pendiente de sus respectivas rectas de regresión, todas menores a 1. Por otra parte, los errores de comisión están confinados prácticamente a aquellos sitios que tiene un número de censados menor a 15 y son más pronunciados, lo que se refleja en un porcentaje de detección promedio del 105.4 %. El error absoluto de la detección comprende el promedio de los valores absolutos de los errores de omisión y comisión expresados en porcentaje; naturalmente es mayor que el que puede deducirse del porcentaje de detección, puesto que en este último los errores de omisión y comisión se compensan. Por ejemplo, en el caso del oyamel, ya que el porcentaje de detección promedio es del 68.1 %, el error promedio de la detección es del 31.9 %, sin embargo el error absoluto de la detección es del 54.5 % (mayor), y esto es debido a que al calcular el

porcentaje de detección algunos individuos que se detectaron de más (errores de comisión) en unos sitios compensan los individuos que no se detectaron en otros (errores de omisión), contrarrestándose, aunque no del todo (por eso el porcentaje de detección es menor al 100 %), mientras que el error absoluto promedia la magnitud (esto es, sin considerar el tipo de error) de todos los errores.

Matorral y oyamel son los tipos de vegetación que presentan mayores errores, lo cual es crítico en el caso de esta última por ser una de las vegetaciones dominantes en la muestra. En matorral se detectaron en promedio casi el 70 % más de los árboles efectivamente presentes (el error de comisión supera al error de omisión), y considerando que de cada dos individuos identificados uno era erróneo (error absoluto de la detección de cerca del 100 %). El mejor desempeño fue para pastizal, en buen grado por la poca cantidad de árboles en esas zonas, lo cual ayuda a cometer menos errores; los de comisión fueron ligeramente mayores, presentando un error promedio cercano al 30 %. Mención aparte es el bosque mixto, que presentó el segundo mayor error promedio de detección, sin embargo el porcentaje de detección promedio fue el mejor (105.4 %), lo cual denota un alto equilibrio entre los errores de comisión y omisión.

Respecto a los errores de altura media, no se les presta demasiada atención, pues al ser fuertes los errores de detección, no se considera conveniente esperar valores muy parecidos de altura media en sitios en donde la relación de árboles identificados con los censados es de cerca del 40 %. Además, en estos errores también va reflejado el hecho de que difícilmente se habrán obtenido puntos precisamente de las partes más altas de los árboles (especialmente con la densidad de puntos de la nube utilizada), los errores del proceso de segmentación suelo-no suelo, y los del cálculo de la altura sobre el suelo. Anteriormente se ha mencionado que por la diferencia de tres años entre la toma de muestras de campo y el levantamiento lidar, las alturas medias obtenidas de la estimación estarían subestimando las censadas en campo. Tampoco se considera confiable cuantificar este sesgo debido al pobre desempeño de la detección. De manera similar a la detección, en el error promedio de altura se compensan los errores de omisión y comisión, y en el error absoluto de altura sí se ven reflejados.

De manera general se atienden dos aspectos en los resultados: el desempeño de la detección y la exactitud de la detección en relación con otras variables (altura y diámetro normal de los árboles). Los resultados y las conclusiones descritas están apoyadas en la inspección visual de las gráficas de dispersión y los valores de los parámetros de las ecuaciones de regresión entre las variables analizadas.

### Desempeño de la detección.

*-Detección general debajo del 50 %.* En general se observa la tendencia a subestimar el número de árboles censados, identificando alrededor de la mitad de ellos. Aunque existe sobreestimación muy marcada en algunos sitios, esta ocurre en minoría de ellos y en una región claramente delimitada (en sitios con un número de individuos censados menor a 15) y la propensión general es a subestimar. Como se ha mencionado, pastizal fueron los sitios con mejor desempeño (error absoluto de la detección de 28.1 %), en este caso, la recta de regresión que relaciona los individuos detectados con los censados se parece mucho visualmente a la recta  $y=x$  (figura 4.1), que es la recta que representa a los puntos en donde el número de árboles detectados es igual el número de censados. Sin embargo, descartando matorral, que fueron de los sitios con peores desempeños (error absoluto de la detección de 107.6 %), de los restantes 4 tipos de vegetación en 3 de ellos se observa la tendencia de detectar sólo alrededor de la mitad de los individuos censados (figuras 4.2, 4.3 y 4.4), siendo dos de ellos (oyamel y bosque inducido) los que presentan mayor densidad de árboles.

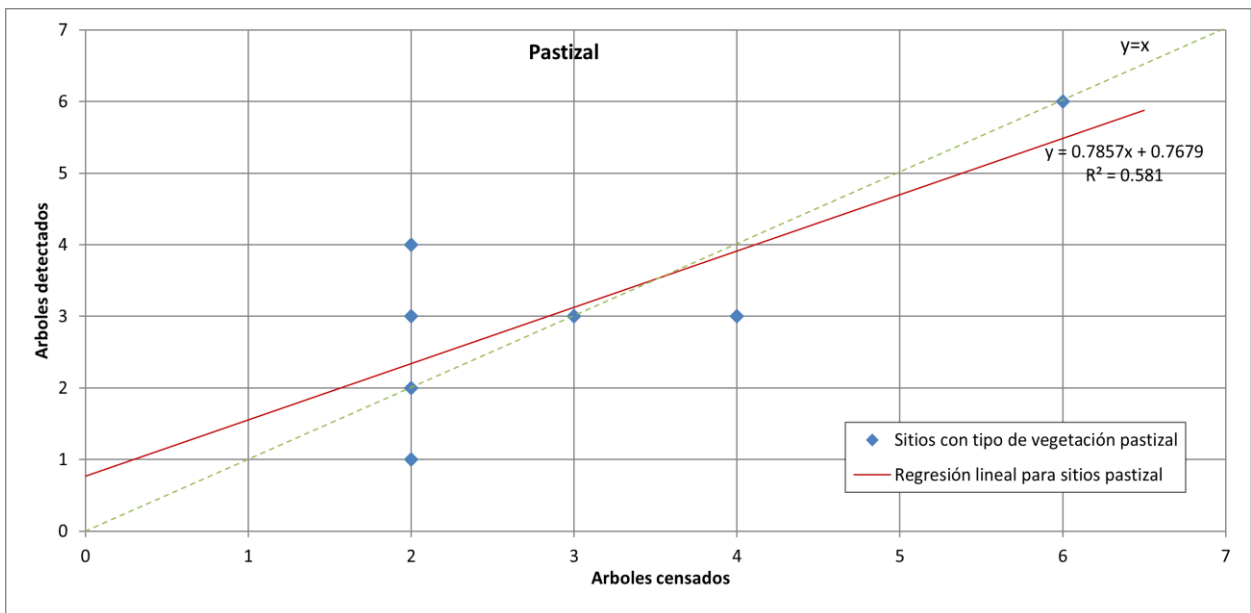


Figura 4.1. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación pastizal. La línea punteada verde es la recta que contiene a todos los puntos en donde el número de detectados es igual al de censados, cuya ecuación es  $y=x$  (pendiente 1 y ordenada al origen 0). La recta roja es la regresión, que tiene una pendiente de 0.78 y corta al eje de las ordenadas en 0.77.

*-En sitios con bajo número de árboles censados se presentó sobredetección, en sitios con alto número de árboles censados se presentó subdetección.* Esto es evidente en los tipos de vegetación que presentan sitios tanto con pocos individuos censados como con muchos. Se puede observar que en el conjunto de los sitios con menos a 10 árboles censados se presenta sobredetección (aunque también hay

subdetección), y en los sitios con más censados se presenta casi siempre subdetección. Las figuras 4.5, 4.6 y 4.7 presentan ejemplo de esto para los sitios de oyamel, pino y bosque inducido respectivamente, que son los tipos de vegetación en donde se observó más claro este comportamiento.

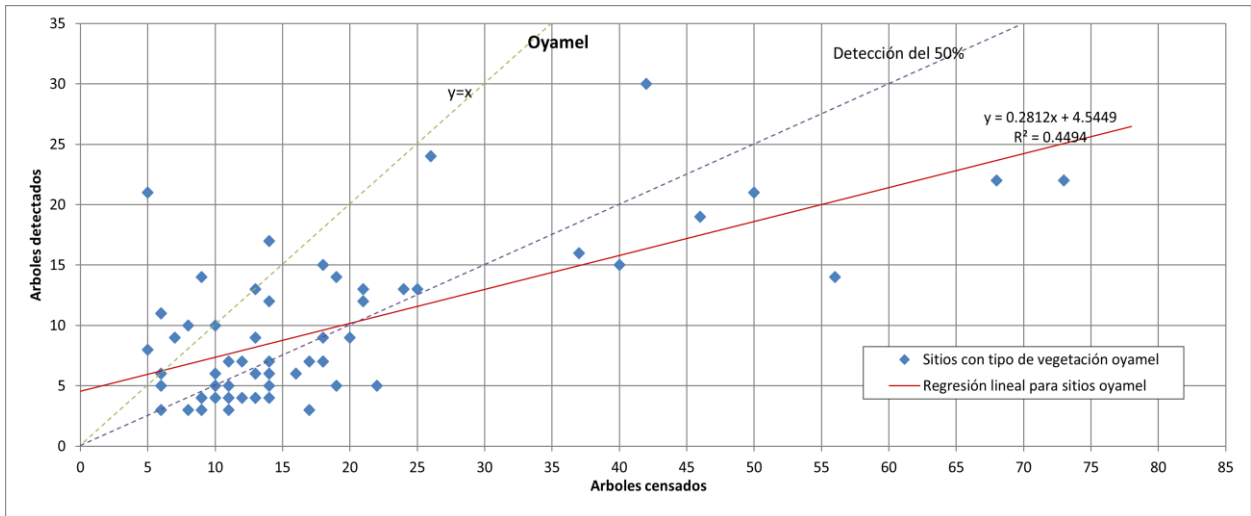


Figura 4.2. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación oyamel. La línea punteada verde es la recta que contiene a todos los puntos en donde el número de detectados es igual al de censados, cuya ecuación es  $y=x$  (pendiente 1 y ordenada al origen 0). La línea punteada púrpura contiene a todos los puntos en donde el número de detectados es igual a la mitad de los censados (pendiente 0.5 y ordenada al origen 0) Los parámetros de la recta de regresión (roja) que relaciona los censados con los detectados se acercan más a los de la línea púrpura.

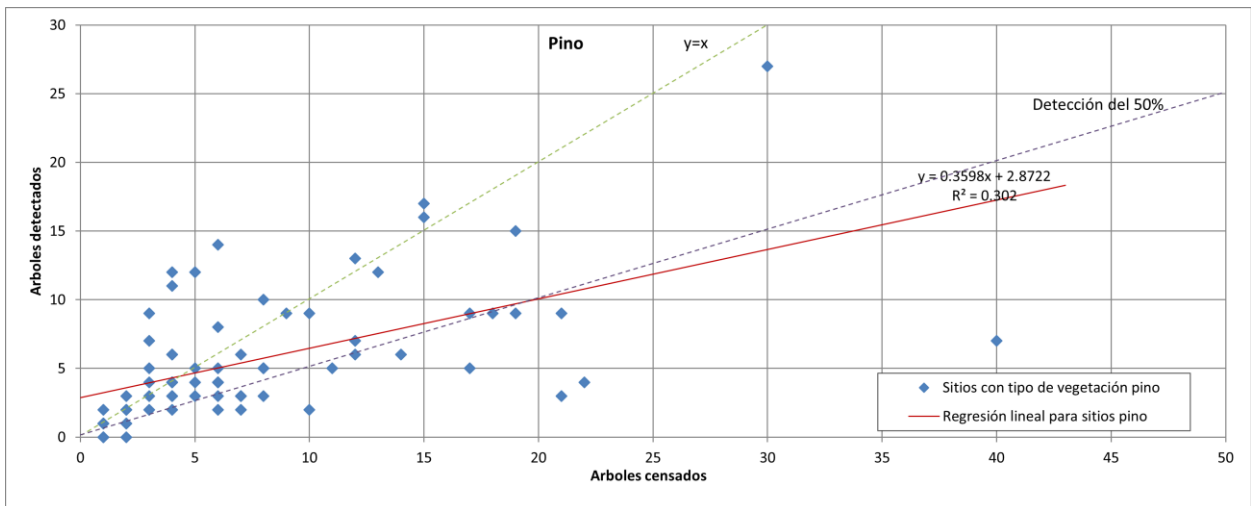


Figura 4.3. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación pino. La línea punteada verde es la recta que contiene a todos los puntos en donde el número de detectados es igual al de censados, cuya ecuación es  $y=x$  (pendiente 1 y ordenada al origen 0). La línea punteada púrpura contiene a todos los puntos en donde el número de detectados es igual a la mitad de los censados (pendiente 0.5 y ordenada al origen 0) Los parámetros de la recta de regresión (roja) que relaciona los censados con los detectados se acercan más a los de la línea púrpura.

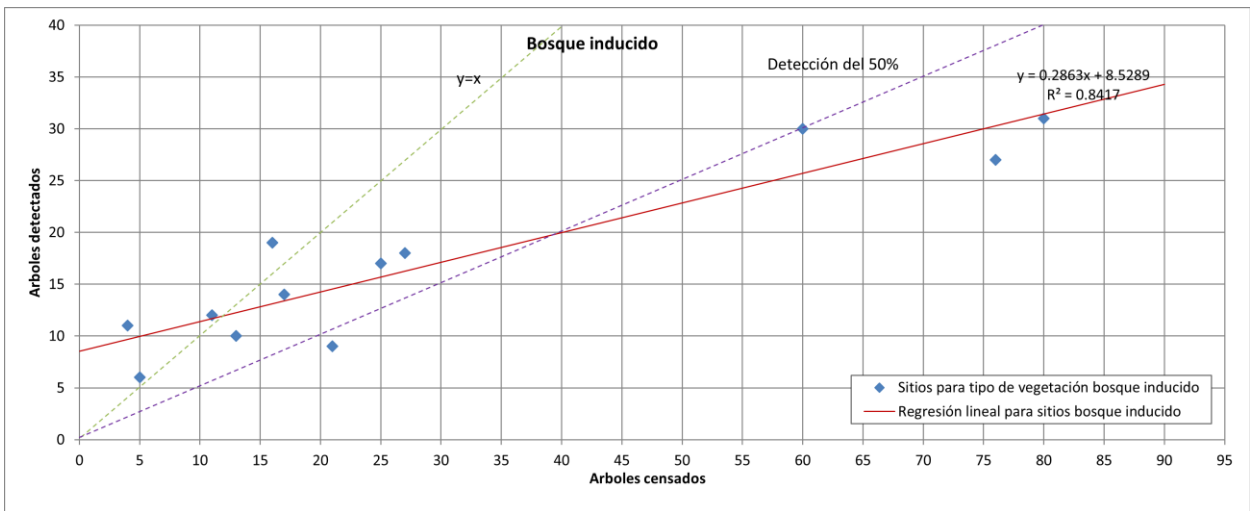


Figura 4.4. Gráfica de dispersión de árboles censados contra árboles detectados en los sitios con tipo de vegetación bosque inducido. La línea punteada verde es la recta que contiene a todos los puntos en donde el número de detectados es igual al de censados, cuya ecuación es  $y=x$  (pendiente 1 y ordenada al origen 0). La línea punteada púrpura contiene a todos los puntos en donde el número de detectados es igual a la mitad de los censados (pendiente 0.5 y ordenada al origen 0) Los parámetros de la recta de regresión (roja) que relaciona los censados con los detectados se acercan más a los de la línea púrpura.

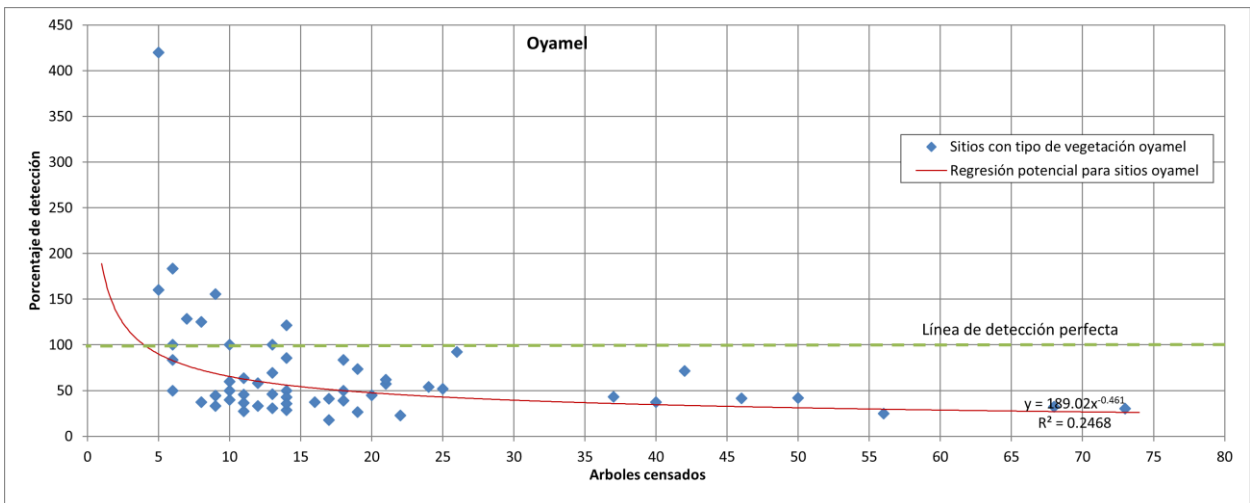


Figura 4.5. Gráfica de dispersión de árboles censados contra porcentaje de detección en los sitios con tipo de vegetación oyamel. La línea punteada verde horizontal representa los puntos en donde el número de censados es igual al número de detectados. Por encima de ella están los sitios con más detectados que censados, y por debajo de ella los sitios con más censados que detectados.

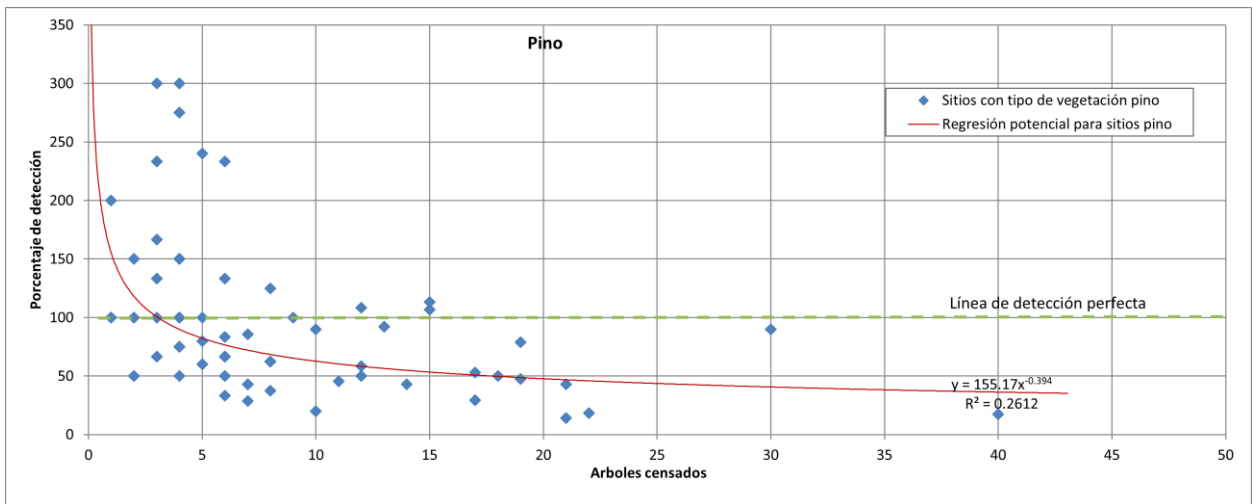


Figura 4.6. Gráfica de dispersión de árboles censados contra porcentaje de detección en los sitios con tipo de vegetación pino. La línea punteada verde horizontal representa los puntos en dónde el número de censados es igual al número de detectados. Por encima de ella están los sitios con más detectados que censados, y por debajo de ella los sitios con más censados que detectados.

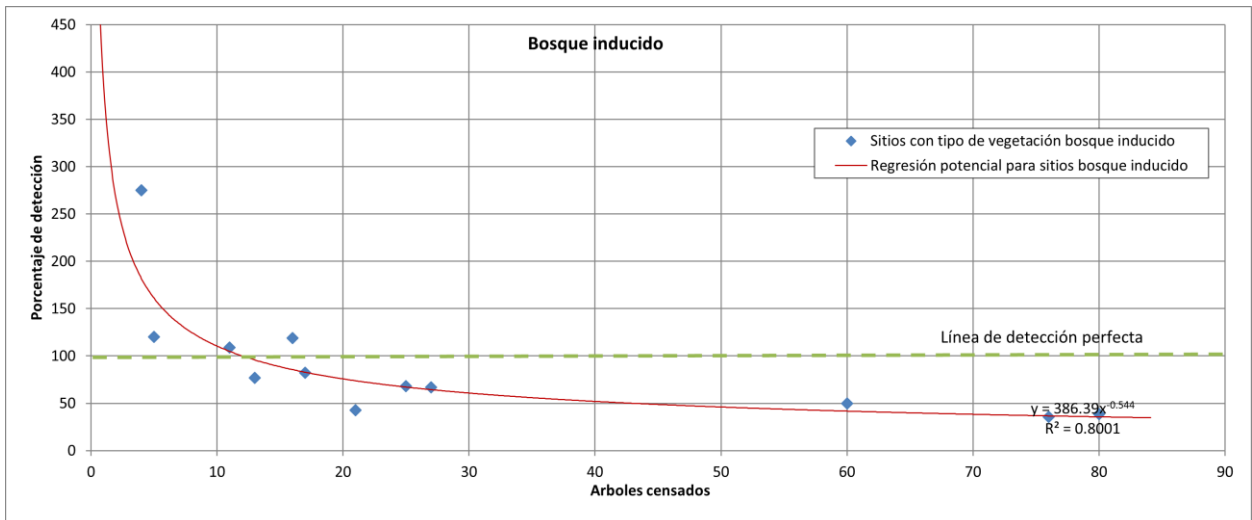


Figura 4.7. Gráfica de dispersión de árboles censados contra porcentaje de detección en los sitios con tipo de vegetación bosque inducido. La línea punteada verde horizontal representa los puntos en dónde el número de censados es igual al número de detectados. Por encima de ella están los sitios con más detectados que censados, y por debajo de ella los sitios con más censados que detectados.

#### Relación de la detección con las otras variables censadas.

*-La altura de los árboles no influye en la detección.* Al inspeccionar visualmente las gráficas de dispersión entre la altura promedio censada y el porcentaje de detección no se puede apreciar algún patrón de relación. Se pueden ver sitios con subdetección y sobredetección a lo largo de todo el espectro de alturas medias de los sitios. Como ejemplo se presenta únicamente la gráfica de dispersión para pino (figura 4.8), ya que los demás tipos de vegetación tampoco muestran algún patrón de relación.



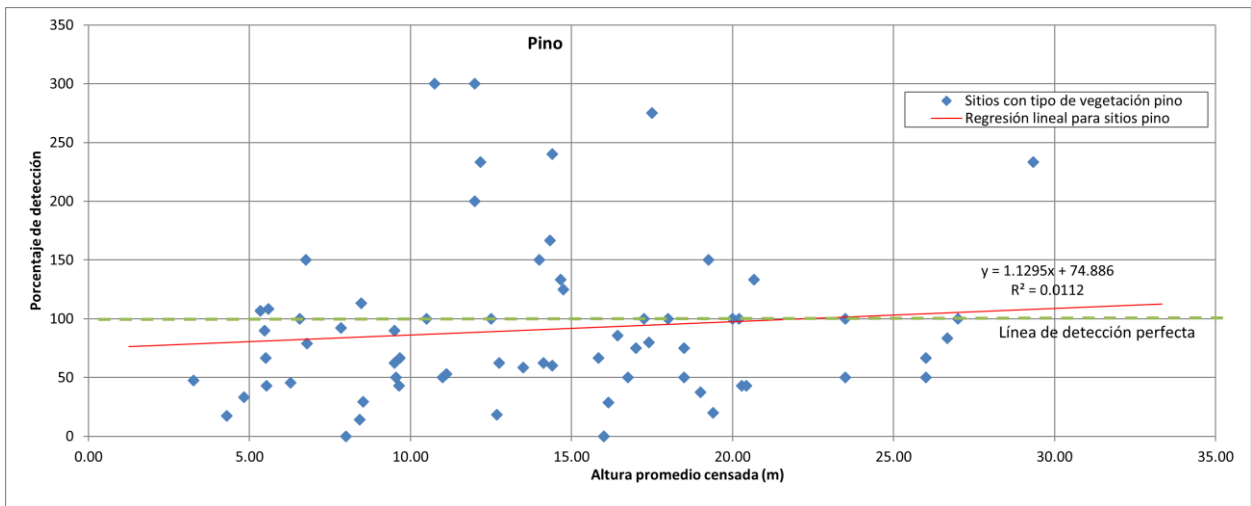


Figura 4.8. Gráfica de dispersión de altura promedio censada contra porcentaje de detección en los sitios con tipo de vegetación pino. En la línea punteada verde caen los sitios donde la detección fue perfecta, por debajo de ella los sitios con subdetección y por encima los con sobredetección. No se puede apreciar un patrón claro.

*-Sobreestimación de altura en sitios con subdetección.* Las gráficas de diferencia en la estimación de altura media con respecto al porcentaje de detección muestran una mayor cantidad de sitios que presentan sobreestimación de altura, lo cual es más marcado en sitios con subdetección, que, como se ha visto, típicamente son las zonas con un alto número de árboles censados. Se presentan las gráficas de porcentaje de detección contra el error de la altura media estimada para oyamel y pino (figuras 4.9 y 4.10), que son los tipos de vegetación más abundantes en las muestras y en donde se puede apreciar mejor este comportamiento. En el caso del pino, se nota más marcada esta tendencia, pues la mayoría de los sitios con subdetección (a la izquierda de la línea punteada verde), sobreestiman la altura. Es de notar que existen varios sitios con detección perfecta (sobre la línea punteada verde) y la gran mayoría también la sobreestiman.

*-El diámetro normal no influye en la detección.* En las gráficas de dispersión de diámetro normal promedio censado contra porcentaje de detección no se observa alguna relación de comportamiento entre las magnitudes de los errores y los diámetros. Distribuidos a lo largo de todo el espectro de valores censados de diámetro normal promedio existen sitios que presentan errores de detección cero (detección perfecta) y sitios con errores fuertes, mayores al 50 %. El diámetro normal es uno de los parámetros más utilizados en la estimación forestal y caracterizador de los árboles, por lo tanto se esperaría que tuviera alguna relación, sin embargo, no lo hay, y más aún si tampoco existe relación entre detección y altura (ver figura 4.8). Como ejemplo se presenta la gráfica de diámetro normal promedio censado contra porcentaje de detección para pino (figura 4.11).

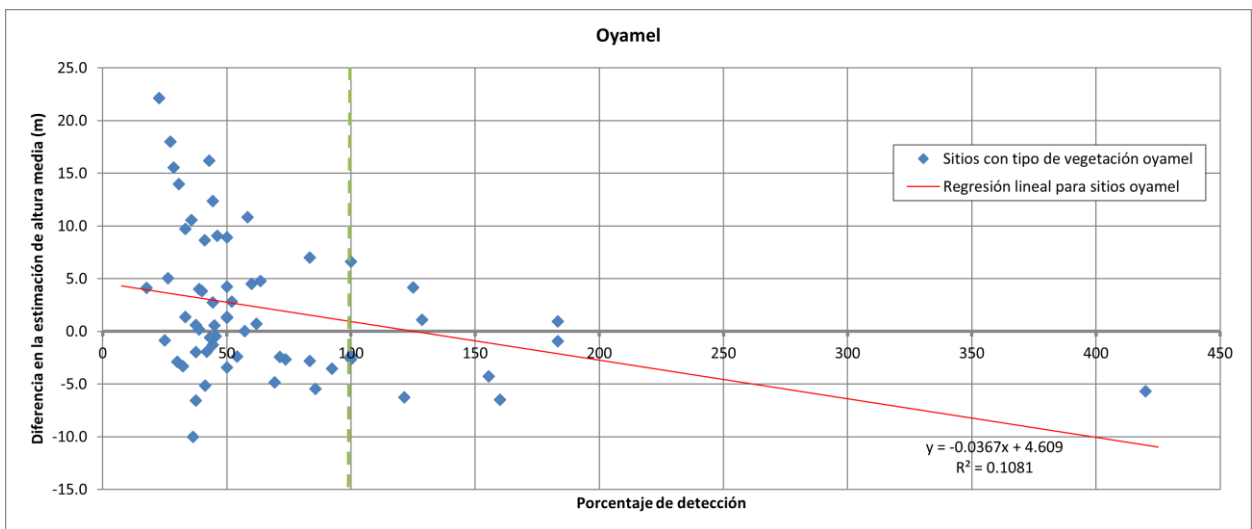


Figura 4.9. Gráfica de dispersión de porcentaje de detección contra el error en la estimación de la altura media en los sitios con tipo de vegetación oyamel. La línea verde vertical representa los puntos con detección perfecta. Como se puede apreciar, la mayoría de los puntos están por encima del eje horizontal que representa cero diferencia en la estimación de altura media, especialmente los que representan subdetección.

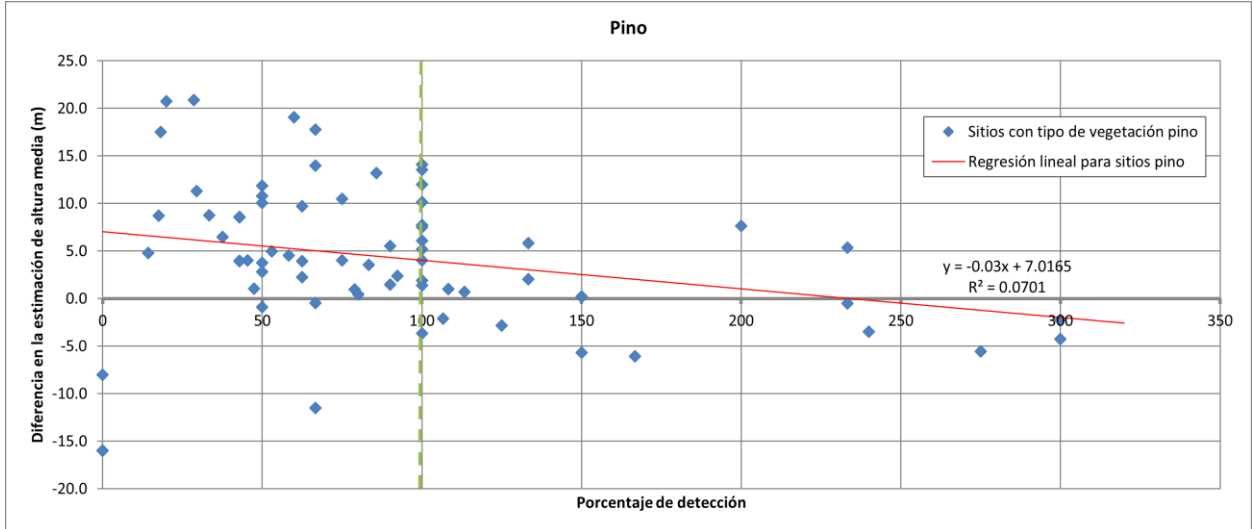


Figura 4.10. Gráfica de dispersión de porcentaje de detección contra el error en la estimación de la altura media en los sitios con tipo de vegetación pino. La línea verde vertical representa los puntos con detección perfecta. Como se puede apreciar, la mayoría de los puntos están por encima del eje horizontal que representa cero diferencia en la estimación de altura media, especialmente los que representan subdetección.

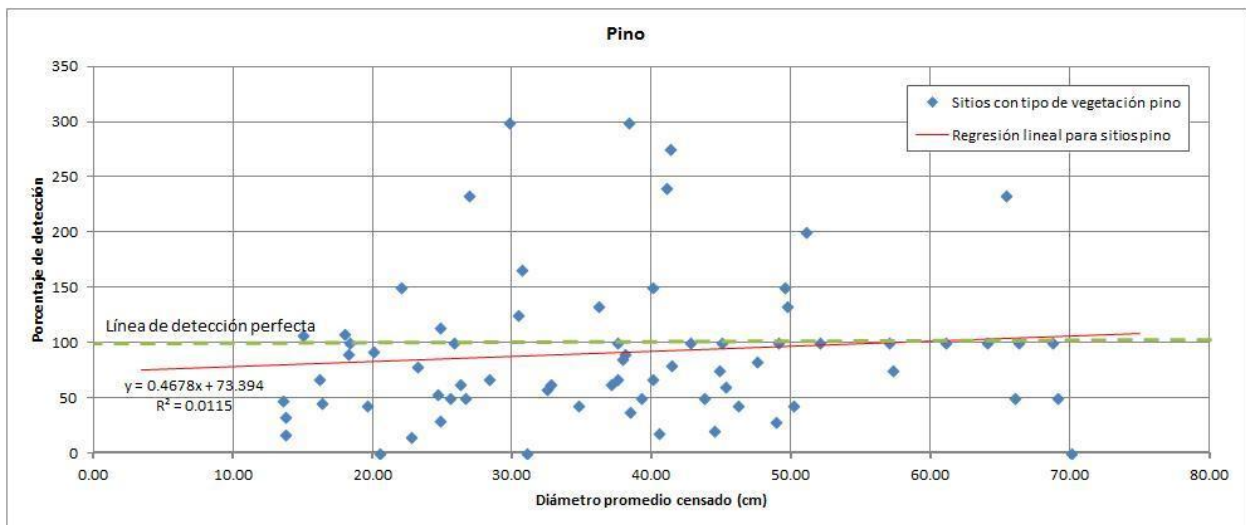


Figura 4.11. Gráfica de dispersión de diámetro normal promedio censado contra porcentaje de detección en los sitios con tipo de vegetación pino. No se observa alguna relación entre estas variables.

Como se ha visto anteriormente, los errores de comisión son muy parecidos en cantidad a los de omisión, lo cual implica aproximadamente un mismo número de árboles detectados contra censados en promedio, entonces ¿por qué se sostiene que se tiene una detección de alrededor del 50 %, e incluso se observa este comportamiento en varias zonas? La respuesta es que son más numerosos los errores de omisión, por eso se observan más de éstos en las gráficas, pero en el menor número de sitios con errores de comisión, éstos son más marcados; al calcularse el índice de detección como porcentaje (como proporción), da más peso a estos sitios al llevar a cabo el promedio, lo que es el contrapeso de los errores de omisión (e incluso los superan ligeramente).

La razón por la que hay sobredetección en zonas con pocos árboles censados, es porque en estas zonas es más factible poder encontrar un mayor número de puntos lidar bajos, cercanos al suelo. Esto debido a que al haber más espaciamiento entre árboles, son más visibles los arbustos pequeños que pudieran originar retornos cercanos al suelo, pero lo suficientemente alejados de él para ser considerados como puntos de vegetación por la segmentación suelo-no suelo. También puede darse el caso de árboles correctamente identificados en la nube, pero que no fueron censados por tener un diámetro normal menor a 7.5 cm, lo cual materializa un error de comisión.

La razón por la cual existe subdetección en las zonas con muchos árboles, es porque, al ser lugares con una densidad fuerte de individuos, sólo es posible identificar aquellos que pertenecen a las capas dominantes por ser los más visibles; los de las capas inferiores no alcanzan a ser detectados por el sensor, y si llegan a producir retornos, es prácticamente imposible determinar justificadamente que ahí existe un árbol, cubierto por otros más grandes. En otros casos, árboles medianos o pequeños, junto a

árboles grandes, que sí son caracterizados por la nube de puntos, pueden llegar a ser confundidos con partes del grande.

La diferencia de tiempo entre el vuelo lidar y el levantamiento de campo es un factor que contribuye a disminuir la sobredetección y aumentar la subdetección. Con los datos lidar tomados en 2007 se detectaron  $N_d$  número de árboles, magnitud que se compararía con los  $N_r$  número de árboles realmente presentes con diámetro normal mayor a 7.5 cm, si los datos de campo se hubieran tomado ese mismo año. Ya que se tomaron en 2010, es probable que  $N_r$  haya aumentado debido al crecimiento, y entonces si en un determinado sitio  $N_d < N_r$  (subdetección), esta diferencia es mayor con los datos tomados en 2010 que con los que se hubieran tomado en 2007, y si en un determinado sitio  $N_d > N_r$  (sobredetección), esta diferencia es menor que en el caso de haberse tomado los datos de campo en 2007.

Con respecto a la relación de la altura media y diámetro promedio con la detección, uno esperaría que el tamaño dominante de los árboles influya en la cantidad de detectados, que en zonas con árboles grandes (los cuales con mayor probabilidad podrían ocultar o semiocultar otros individuos, a comparación de los relativamente pequeños) fuera peculiar que se presentara subdetección, y en zonas con árboles pequeños, fuera más probable que la detección fuera más exacta o se presentara sobredetección; sin embargo, los análisis no muestran relación alguna. El tamaño promedio de los individuos no es necesariamente un factor determinante del desempeño del algoritmo de identificación. Sin embargo, sí hay relación entre la densidad de árboles y la detección, a mayor densidad hay más subdetección y a menor densidad hay más sobredetección, como se ha visto anteriormente en las figuras 4.5, 4.6 y 4.7 (que presentan número de árboles censados contra porcentaje de detección, pero ya que todas las muestras tienen la misma área, la densidad de árboles es proporcional al número de ellos), independientemente del tamaño de ellos.

La sobreestimación de altura presente en los resultados es lógica, pues al haber una mayor cantidad de sitios con subdetección, en los cuales típicamente se detectan los árboles mayores, entonces el promedio de altura detectado será mayor que el promedio de los censados, pues en estos incluyen más árboles pequeños. Y lógico es también el efecto de que al aumentar la precisión de la detección, disminuya el error en altura, pues al haber más árboles detectados correctamente naturalmente que el promedio de su tamaño será más cercano al promedio de tamaño de los efectivamente reales.

La densidad de puntos es crítica especialmente en aquellos sitios que presentan alrededor de 20 árboles censados. En el caso de presentar menos censados, se empieza a presentar sobredetección en varias

muestras, atribuida a árboles pequeños y arbustos; en el caso de presentar más censados, se empieza a presentar subdetección, atribuida a la densidad individuos que ocasiona que algunos no se puedan detectar. Entonces, considerando aquellos sitios con 15 a 25 censados, y considerando la densidad de 0.5 pts/m<sup>2</sup>, esto resulta en tener cerca de 200 puntos por muestra (cada muestra tiene un área de 400 m<sup>2</sup>), que implica un promedio de 10 retornos por árbol (aproximadamente) para caracterizarlos; sin embargo, se encuentran factores que disuelven este caso ideal:

- a) La presencia de numerosos árboles grandes. Dividiendo el área de la muestra entre 20 árboles, supuestos uniformemente distribuidos y de igual tamaño, sería visible un área de la copa de cada uno de ellos de 20 m<sup>2</sup>, esto implica que para poder ser visible la copa completa esta debería ser de un tamaño cuyo radio máximo sea de 2.5 m. Si se encuentran árboles de copa mayor, esto implicaría el tapar o invadir la copa adyacente, o necesariamente tener un determinado número de árboles de radio de copa menor a 2.5 m para poder mantener la visibilidad de todos.
- b) En muchas ocasiones no se alcanzan los 200 puntos por muestra.
- c) Difícilmente los puntos dentro de la muestra están distribuidos equitativamente en los árboles presentes en el sitio.

La presencia de estos factores siempre deja individuos con pocos retornos, difíciles de detectar.

En la tabla 4.2 se presenta una síntesis de algunos de los ensayos hechos respecto a la identificación de árboles con objeto de contrastarlos con el procesamiento utilizado en el presente trabajo.

La exactitud de la detección obtenida en el presente es pobre, cerca del 40 %, la que lo colocaría en la tabla en la última posición, en gran medida causada por la sobredetección y subdetección antes mencionada.

Como se ha mencionado antes, el algoritmo de agrupación y el método de identificación de árboles se tomaron del trabajo de Morsdorf et al. (2003), y por tanto comparar los resultados obtenidos con respecto a este, es una de las cuestiones inmediatas. Morsdorf y colegas obtuvieron una detección del 60 %, por encima del 42 % obtenido en esta tesis, con la diferencia de que ellos llevaron a cabo el proceso de reconocimiento de máximos locales sobre una superficie y no sobre la nube, además de contar con una densidad de puntos bastante superior.

<b>Autor</b>	<b>Densidad de puntos</b>	<b>Método de identificación de árboles</b>	<b>Método de segmentación</b>	<b>Exactitud de la detección</b>
Solberg et al. (2006)	5/m <sup>2</sup>	Máximos locales.	Construcción de región, incluyendo un algoritmo de control de forma.	93 % *
Roberts et al. (2005)	4-20/m <sup>2</sup>	Máximos locales.	Radios calculados en cuatro direcciones. Límite de la copa tomando el pixel de menor altura en cada dirección, dentro de una ventana de 3.4 m.	83 %
Heurich (2008)	10/m <sup>2</sup>	Ajuste de superficie parabólica al 30% más elevado de los datos.	Ajuste de superficie parabólica al 30% más elevado de los datos.	77 %
Vega y Durrieu (2011)	5.5/m <sup>2</sup>	Máximos locales.	Ajuste de superficie elíptica a la forma de la copa.	74 %
Persson et al. (2002)	5/m <sup>2</sup>	Máximos locales.	Agrupación de píxeles que suban hacia el mismo máximo.	71 %
Holmgren y Persson (2004)	5/m <sup>2</sup>	Obtenidos de la segmentación.	Superficie parabólica ajustada a las elevaciones del modelo digital del dosel.	71 %
Yu et al. (2011)	2.6/m <sup>2</sup>	Máximos locales.	Basado en llenado de cuencas.	69 %
Schardt et al. (2002)	4-5/m <sup>2</sup>	Máximos locales.	Basado en llenado de cuencas.	62 %
Hirata et al. (2009)	40.5/m <sup>2</sup>	Identificados en la segmentación.	Basado en llenado de cuencas.	60 % **
Morsdorf et al. (2003)	10/m <sup>2</sup>	Máximos locales.	Algoritmo de agrupación K-means.	60 %
Rahman et al. (2009)	70/m <sup>2</sup>	Ubicación del árbol en el punto con mayor densidad de puntos altos, dentro del área delimitada por los límites de la copa.	Delimitación basada en densidad de puntos altos, filtrado de la nube para dejar fuera aquellos puntos de vegetación de capas inferiores a la copa.	60 %
Pitkanen et al. (2004)	10/m <sup>2</sup>	Máximos locales en el modelo digital del dosel crudo.***	No se segmentan copas.	49 % ***

Tabla 4.2. Estudios forestales que incluyen la identificación de árboles individuales, ordenados en función de la exactitud de la detección. Coloreados se encuentran los métodos más comunes: rojo para construcción de región, anaranjado para delimitación basada en cuencas o similar, amarillo para el ajuste de alguna superficie a las copas, y azul para la identificación de árboles por máximos locales.

\*En este estudio se hace la distinción entre árboles dominantes, codominantes, subdominantes, suprimidos y muertos. El porcentaje de 93% es para los dominantes; el promedio para las 5 clases es del 44%.

\*\*En este estudio se lleva a cabo la detección en tres tipos de zonas: con mucho adelgazamiento (con tala de árboles para obtener un espaciamiento entre ellos amplio y regular), con adelgazamiento medio y sin adelgazamiento. El 60% es para la zona sin adelgazamiento.

\*\*\*Se prueban cinco métodos diferentes de identificación de árboles, el 49% es para el mejor método.

La mayoría de las propuestas mostradas en la tabla 4.2 realizan algún tipo de transformación de los datos de forma que no trabajan directamente sobre la nube de puntos. El único que no lo hace son es el estudio de Rahman et al. (2009), lo cual materializa una punto en común entre este y el presente trabajo. Rahman et al. (2009) proponen una alternativa bastante más sofisticada y con una densidad de

puntos superior, entregando resultados en la misma medida superiores al presente que los obtenidos por Morsdorf et al. (2003).

Con respecto a la densidad de puntos, el presente estudio se llevó a cabo con una densidad de puntos baja y la mayoría de los antes citados en la tabla 4.2 utilizan densidades de 5 pts/m<sup>2</sup> o superiores, lo cual los pone en ventaja. La propuesta más cercana en este sentido es la de Yu et al. (2011), con 2.6 pts/m<sup>2</sup>. Esta última también identifica árboles mediante máximos locales, aunque emplea un algoritmo de llenado de cuencas para segmentar las copas, presentando también una exactitud superior al trabajo de la presente tesis (69 % de individuos bien identificados).

El estudio de Solberg et al. (2006) es el único que utiliza un proceso de construcción de región en la segmentación y el que presenta mejores resultados. Sin embargo, la detección del 93 % es sólo tomando en cuenta los árboles dominantes, considerando todos los árboles dentro de su área de estudio, esta decae al 44 %, muy cercana a la precisión alcanzada en el presente ensayo y de igual forma empleando la identificación por máximos locales, a pesar de que Solberg et al. (2005) emplean una densidad de 5 pts/m<sup>2</sup>.

#### 4.2 ESTIMACIÓN DE CARBONO POR DISTRIBUCIÓN DE ALTURAS DEL DOSEL

Se generaron modelos de predicción para cantidad de carbono (Ton/Ha), número de árboles, altura media (m) y diámetro normal promedio (cm) para todos los tipos de vegetación; y modelos de predicción de cantidad de carbono para las coberturas oyamel y pino, que son las dominantes. Comparándolos con los valores obtenidos en campo se determinó la calidad de predicción. A continuación se muestran los modelos que incluyen todos los tipos de vegetación, primero (ecuaciones de la 4.1 a la 4.4), y los dos últimos de estimación de carbono para oyamel (ecuación 4.5) y pino (ecuación 4.6):

$$\text{Carbono(Ton / Ha)} = -14.3857 + 6.9860e^{0.09498 \cdot H90} + 7.2603e^{1378 \cdot \text{SIDrel}} \dots\dots\dots(4.1)$$

RMSE = 60.0051Ton / Ha  
R2 = 0.61

$$\text{No.Árboles} = 25.1456 + 9.8142(\text{SIDpesada}) - 29.6152(\text{Ps1}) - 11.0472(\text{SID2}) + 9.9540(\text{SIDtodos}) \dots\dots\dots(4.2)$$

RMSE = 9.7361árboles  
R2 = 0.50

$$\text{Alturamedia(m)} = 4.5048 + 0.5817(H2) \dots\dots\dots(4.3)$$

RMSE = 4.4543m  
R2 = 0.55

$$\text{Diámetronormalmedio(cm)} = 1.7060 + 21.9507(H1) - 20.0722(H95) + 19.3761(\text{P1s}) \dots\dots\dots(4.4)$$

RMSE = 9.9588cm  
R2 = 0.57

$$\text{Carbono(Ton / Ha)} = 14.0701 + 6.9393(\text{SID2}) + 4.0412e^{0.09216 \cdot H1} \dots\dots\dots(4.5)$$

RMSE = 42.9693Ton / Ha  
R<sup>2</sup> = 0.63

$$\text{Carbono(Ton / Ha)} = -2.3944 + 368.2161e^{1282 \cdot \text{SIDrelcorreg}} - 195.2750e^{1378 \cdot \text{SIDrel}} + 5.1234e^{0.09498 \cdot H90} \\ + 524.1922e^{-2.336 \cdot \text{Ps1}} + 0.00001e^{-16.83 \cdot \text{Ps1}} \dots\dots\dots(4.6)$$

RMSE = 40.7421Ton / Ha  
R<sup>2</sup> = 0.90

Donde:

H1 es la altura media de los primeros retornos vegetación (m).

H2 es la altura media de los primeros retornos vegetación mayores a 2 m (m).

H90 es la altura media del 90% más alto de los retornos vegetación (m).

H95 es la altura media del 95% más alto de los retornos vegetación (m).

SIDtodos es la suma de intensidades de todos los retornos vegetación entre el área (suma de intensidades/m2).

SID2 es la suma de intensidades de todos los retornos vegetación mayores a 2 m de altura, entre el área (suma de intensidades/m2).

SIDrel es la suma de intensidades de todos los retornos vegetación entre la suma de intensidades de todos los retornos, entre el área (suma de intensidades/m2).

SIDrelcorreg es la SIDrel corregida por reflectancia del suelo. Todas las intensidades de los retornos del suelo se multiplican por dos (suma de intensidades/m2).



SIDpesada es la suma de intensidades de todos los retornos vegetación mayores a 2 m de altura, en donde cada intensidad se multiplica por el inverso de su altura; todo dividido entre el área (suma de intensidades/m<sup>2</sup>).

Ps1 es la proporción de primeros retornos del suelo respecto a todos los primeros retornos (sin unidades).

P1s es la proporción de primeros retornos del suelo respecto a todos los retornos suelo (sin unidades).

Después de llevar a cabo la exploración de las relaciones entre las variables dependientes y las predictoras, se observó que únicamente el carbono presenta algún comportamiento no lineal con varias de las predictoras, 17 para ser exactos. Presenta relaciones con todas las variables caracterizadoras de altura y con la mayoría de las que caracterizan la densidad del follaje; fuera de esas, también presenta correlación con la desviación estándar de los primeros retornos vegetación (DE2) y con la proporción de primeros retornos del suelo con respecto a todos los primeros retornos (Ps1). Todas las ecuaciones que modelaron mejor estas relaciones fueron exponenciales, de la forma  $y = \alpha * \exp(\beta * x)$ , a excepción de la relacionada con Ps1, que fue de la forma  $y = \alpha * \exp(\beta * x) + \epsilon * \exp(\delta * x)$ , y todas ellas presentando un coeficiente de determinación de entre 0.4 y 0.5.

Cinco de las seis variables predictoras finales en los modelos de estimación de carbono sufrieron una transformación antes de utilizarse como argumentos de entrada en el proceso de regresión, ya que no presentaron una relación lineal con la cantidad de carbono en cada sitio. En las figuras 4.12 a 4.16 se muestran las gráficas de dispersión de cada una de ellas, a saber: H1, H90, SIDrel, SIDrelcorreg y Ps1, contra la cantidad de carbono, y la regresión exponencial que mejor se ajusta a los datos.

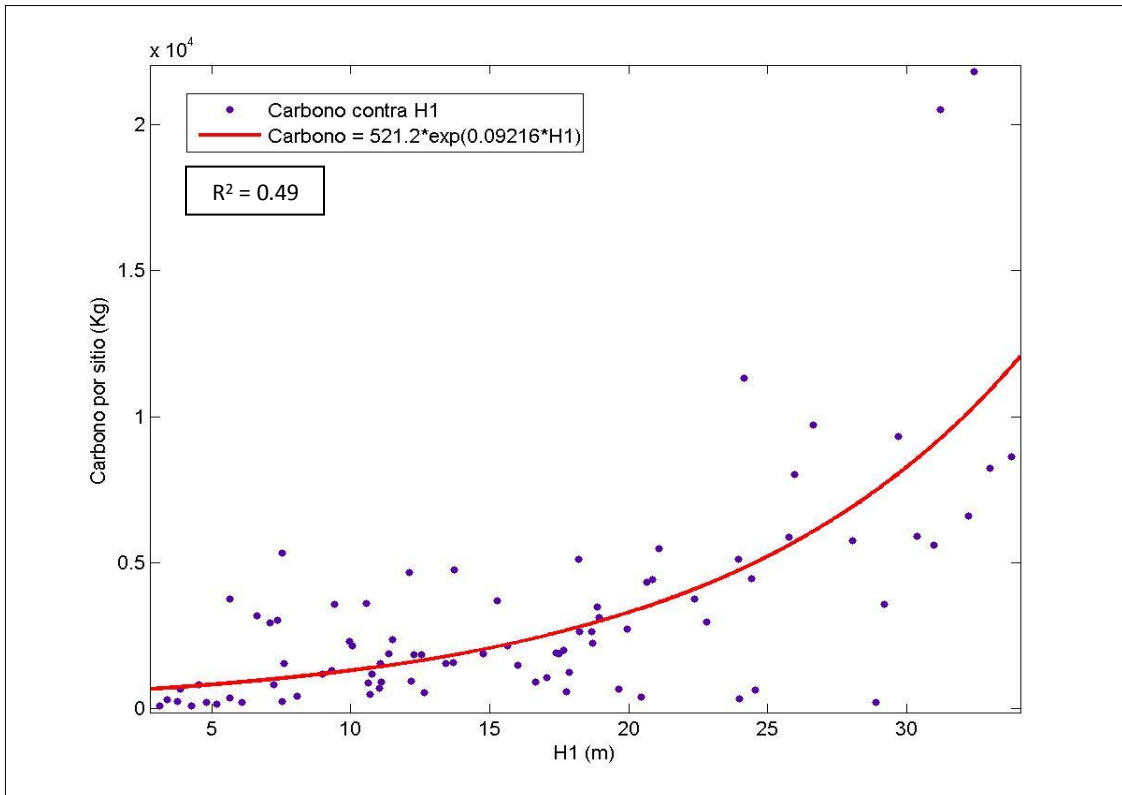


Figura 4.12. Gráfica de dispersión de la variable carbono contra la variable predictora H1. Se muestra la regresión exponencial (línea roja) y su coeficiente de determinación.

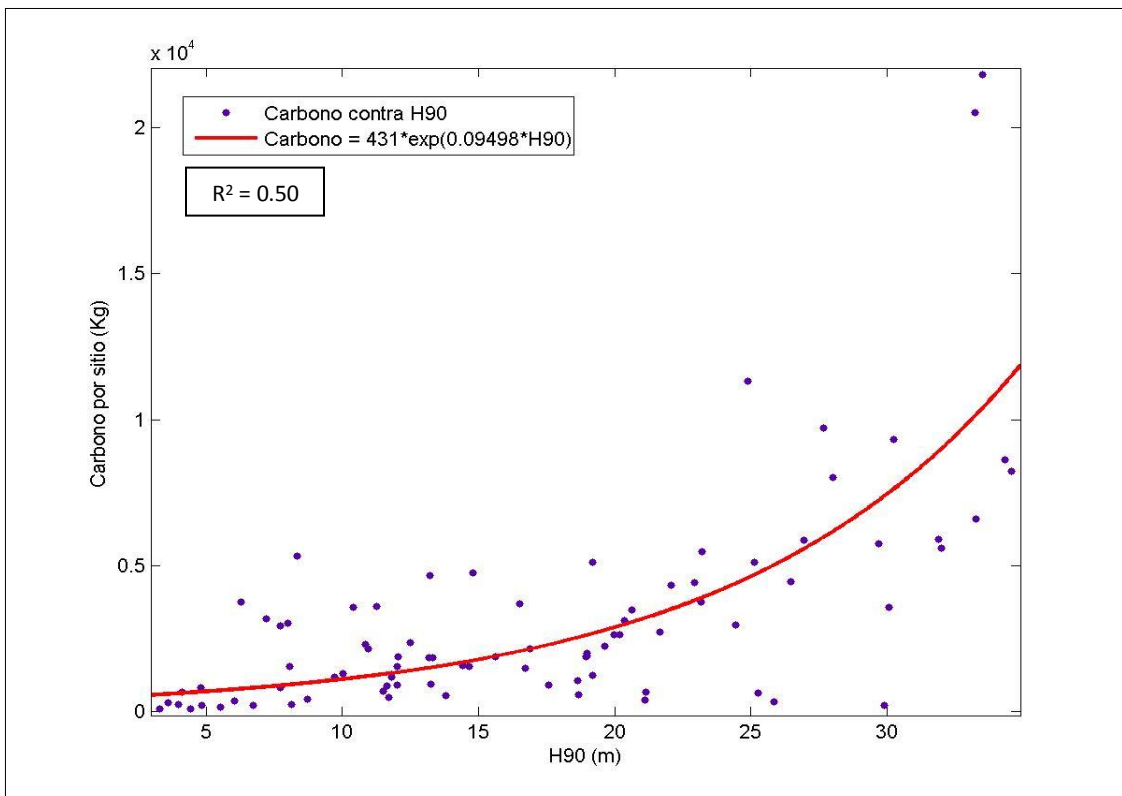


Figura 4.13. Gráfica de dispersión de la variable carbono contra la variable predictora H90. Se muestra la regresión exponencial (línea roja) y su coeficiente de determinación.

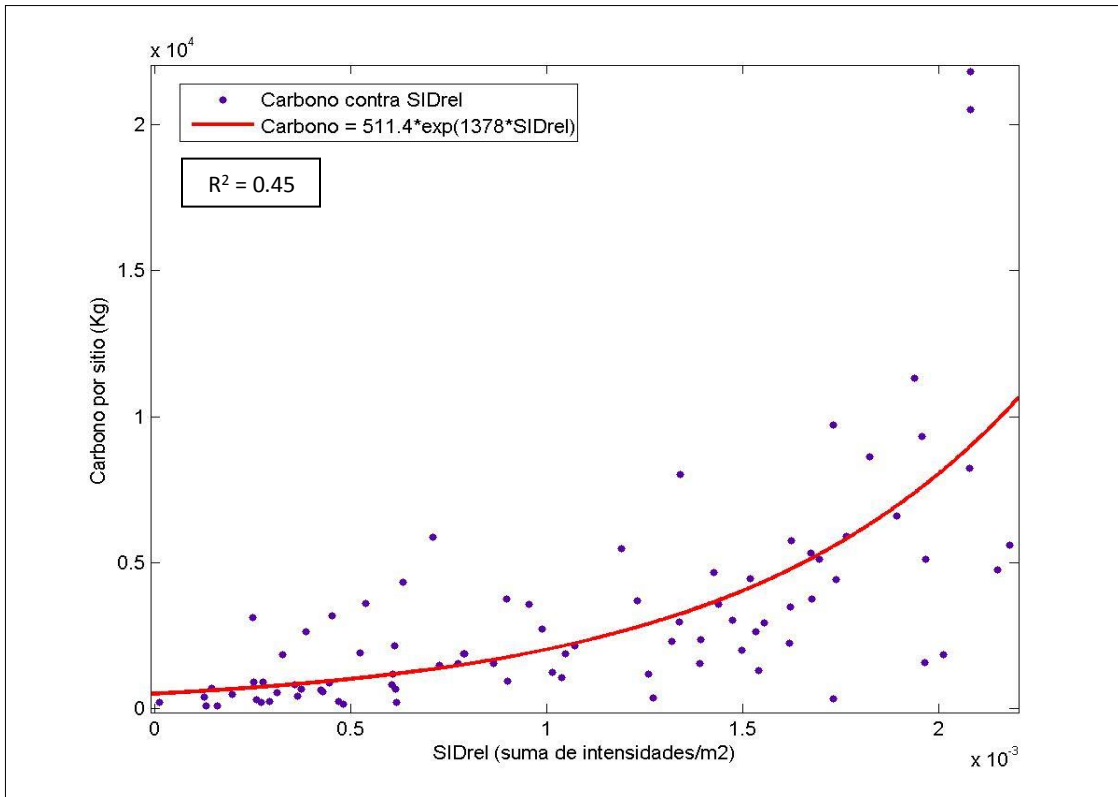


Figura 4.14. Gráfica de dispersión de la variable carbono contra la variable predictora SIDrel. Se muestra la regresión exponencial (línea roja) y su coeficiente de determinación.

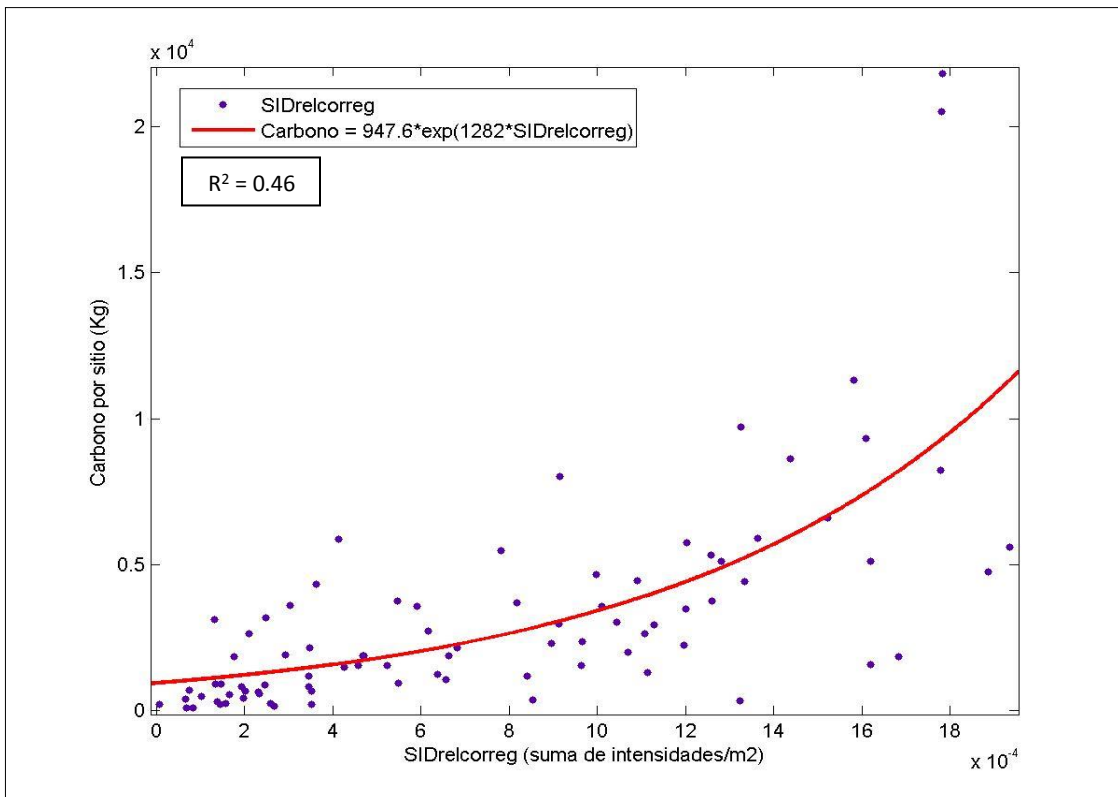


Figura 4.15. Gráfica de dispersión de la variable carbono contra la variable predictora SIDrelcorreg. Se muestra la regresión exponencial (línea roja) y su coeficiente de determinación.

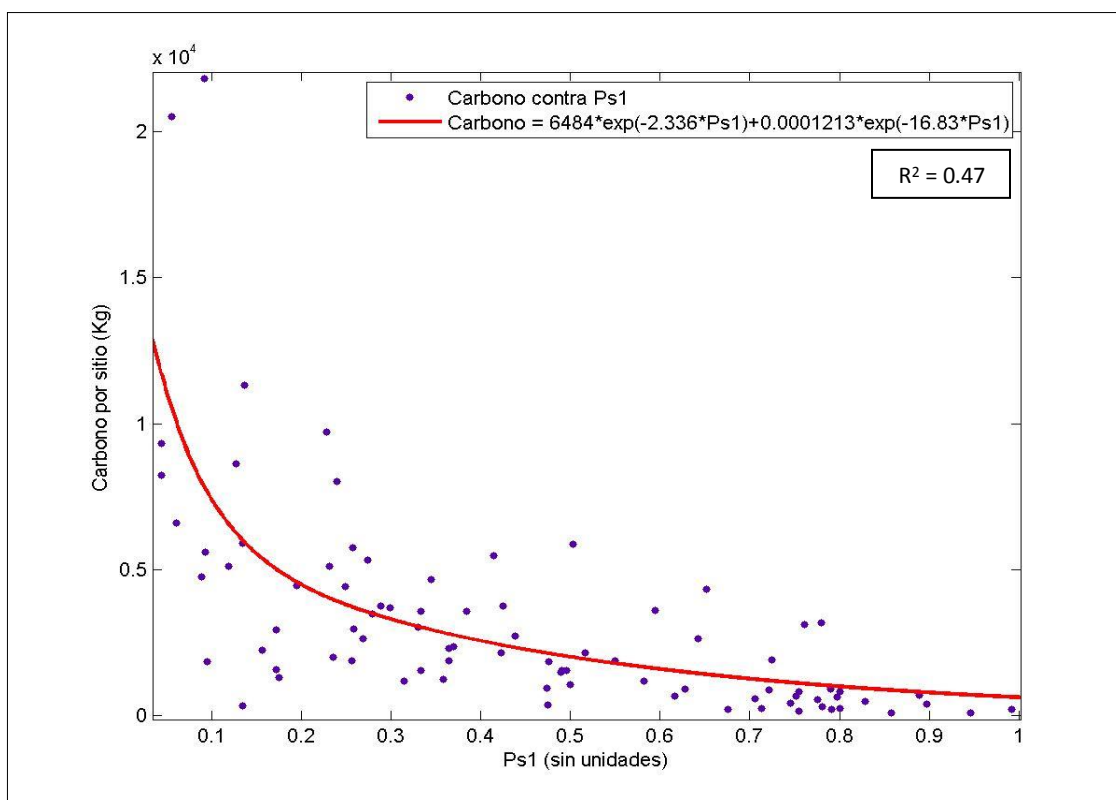


Figura 4.16. Gráfica de dispersión de la variable carbono contra la variable predictora Ps1. Se muestra la regresión exponencial (línea roja) y su coeficiente de determinación.

Las variables transformadas se construyeron entonces a partir de las ecuaciones que mejor modelaron esta relación:

$$H1^* = 521.2e^{0.09216 \cdot H1} \quad (4.7)$$

$$H90^* = 431e^{0.09498 \cdot H90} \quad (4.8)$$

$$SIDrel^* = 511.4e^{1378 \cdot SIDrel} \quad (4.9)$$

$$SIDrelcorreg^* = 947.6e^{1282 \cdot SIDrelcorreg} \quad (4.10)$$

$$Ps1^* = 6484e^{-2.336 \cdot Ps1} + 0.0001213e^{-16.83 \cdot Ps1} \quad (4.11)$$

Finalmente  $H1^*$ ,  $H90^*$ ,  $SIDrel^*$ ,  $SIDrelcorreg^*$  y  $Ps1^*$  sí se relacionaron linealmente con el carbono en el modelo de predicción, que al transformarse nuevamente a las variables originales  $H1$ ,  $H90$ ,  $SIDrel$ ,  $SIDrelcorreg$  y  $Ps1$  arrojaron los modelos 4.1, 4.5 y 4.6.

Los modelos de estimación de carbono incluyen una variable predictora de altura y otra de densidad de follaje, para el caso general y el de oyamel, y dos variables predictoras de altura y una de densidad de follaje, para el caso del pino, lo cual es coherente. Las tres magnitudes necesarias para calcular la

cantidad de materia de un cuerpo son el largo, ancho, alto (las cuales definen el tamaño del cuerpo) y su densidad, reflejadas también en el cálculo de la cantidad de carbono: la altura implica una magnitud y los límites establecidos de antemano de la muestra (circunferencia de 400 m<sup>2</sup>) implican las otras dos magnitudes que establecen el tamaño, además de la densidad del follaje. Es de esperarse que los modelos hubieran incluido alguna variable caracterizadora de la cantidad de cobertura del área, sin embargo, ya que la gran mayoría de los sitios no presenta muchas áreas abiertas, el algoritmo sólo incluyó una (Ps1) en la construcción del modelo de carbono para pino.

El modelo de número de árboles es el más complejo, incluye tres variables caracterizadoras de densidad de follaje y una de la cantidad de cobertura, en este caso sí se incluye una variable de este tipo; probablemente en este caso sea más crítico conocer la cantidad de áreas cubiertas y la cantidad de espacios abiertos con más precisión. El hecho de que haya tres predictoras que dependan de la cantidad de follaje puede explicarse por el grado de heterogeneidad de las muestras en cuanto a especies y forma de los árboles, pues estas características, e incluso la proporción en la que están presentes en determinada área influyen en el espaciamiento entre individuos y por lo tanto en la cantidad de ellos. Que el modelo no haya incluido ninguna variable predictora de altura, muestra que el número de árboles no depende tanto de la altura predominante en la zona.

Sobre el modelo de altura media no hay muchas consideraciones que hacer, la variable dependiente altura media presenta relaciones lineales con todas las predictoras de altura, y con la que presentó mayor correlación fue con H2, altura media de los primeros retornos vegetación mayores a 2 m, descartando los retornos más bajos que introducen ruido en la predicción.

El diámetro normal está relacionado con el tamaño del árbol, por lo tanto es lógico que existan dos variables predictoras de altura en su modelo de predicción. La inclusión de una variable de cantidad de cobertura vegetal se justifica por el hecho de que al ser la variable dependiente un valor que representa un promedio, es necesario introducir un índice de la cantidad de árboles presentes en el área. Como se observa en los modelos, ninguno de ellos incluyó alguna variable caracterizadora de la variabilidad de altura, lo cual manifiesta que, al menos en estas áreas, la presencia tanto de árboles pequeños como de grandes no tiene mucha influencia en la predicción de las variables dependientes.

Para evaluar el desempeño de los modelos se calculó el coeficiente de determinación ( $R^2$ ), el coeficiente de correlación de Pearson y la raíz del error medio cuadrático (RMSE, del inglés Root Mean Square Error) de cada uno, tanto para los datos de entrenamiento como para los de prueba. En el caso de los modelos

generales (los construidos para todos los tipos de vegetación en conjunto), estos índices también se calcularon por tipo de vegetación para el segmento de prueba. En la tabla 4.3 se muestran valores para los modelos generales y en la tabla 4.4 para los modelos de carbono para oyamel y pino.

Indicadores de la calidad de los modelos de predicción para todos los tipos de vegetación						
CARBONO						
	Segmento de entrenamiento	Segmento de prueba	Oyamel	Pino	Pastizal y matorral	Bosque mixto e inducido
R <sup>2</sup>	0.6091	0.5804	0.5424	0.7065	0.0648	0.0135
Coef. de Pearson	0.7804	0.7618	0.7364	0.8405	-0.2545	0.1163
RMSE	60.0051 Ton/Ha	47.7377 Ton/Ha	51.2635 Ton/Ha	39.2078 Ton/Ha	5.4906 Ton/Ha	79.1885 Ton/Ha
NÚMERO DE ÁRBOLES						
	Segmento de entrenamiento	Segmento de prueba	Oyamel	Pino	Pastizal y matorral	Bosque mixto e inducido
R <sup>2</sup>	0.4965	0.6368	0.6174	0.3371	0.1280	0.8192
Coef. de Pearson	0.7047	0.7980	0.7856	0.5806	-0.3578	0.9051
RMSE	9.7361 árboles	9.4056 árboles	10.6996 árboles	6.2238 árboles	9.7147 árboles	14.2795 árboles
ALTURA MEDIA						
	Segmento de entrenamiento	Segmento de prueba	Oyamel	Pino	Pastizal y matorral	Bosque mixto e inducido
R <sup>2</sup>	0.5528	0.5887	0.7087	0.3992	0.8058	0.0058
Coef. de Pearson	0.7435	0.7673	0.8418	0.6318	0.8977	-0.0762
RMSE	4.4543 m	4.2104 m	3.7443 m	5.0751 m	2.2133 m	4.0938 m
DIÁMETRO NORMAL MEDIO						
	Segmento de entrenamiento	Segmento de prueba	Oyamel	Pino	Pastizal y matorral	Bosque mixto e inducido
R <sup>2</sup>	0.5732	0.5640	0.5740	0.4967	0.0158	0.0422
Coef. de Pearson	0.7571	0.7510	0.7576	0.7048	-0.1258	-0.2056
RMSE	9.9588 cm	9.2933 cm	7.3814 cm	11.2727 cm	7.4876 cm	9.5163 cm

Tabla 4.3. Coeficiente de determinación (R<sup>2</sup>), coeficiente de correlación de Pearson y la raíz del error medio cuadrático (RMSE) de los modelos, tanto en el segmento de entrenamiento como de prueba para todos los sitios, y divididos por vegetación para el segmento de prueba únicamente. En anaranjado se somborean los valores menores a 0.4 y en verde los mayores a 0.70, de los coeficientes.

Para el caso de los modelos generales, no existe gran diferencia entre los índices calculados para el segmento de entrenamiento como para el segmento de prueba. Esto es indicador de consistencia y de que la distribución de los sitios en entrenamiento y prueba fue equilibrada, quedando grupos con características similares. Los índices son parecidos para cada caso, la mayor diferencia es en el número de árboles, en donde en el segmento de prueba el ajuste es mejor alrededor del 10 %. El RMSE

disminuye un poco para el segmento de prueba en relación al de entrenamiento para las cuatro variables dependientes.

<b>Indicadores de la calidad de los modelos de predicción de carbono para los tipos de vegetación oyamel y pino</b>				
	<b>Oyamel (Segmento de entrenamiento)</b>	<b>Oyamel (Segmento de prueba)</b>	<b>Pino (Segmento de entrenamiento)</b>	<b>Pino (Segmento de prueba)</b>
<b>R<sup>2</sup></b>	0.6254	0.3595	0.8985	0.6982
<b>Coef. de Pearson</b>	0.7908	0.5995	0.9479	0.8356
<b>RMSE</b>	42.9693 Ton/Ha	31.4810 Ton/Ha	40.7421 Ton/Ha	152.4280 Ton/Ha

Tabla 4.4. Coeficiente de determinación ( $R^2$ ), coeficiente de correlación de Pearson y la raíz del error medio cuadrático (RMSE) de los modelos de estimación de carbono para oyamel y pino. En anaranjado se somborean los valores menores a 0.4 y en verde los mayores a 0.70, de los coeficientes.

Observando la distribución de los índices entre los tipos de vegetación, se nota una tendencia en los resultados. En cada variable estimada los índices disminuyen para dos o tres tipos de vegetación, con respecto a los índices calculados para todas las coberturas en conjunto. En el caso del carbono sólo mejoran para el pino, en número de árboles sólo en bosque mixto e inducido, para diámetro normal sólo en oyamel se conservan los resultados, y en altura media en dos clases de vegetación aumentan (oyamel, y pastizal y matorral) y en las otras dos clases disminuyen. Claramente los modelos tienen mejor desempeño para sitios con oyamel y pino, lógicamente al ser éstos los sitios más numerosos y que más pesan en la regresión.

Para oyamel se tienen los mejores resultados, mejor que para pino, que son los sitios más numerosos incluso. Los índices para carbono y número de árboles disminuyen ligeramente con respecto a los índices para calculados todos los tipos de vegetación en conjunto, para diámetro normal se mantienen igual y para altura media aumentan en buen grado.

En los sitios con tipo de vegetación pino el desempeño de los modelos generales con respecto al desempeño para todos los sitios es mejor para carbono, para diámetro normal disminuye y para número de árboles y altura media disminuyen considerablemente. Es de notar que la variable de mayor importancia para el estudio, el carbono, y que con las especies más comunes, oyamel y pino, los modelos se desempeñen bien, lo cual hace rescatable esta propuesta de estimación.

Para la clase pastizal y matorral y la clase bosque mixto e inducido se obtuvieron resultados malos. Cada una de estas clases presentó buen desempeño sólo en una de las cuatro variables a estimar, altura

media para el caso de pastizal y matorral y número de árboles para el caso de bosque mixto e inducido, ambos con  $R^2$  de 0.8 y coeficiente de Pearson de 0.9. En todos los demás modelos se mostraron predicciones muy pobres, esto debido a que los sitios con estos tipos de vegetación son minoría (39, comparados con los 131 de oyamel y pino), teniendo poco peso en la construcción de la regresión.

En los modelos de carbono construidos para oyamel y pino por separado (ecuaciones 4.5 y 4.6, respectivamente, tabla 4.4), el  $R^2$  y el coeficiente de Pearson calculados para el segmento de prueba estuvieron por debajo de los calculados para el segmento de entrenamiento, con diferencias de 0.27 ( $R^2$ ) y 0.19 (Pearson) para oyamel y 0.20 ( $R^2$ ) y 0.11 (Pearson) para pino. EL RMSE fue bastante mayor en el segmento de prueba que en el de entrenamiento para pino (diferencia de 111.69 Ton/Ha). El único índice donde hubo una mejora en el segmento de prueba fue en el RMSE de oyamel, que presentó 31.48 Ton/Ha, en comparación del 42.97 Ton/Ha para el grupo de entrenamiento. Esto manifiesta poca consistencia en el comportamiento de los modelos.

Los modelos de carbono para oyamel y pino no mejoraron consistentemente el desempeño del modelo general de carbono. Tomando como parámetro el  $R^2$ , el valor de este índice para el segmento de entrenamiento del modelo de oyamel (0.6254) mejoró con respecto al de entrenamiento para el modelo general (0.6091) con una diferencia de 0.02, aunque en el segmento de prueba (0.3595) fue más bajo por 0.18 con respecto al del modelo general calculado para oyamel (0.5424). En el caso del pino, el  $R^2$  para el segmento de entrenamiento (0.8985) mejoró con respecto al de entrenamiento para el modelo general (0.6091) en 0.29, pero los valores para el segmento de prueba fueron prácticamente iguales (0.6983 para el modelo de pino, 0.7065 para el modelo general calculado para pino).

En la literatura es difícil encontrar propuestas de estimación de almacenamiento de carbono en específico utilizando alguna propuesta de distribución de las alturas del dosel, por esto, se hace una comparación con algunos estudios que estiman biomasa, por ser una variable similar al carbono, pues en ambos casos se trata de parámetros de masa. En la tabla 4.5 se muestran cuatro trabajos de generación de modelos de estimación de biomasa, cada uno diferente; considerando que el  $R^2$  global de estimación de carbono en este caso es de 0.58 para el segmento de prueba, éste se colocaría debajo de los mejores resultados de la tabla.



Autor	Densidad de puntos	Variables predictoras	Método de obtención de los modelos de estimación	R <sup>2</sup> del modelo de estimación de biomasa
Hall et al. (2005)	1.2/m <sup>2</sup>	Estadísticas de la altura, intensidad y tipo de retorno	Regresión múltiple	0.74
Jochem et al. (2010)	0.9 – 2.7/m <sup>2</sup>	Volumen del dosel, calculado a partir de la nube y de los datos rasterizados	Regresión lineal simple	Biomasa calculada a partir de la nube, 0.70 Biomasa calculada a partir de los datos rasterizados, 0.72
Bortolot (2006)	1/m <sup>2</sup>	Porcentaje de píxeles perteneciente a grupo de árboles, porcentaje de píxeles interiores de grupo, altura media del dosel y desviación estándar de la altura del dosel	Regresión lineal múltiple (2 variables predictoras)	De 0.04 a 0.62
Lefsky et al. (1999)	Datos de onda completa	Variables de altura	Regresión univariable	0.80 para el segmento de entrenamiento, 0.33 para el segmento de prueba

Tabla 4.5. Estudios de estimación de biomasa basados en la distribución de las alturas del dosel, ordenados de acuerdo al valor más alto de su coeficiente de determinación. En la mayoría de ellos no sólo se estima la biomasa, sino también otros parámetros, sin embargo no se muestran.

Cada uno de los estudios mostrados en la tabla 4.5 tiene sus peculiaridades. En el caso de Lefsky et al. (1999) se utilizan datos lidar de onda completa, no discretos, de los cuales se obtienen una serie de variables de altura y su variabilidad, para relacionarlas una a una con área basal y biomasa. La variable predictora de mejor respuesta fue la media cuadrática de la altura del dosel –una variable de altura– de manera similar al modelo de carbono propuesto, pues también incluye una variable de altura (altura media del 90% más alto de los retornos vegetación). Es de esperarse que un modelo multivariable en la mayoría de los casos podría ofrecer mejores estimaciones, sin embargo, en el trabajo de Lefsky et al. (1999) se prueba utilizar una regresión stepwise sin mejorar los resultados; en el caso de la propuesta trabajada en la presente tesis se partió del hecho de que una regresión múltiple sería lo más conveniente, sin reparar en generar modelos univariados, obteniendo un mejor R<sup>2</sup>. Llama la atención la gran diferencia entre la R<sup>2</sup> del segmento de entrenamiento y la del segmento de prueba en el estudio de Lefsky et al. (1999), esto es debido a que las características del segmento de entrenamiento tienen un espectro de variación más alto que las de prueba (Lefsky et al., 1999:83), caso contrario a los resultados mostrados anteriormente (tabla 4.3), que presentan consistencia en ambos segmentos.

Bortolot (2006) propone una propuesta “intermedia” entre la identificación de árboles individuales y la distribución de alturas del dosel, pues identifica conglomerados de árboles en los datos convertidos a

raster. A partir de esos grupos genera un par de variables predictoras poco usuales: el porcentaje de píxeles pertenecientes a grupos de árboles y el porcentaje de píxeles interiores a los grupos, que son variables en función de áreas, sin tomar en cuenta la altura, lo cual es una diferencia fundamental con respecto al presente estudio. Finalmente construye modelos lineales de dos variables predictoras. En el caso del presente trabajo no se identifican árboles de ninguna manera, se emplean una serie de variables predictoras más sofisticadas y no se limitan los modelos de predicción a dos variables independientes, aún así los modelos con mejor  $R^2$  de Bortolot (2006) se equiparan con el conseguido por este.

El caso de Jochem et al. (2010) es una propuesta más simple, únicamente se limita a calcular el volumen de la vegetación por área a partir tanto de la nube directamente como de los datos convertidos a raster, para después proponer un modelo de relación de este con biomasa (regresión simple). Existe cierta coherencia entre el trabajo de Jochem et al. (2010) y el modelo propuesto en el presente, pues ambos toman en cuenta el volumen de la vegetación (a final de cuentas, si de una parcela circular se obtiene una variable de altura media [H90], esto representa un volumen), no obstante con el agregado, en el caso del presente, de una variable caracterizadora de densidad de follaje (SIDrel). La propuesta de Jochem et al. (2010) es bastante más sencilla y obtiene mejores resultados que los conseguidos por la de esta tesis, sin embargo, esto se atribuye en gran medida a la mayor heterogeneidad de los datos de campo de la presente; en el caso de Jochem et al. (2010) la presencia de Pícea Noruega en el área de estudio es del 96 %.

El estudio de Hall et al. (2005) es el único de los presentados que trabaja completamente con datos de la nube, tiene una cantidad de variables más extensa que toma en cuenta también los valores de intensidad de los retornos y es el que presenta la mayor precisión para la estimación de biomasa. Estas son razones para haber elegido las mismas variables de éste como entrada en la regresión propuesta en la presente tesis. Al comparar los  $R^2$  obtenidos, el trabajo de Hall et al. (2005) presenta considerablemente mejores resultados (0.74 contra 0.58), empero su área de estudio es más homogénea: presencia de sólo dos especies –pino ponderosa y abeto douglas– con altura media promedio de 12.6 m, altura mínima promedio de 8.1 m y altura máxima promedio de 17.2 m, por unidad de muestreo; mientras que en el presente se censaron 23 especies (aunque dos son dominantes, oyamel y pino, en el 77 % de las muestras), con una altura media promedio de 13.3 m, una altura mínima promedio de 1.3 m y una altura máxima promedio de 31.1 m.

## 5. CONCLUSIONES

### 5.1 IDENTIFICACIÓN DE ÁRBOLES INDIVIDUALES

Se propuso una segmentación de árboles basada en un algoritmo de agrupación, k-means, identificando los individuos por máximos locales utilizando una ventana de búsqueda dinámica sobre la nube de puntos, como propuesta a ser insertada en alguna metodología de estimación de parámetros forestales. El resultado de este método es que detecta el 42 % de los árboles correctamente, habiendo mayor número de muestras donde se detectan alrededor de la mitad de los árboles y menor número de muestras en donde se detectan más de los existentes, con errores de comisión en muchos casos mayores al 50 % (se detectan más del doble de los censados).

Por todo esto, se concluye que la segmentación de árboles individuales, como candidata a ser integrada a una metodología de estimación de variables forestales, con datos lidar de baja densidad de puntos (0.5 pts/m<sup>2</sup>) y trabajada directamente sobre la nube, no es viable, pues no entrega buenos resultados; la detección del 42 % está por debajo del 60 % de los trabajos ya hechos anteriormente (Morsdorf et al., 2003; e Hirata et al., 2009). No es confiable obtener alguna estimación de densidad de árboles, altura por área, o como nos interesa en este caso, cantidad de carbono, cuando se están detectando correctamente menos de la mitad de los árboles realmente existentes en campo.

Un eje fundamental en el desempeño de los métodos es la densidad de puntos, especialmente para las propuestas de segmentación de árboles; la mayoría de los estudios de este tipo utilizan datos con valores de 5 pts/m<sup>2</sup> o superiores. Aquí se utilizó una baja densidad, lo cual es una desventaja, y deben tomarse con precaución los resultados, especialmente cuando se comparan con los de otros trabajos. Los errores en la estimación de altura media promedio, por ejemplo, pierden valor representativo al no tener una identificación satisfactoria.

Una situación que debe ser analizada más afondo para develar sus causas, es el hecho de que las muestras para tipo de vegetación matorral hayan tenido los peores resultados y las muestras para pastizal hayan tenido los mejores, siendo ambos tipos de vegetación muy similares en cuanto a número de árboles y altura. Es probable que los resultados estén relacionados con el bajo número de muestras de estas vegetaciones, quizá contando con más sitios los elementos que provocan estas diferencias de desempeño pudieran haber sido evidentes.

Existen varias oportunidades de mejorar el proceso de segmentación aquí propuesto, especialmente en aquellas zonas abiertas en donde se obtuvieron muchos errores de comisión. Podrían hacerse modificaciones en la forma de identificar los máximos, en los criterios de elección de tamaño de la ventana de búsqueda dinámica, podrían establecerse filtros para descartar los máximos que probablemente no representen árboles o incluso refinar la identificación después de llevar a cabo la agrupación mediante el análisis de la forma de los grupos ya formados.

En cuanto a las zonas donde se presentó subdetección, la mejora de los resultados es más complicada. Muy probablemente la forma más inmediata y adecuada para hacerlo es mediante el aumento de la densidad de puntos, elemento que queda fuera de los alcances de este método. En general para el aumento de la precisión de la detección una alternativa viable podría ser la división de las muestras por tipo de vegetación y la calibración de los parámetros de los algoritmos en función de esta.

Sería interesante además probar con otros métodos de segmentación usando los mismos datos lidar y de campo, transformándolos a raster o imagen, o incluso trabajando conjuntamente con imágenes satelitales. Esto podría proporcionar una evaluación más objetiva de las fortalezas y debilidades de la propuesta presentada en este documento.

## 5.2 ESTIMACIÓN DE CARBONO POR DISTRIBUCIÓN DE ALTURAS DEL DOSEL

Mediante 25 variables predictoras caracterizadoras de la altura del dosel, la variabilidad de altura, la densidad del follaje y la cantidad de cobertura vegetal se diseñaron modelos de estimación de carbono, número de árboles, altura media y diámetro normal promedio por muestra, mediante la aplicación de una regresión múltiple stepwise. Las variables predictoras utilizadas estuvieron basadas en el estudio de Hall et al. (2005), obtenidas directamente de la nube y presentando buenos resultados.

Los coeficientes de determinación en el segmento de prueba de los modelos generales fueron 0.58, 0.64, 0.59 y 0.56 para carbono, número de árboles, altura media y diámetro normal medio, respectivamente, mientras que para los modelos de carbono para oyamel y pino fueron de 0.36 y 0.70, respectivamente. Carbono es el parámetro que mayor interés reviste, y para los tipos de vegetación predominante tuvo  $R^2$  de 0.54 (oyamel) y 0.71 (pino) calculados con los modelos generales. Las variables predictoras finales incluidas en los modelos fueron coherentes en cuanto a su relación con las dependientes, y se encontró consistencia en el comportamiento de los modelos generales en los segmentos de entrenamiento y de prueba, y poca consistencia en los modelos de carbono para oyamel y pino.

Se consideran con buen desempeño el modelo general de estimación de carbono (ecuación 4.1) utilizado sólo para tipo de cobertura pino ( $R^2$  de 0.71), el modelo de estimación de carbono para pino (ecuación 4.6) ( $R^2$  de 0.70) y el modelo general de estimación de altura media (ecuación 4.3) utilizado sólo para tipo de cobertura oyamel ( $R^2$  de 0.71). En la estimación de carbono para pino, se considera mejor el modelo general que el diseñado en específico para pino, pues tiene un  $R^2$  mayor (0.71 contra 0.70), tiene menos variables predictoras (2 contra 4) y menor RMSE (39.21 Ton/Ha contra 152.43 Ton/Ha, en el segmento de prueba).

Por lo anterior, se considera factible la estimación del carbono para pino y de altura media para oyamel, utilizando los modelos aquí propuestos, con una densidad mínima de 0.5 pts/m<sup>2</sup> e incluyendo datos de intensidad y número de retorno. Los resultados muestran que los modelos explican el 71 % de la varianza de los parámetros forestales estimados y la correlación entre éstos y las variables predictoras es del 84 %. El buen desempeño de los modelos de altura media para pastizal y matorral y número de árboles para bosque mixto e inducido se deben tomar con precaución por el bajo número de muestras en estos tipos de vegetación.

La heterogeneidad de la muestra mina las posibilidades de obtener modelos con mayor poder predictivo. En el caso de los modelos generales esto es una desventaja pues las especies que son minoría introducen ruido en el proceso de selección de las variables y construcción de los modelos. Sólo para oyamel y pino se construyeron modelos específicos y los resultados no fueron consistentemente mejores, probablemente contando con un número de muestras mayor sí lo sean. Para el caso de las demás especies, el número de muestras no permitió construir modelos específicos.

Por otro lado, probar otro método de selección de variables predictoras podría también mejorar los modelos ya que el stepwise es un método sensible a las variables iniciales incluidas; incluso podrían probar modelos más simples, que podrían entregar mejores estimaciones.

## 6. BIBLIOGRAFÍA

- Acosta M., M., Vargas H., J., Velázquez M., A., y Etchevers B., J. D., 2002, Estimación de la biomasa aérea mediante el uso de relaciones alométricas en seis especies arbóreas en Oaxaca, México. *Agrociencia*, **36**(6), 725-736.
- Acuña F., E., 2011, *Análisis de Regresión*, Departamento de Matemáticas, Recinto Universitario de Mayaguez, Universidad de Puerto Rico, Puerto Rico, 245 p.
- Arellano, J., Chávez, J., y Durón, J., 2005, *Línea de producción LiDAR*, Aguascalientes, Aguascalientes, México, INEGI, 27 p.
- Avendaño H., D. M., Acosta M., M., Carrillo A., F., y Etchevers B., J. D., 2009, Estimación de biomasa y carbono en un bosque de Abies religiosa. *Fitotécnica Mexicana*, **32**(3), 233-238.
- Bortolot, Z. J., 2006, Using tree clusters to derive forest properties from small footprint lidar data. *Photogrammetric Engineering & Remote Sensing*, **72**(12), 1389-1397.
- Bortolot, Z. J., y Wynne, R. H., 2005, Estimating forest biomass using small footprint LiDAR data: an individual tree-based approach that incorporates training data, *ISPRS Journal of Photogrammetry and Remote Sensing*, **59**(6), 342-360.
- Brantberg T., Warner, T., Landenberger, R., y McGraw, J., 2003, Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North América, *Remote Sensing of Environment*, **85**, 290-303.
- The British Columbia for Applied Remote Sensing, Modeling and Simulation (BC-CARMS), 2006, *LiDAR – Overview of Technology, Applications, Market Features & Industry*, University of Victoria, Victoria, Canada, 49 p.
- Burtch, R., 2002, Lidar Principles and Applications, Proceedings IMAGIN Conference, The Center for Photogrammetric Training, Ferris State University, Michigan, 13 p.
- Comisión Nacional Forestal (CONAFOR), S/F, *Manual y procedimientos para el muestreo de campo para el Inventario Nacional Forestal y de Suelos*. México. 140 p.
- Cowie, A., Schneider, U., y Montanarella, L., 2007, Potential synergies between existing multilateral environments agreements in the implementation of land use, land use change & forestry activities. *Environmental Science & Policy*, **10**, 353-352.
- Draper, N. R., y H. Smith. 1998, *Applied Regression Analysis*, Hoboken, New Jersey, Wiley-Interscience, 706 p.
- Fernandez, J., Singhanía, A., Caceres, J., Slatton, K., Starek, M., y Kumar, R., 2007, *An Overview of Lidar Point Cloud Processing Software*, Geosensing Engineering and Mapping (GEM), Civil and Coastal Engineering Department, University of Florida, Florida, 27 p.
- Gill, S.J., Biging, G.S., y Murphy, E.C., 2000, Modeling conifer tree crown radius and estimating canopy cover. *Forest Ecology and Management*, **126**, 405–416.
- Hall, S. A., Burke, I. C., Box, D. O., Kaufmann, M. R., y Stoker, J. R., 2005, Estimating stand structure using discrete return lidar: an example for low density, fire prone ponderosa pine forests. *Forest Ecology and Management*, **208**(1-3), 189-209.
- Heinzel, J., y Koch, B., 2011, Exploring full-waveform LiDAR parameters for tree species classification, *International Journal of Applied Earth Observation and Geoinformation*, **13**(1), 152-160.

- Heurich, M., 2008, Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management*, **255**(7), 2416-2433.
- Hirata, Y., Furuya, N., Suzuki, M., y Yamamoto, H., 2009, Airborne laser scanning in forest management: individual tree identification and laser pulse penetration in a stand with different levels of thinning, *Forest Ecology and Management*, **258**(5), 752-760.
- Hollaus, M., Mücke, W., Höfle, B., Dorigo, W., Pfeifer, N., Wagner, W., Bauerhansl, C., y Regner, B., 2009, Tree species classification based on full-waveform airborne laser scanning data, In Proceedings of Silvilaser 2009, College Station, Texas, USA, 14–16 October 2009, 54-62.
- Holmgren, J., y Persson, A., 2004, Identifying species of individual trees using airborne laser scanning. *Remote Sensing of Environment*, **90**(4), 415-423.
- Hyypä, J., y Inkinen, M., 1999, Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, **16**, 27-42.
- Hyypä, J., Mielonen, T., Hyypä, H., Maltamo, M., Yu, X., Honkavaara, E., y Kaartinen, H., 2005, Using individual tree crown approach for forest volumen extraction with aerial images and laser point clouds. In Proceeding of ISPRS Workshop Laser Scanning 2005, September 12-14, 2005, Enschede, The Netherlands, **XXXVI**, Part 3/W19, 144-149.
- Hyypä J., Hyypä, H., Leckies, D., Gougeons, F., Yu, X., y Maltamo, M., 2008, Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forest, *International Journal of Remote Sensing*, **29**(5), 1339-1366.
- Intergovernmental Panel on Climate Change (IPCC), 2007, *Climate Change 2007: Synthesis Report*. Geneva, Switzerland, 104 p.
- Jenkins, L. G., 2006, Key drivers in determining lidar sensor selection. 5th FIG Regional Conference—Promoting Land Administration and Good Governance, Accra, Ghana, March 8–11, 2006.
- Jochem, A., Hollaus, M., Rutzinger, M., Höfle, B., Schadauer, K., y Maier, B., 2010, *Estimation of aboveground biomass using airborne LiDAR data*. Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Vienna, Austria, 9 p.
- Kato, A., Moskal, L. M., Schiess, P., Swanson, M. E., Calhoun, D., y Stuetzle, W., 2009, Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment*, **113**(6), 1148-1162.
- Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., y Shugart, H. H., 1999, Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA, *Remote Sensing of Environment*, **67**(1), 83-98.
- Lim, K., Treitz, P., Wulder, M., St.Onge, B., y Flood, M., 2003, LIDAR remote sensing of forest structure. *Progress in Physical Geography*, **27**(1), 88–106.
- Lohani, B., 2008, *Airborne Altimetric LiDAR: Principle, Data Collection, Processing and Applications*, Indian Institute of Technology, Kanpur, India, 32 p.
- Malleux O., J., 1970, Estudio de la relación D.A.P. con el diámetro de copa en un bosque húmedo subtropical. *Revista Forestal del Perú*, **4**(1-2), 1-5.

- Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., y Pitkänen, J., 2005, Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, **216**, 41–50.
- Maltamo, M., Eerikäinen, K., Packalen, P., y Hyyppä, J., 2006a, Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, **79**, 217–229.
- Moffiet, T., Mengersen, K., Witte, C., King, R., y Denham, R., 2005, Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing*, **59**, 289–309.
- Morsdorf, F., Meier, E., Allgöwer, B., y Nüesch, D., 2003, Clustering in airborne laser scanning raw data for segmentation of single trees. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **34**, Part 3/W13, 27–33.
- Næsset, E., y Økland, T., 2002, Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, **79**, 105–115.
- National Center for Policy Analysis (NCPA), 2007, *A Global Warming Primer*, Estados Unidos, 39 p.
- Nilsson, M., 1996, Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, **56**, 1–7.
- Ørka, H. O., Næsset, E., y Bollandsås, O. M., 2009, Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, **113**(6), 1163–1174.
- Persson, Å., Holmgren, J., y Söderman, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*, **68**(9), 925–932.
- Pfeiffer, N., y Briese, C., 2007, Geometrical aspects of airborne laser scanning and terrestrial laser scanning. *International Archives of Photogrammetry and Remote Sensing*, **36**(3-W52), 311–319.
- Pitkänen, J., Maltamo, M., Hyyppä, J., y Yu, X., 2004, Adaptive methods for individual tree detection on airborne laser based height model. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XXXVI**(8/W2), 187–191.
- Popescu, Sorin C., 2007, Estimating biomass of individual pine trees using airborne lidar. *Biomass & Bioenergy*, **32**, 646–655.
- Popescu, S., Wynne, R., y Nelson, R., 2003, Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, **29**(5), 564–577.
- Procuraduría Ambiental y de Ordenamiento Territorial del Distrito Federal, 2010, *Estimación de captura de carbono como indicador del estatus del derecho de los habitantes del Distrito Federal a gozar de áreas verdes urbanas adecuadas para su desarrollo, salud y bienestar*. México, 48 pags.
- Puttonen, E., Jaakkola, A., Litkey, P., y Hyyppä, J., 2011, Tree classification with fused mobile laser scanning and hyperspectral data. *Sensors 2011*, **11**, 5158–5182.
- Quinto M., H., y Moreno H., F., 2010, Crecimiento de árboles en un bosque pluvial tropical del chocó y sus posibles efectos sobre las líneas de energía. *Revista de Biología e Ciências da Terra*, **10**(2), 12–28.
- Pyysalo, U., y Hyyppä, H., 2002. Reconstructing tree crowns from laser scanner data for feature extraction. In ISPRS Commission III, Symposium 2002, September 9 - 13, 2002, Graz, Austria, 4 p.



- Rahman, M.Z.A., Gorte B. G. H., y Bucksch A.K., 2009, *A new method for individual tree delineation and undergrowth removal from high resolution airborne lidar*, Delft Institute of Earth Observation and Space Systems (DEOS), Delft University of Technology, Países Bajos, 6 p.
- Real, E. A., 2011, *El modelado geomático del lidar: de la fusión svm a la noción de prosdetección*, Tesis de Maestría, Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo" A.C., México, 74 p.
- Roberts, S. D., Dean, T. J., Evans, D. L., McCombs, J. W., Harrington, R. L., y Glass, P. A., 2005, Estimating individual tree leaf area in loblolly pine plantations using LiDAR-derived measurements of height and crown dimensions, *Forest Ecology and Management*, **213**(1-3), 54-70.
- Schardt, M., Ziegler, M., Wimmer, A., Wack, R., y Hyypä, J., 2002, Assessment of forest parameters by means of laser scanning. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, **34**, 302-309.
- Shan, J., y Toth, C. K., 2009, *Topographic laser ranging and scanning: principles and processing*, Estados Unidos, Taylor & Francis Group, 590 p.
- Simard, S. W., y Austin, M. E., 2010, *Climate Change and Variability*, Croacia, Sclyo, 486 p.
- Secretaría del Medio Ambiente del Distrito Federal (SMA), 2007, *Agenda Ambiental de la Ciudad de México. Programa de Medio Ambiente, 2007-2012*. México, Distrito Federal, 180 p.
- Secretaría del Medio Ambiente y Recursos Naturales (SEMARNAT), 2005, *Informe de la Situación del Medio Ambiente en México*.  
[http://app1.semarnat.gob.mx/dgeia/informe\\_04/02\\_vegetacion/recuadros/c\\_rec1\\_02.htm](http://app1.semarnat.gob.mx/dgeia/informe_04/02_vegetacion/recuadros/c_rec1_02.htm)  
 Visitado 28/05/2012
- Silván-Cárdenas, J. L., y Wang, L., 2006, A multi-resolution approach for filtering LIDAR altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **61**, 11-22.
- Solberg, S., Næsset, E., y Bollandsas, O., 2006, Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering and Remote Sensing*, **72**(12), 1369–1378.
- Starek, M., y Slatton, K. C., 2005, *Basic Gridding of ALSM Point Data*, Geosensing Engineering and Mapping (GEM), Civil and Coastal Engineering Department, University of Florida, Florida, 8 p.
- Suratno, A., Seielstad, C., y Queen, L., Mapping Tree Species Using LIDAR in Mixed-Coniferous Forests, Proceedings of Silvilaser 2009, 14-16 October, National Center for Landscape Fire Analysis, College of Forestry and Conservation, The University of Montana, 10 p.
- Véga, C., y Durrieu, S., 2011, Multi-level filtering segmentation to measure individual tree parameters based on Lidar data: application to a mountainous forest with heterogeneous stands. *International Journal of Applied Earth Observation and Geoinformation*, **13**(4), 646-656.
- Vitousek, P.M., 1994, Beyond global warming: ecology and global change. *Ecology*, **75**(7), 1861-1876.
- Wagner, W., Ullrich, A., Melzer, T., Briese, C., y Kraus, K., 2004, From single-pulse to full-waveform airborne laser scanners: potential and practical challenges, Proceedings of International Society for Photogrammetry and Remote Sensing, **XXXV**, Part B/3, 12-23 July 2004, Commission 3, Istanbul, Turkey, 6 p.
- Wehr, A., y Lohr, U., 1999. Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**(2/3), 68-82.

- Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., y Viitala, R., 2011, Predicting individual tree attributes from airborne laser point clouds based on the random forests technique, *ISPRS Journal of Photogrammetry and Remote Sensing*, **66**(1), 28-37.
- Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., y Gerard, P. D., 2003, Characterizing vertical forest structure using small-footprint airborne LiDAR, *Remote Sensing of Environment*, **87**(2-3), 171-182.