

**CENTRO DE INVESTIGACIÓN EN GEOGRAFÍA Y
GEOMÁTICA ING. JORGE L. TAMAYO A.C
CentroGeo**

Centro Público de Investigación CONACYT

CARACTERIZACIÓN DE LA ALTURA DE LOS BOSQUES DE
CONÍFERAS EN EL SUELO DE CONSERVACIÓN DE LA CDMX
MEDIANTE LA REFLECTANCIA DEL DOSEL

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN GEOMÁTICA

PRESENTA:

ERICK CORONADO JUÉZ

SUPERVISOR PRINCIPAL:

DR. JOSÉ LUIS SILVÁN CÁRDENAS

Comité Supervisor:

Dra. Alejandra López Caloca

Mtro. Mauricio Galena Pizaña

CDMX, DICIEMBRE, 2017.

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

Introducción

Actualmente uno de los aspectos que se está estudiando y analizando con grandes esfuerzos es el de los efectos del cambio climático en el medio ambiente, por lo que investigadores en distintas áreas se han sumado a estudiar este fenómeno para combatir el efecto invernadero, se ha observado que uno de los gases que contribuyen en mayor medida a esa situación es el CO_2 después del vapor de agua, ya que está relacionado directamente con la actividad humana. En las últimas décadas se ha observado un aumento en la concentración de este gas en la atmósfera, así como un incremento en la temperatura promedio en el mundo.

1.1. Planteamiento del Problema

Existen diversas evidencias directas sobre el calentamiento global, tales como la disminución en el tiempo de congelamiento de algunos lagos o ríos, así como el derretimiento en los polos y el aumento del nivel del mar. Se estima que en el cuarto de siglo comprendido entre 1750 y 1900 la concentración de CO_2 en el ambiente pasó de 280 ppm a 353 ppm respectivamente, se prevé que vaya en aumento a una tasa de 1.8 ppm por año (Bazzaz y Morse, 1990), adicionalmente se ha llegado a cuantificar que aumentará entre 550 y 700 ppm para el 2050 (Watson R.T., 1991).

Diversos estudios basados en modelos de circulación indican que podría existir una correlación entre el cambio climático y los gases de efecto invernadero en la atmósfera (Jones, 1997; Labraga, 1997), a la vez existen algunos autores que mencionan que el aumento del CO_2 derivado de la actividad humana es el principal efecto del cambio climático (Mann y Hughes., 1999; Miller, 1992; Vitousek-P.M., 1994), otros investigadores afirman que existe una relación entre la fluctuación natural del clima y las actividades del hombre, (Barnett, 1999).

Se han logrado esfuerzos por parte de distintos investigadores provenientes de

distintas áreas del conocimiento lo que ha originado que exista un cúmulo de datos los cuales necesitan ser analizados para obtener información que sea de utilidad para combatir los embates de la contaminación ambiental, la Percepción Remota contribuye con técnicas las cuales utilizan tecnología LiDAR como también imágenes satelitales, además de otras técnicas las cuales implican el uso de espectrómetros y radiómetros.

La captación de carbono se logra parcialmente evitando que se quemé madera y sus derivados, la madera en su estado natural es decir en los árboles sin cortar comprende lo que se conoce como biomasa, es por eso que se deben seguir impulsando políticas públicas que promuevan la conservación de los bosques, una de estas es la de los servicios ecosistémicos, la cual impulsa a particulares al aprovechamiento de zonas privadas o federales con fines de ecoturismo, conservación de especies forestales maderables o no maderables ya sean silvestres o no (flora y fauna).

Se sabe que los árboles de mayor altura y diámetro son los que contienen una cantidad mayor de biomasa, esto debido principalmente por su misma materia contenida en el follaje y tronco, por lo que es de suma relevancia conservar los bosques con especies de mayor edad y altura, ya que si estos son talados, se estará corriendo el riesgo que esa biomasa almacenada se reintegre al ambiente en forma de CO_2 . Otro factor de emisiones de CO_2 , son los incendios en zonas boscosas, estos han ido en aumento principalmente por las sequías en zonas tropicales esto en parte se le atribuye al cambio climático, pero también por la acción del hombre, ya que se quemán zonas bosques enteras para aprovecharlas como zonas de cultivo, propiciando un cambio de uso de suelo.

1.2. Área de estudio

La Ciudad de México se encuentra dividida en dos clasificaciones para el tipo de suelo, la primera es suelo urbano (SU) con una extensión de 60,867.9 ha, y el suelo de conservación (SC) con 87,294.36 ha (un 59% del total). La región que se analiza en este estudio se encuentra dentro de las delegaciones Magdalena Contreras, Alvaro Obregón, Cuajimalpa y Tlalpan (1.1). Dentro del SC se presentan diversas problemáticas sociales y políticas dando lugar a litigios entre las comunidades de la Ciudad de México y núcleos agrarios de otras entidades, en consecuencia esto hace difícil la conservación y restauración de los ecosistemas. El SC es de vital importancia ya que este provee un bienestar a los habitantes de la Zona Metropolitana del Valle de México

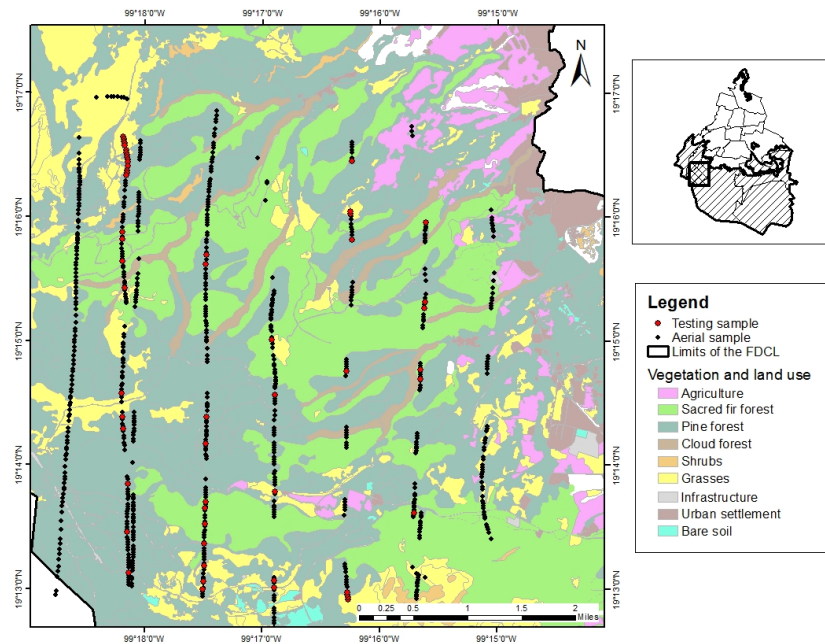


Figura 1.1 Área de estudio, correspondiente al Suelo de Conservación de la Ciudad de México

(ZMVM), por los servicios ecosistémicos que presta. El SC también representa una de las zonas donde se presenta la mayor captación de agua, promoviendo así la recarga de los mantos acuíferos, además de que la vegetación regula los escurrimientos superficiales protegiendo de la erosión hídrica y eólica el suelo. Sirve también como protección de flora y fauna de la región, además de proveer a la población un lugar de esparcimiento. Adicionalmente los bosques prestan el servicio de provisión de leña a las comunidades que habitan la zona, son estos los motivos por los que el suelo de conservación es de suma importancia para la vida en toda la ZMVM y es deber de todos conservarlo y cuidarlo para asegurar la supervivencia de la población (SMA, 2007).

1.3. Objetivo General

Esta tesis hace uso de datos del proyecto CB-2011-01-169069 del FONDO SECTORIAL DE INVESTIGACIÓN PARA LA EDUCACIÓN (SEP-CONACYT)), (Cárdenas, 2016), derivado de esta información se puede plantear un *objetivo particular* de esta tesis el cual es: desarrollar una metodología para la estimación de biomasa, la cual

no dependa de la adquisición o renta de tecnología costosa, de igual manera que se pueda realizar remotamente, sin todo el desgaste físico y económico que implica ir a la zona de estudio y el monitoreo constante que requiere la toma de mediciones in situ.

Dicho esto, el *objetivo general* es relacionar los datos LiDAR los cuales contienen información de la ubicación y altura del dosel de los árboles y las firmas espectrales de distintas regiones de la zona de estudio las cuales se denominaran huellas de aquí en adelante, con el objetivo de localizar ciertas bandas estrechas dentro del espectro electromagnético que respondan a ciertas alturas dentro de la huella. Pero ¿por qué es necesaria la altura de los árboles?, respondiendo esta pregunta principalmente por que si se tiene la altura se puede hacer una estimación de la biomasa con un cierto grado de error, al final se tratará de comparar nuestra estimación de altura contra datos de la CONAFOR a manera de verificación de nuestro modelo.

Hipótesis: La luz que incide sobre la estructura de los árboles genera una respuesta denominada reflectancia, ésta es el cociente entre la luz reflejada y la luz incidente, de tal manera que la reflectancia se puede utilizar para determinar la altura del bosque del suelo de conservación de la Ciudad de México.

1.4. Acerca de:

En el capítulo dos, se dan definiciones formales de los gases de efecto invernadero y cuales son los que deben de preocuparnos ya que se relacionan con el cambio climático, igualmente se define la biomasa. Posteriormente con base a una investigación en distintas fuentes se exponen diferentes formas del cálculo de la biomasa con respecto a volúmenes de existencia, tablas de frecuencias de ocurrencias, ecuaciones alométricas. Se sigue con el cálculo de biomasa con tecnología LiDAR la cuál implica recolección de datos con LiDAR aerotransportado, convirtiéndolo en una estimación no invasiva pero con costos altos.

El capítulo tres trata sobre aspectos técnicos de la estimación que se hizo, aprovechando los datos recolectados por LiDAR y comparándolos con información de imágenes hiperespectrales, con el objetivo de determinar ciertas bandas angostas las cuales dan información sobre la altura de los árboles, con lo que se propone una nueva metodología basada en datos hiperespectrales los cuales nos den una estimación de altura de alguna región de estudio.

En el capítulo cuatro se exponen los resultados obtenidos así como diversas gráfi-

cas y para finalizar en el capítulo cinco se plantea una discusión sobre los resultados y los objetivos alcanzados si es que lograron o no. El capítulo seis trata las conclusiones.

Índice general

Introducción	I
1.1. Planteamiento del Problema	I
1.2. Área de estudio	II
1.3. Objetivo General	III
1.4. Acerca de:	IV
2. Antecedentes: Estado del Arte	1
2.1. Gases de efecto invernadero	1
2.2. Servicios Ecosistémicos	2
2.3. Biomasa	3
2.4. Estimación de Biomasa por volumen de existencias	5
2.4.1. Ecuación General	5
2.5. Estimación de Biomasa por tablas de frecuencias de ocurrencia	6
2.6. Estimación de Biomasa a partir de Ecuaciones Alométricas	8
2.7. Estimación de Biomasa con RADAR	10
2.7.1. Coeficientes de retrodispersión para SAR	11
2.7.2. Datos Interferométricos para SAR	12
2.7.3. Datos Polarimétricos para SAR	12
2.8. Estimación de Biomasa a partir de Datos LiDAR	14
2.9. Estimación de Biomasa con datos HS	15
	VI

3. Fundamentos Teóricos de la Metodología	18
3.1. Aspectos Técnicos	18
3.2. Enfoque de la Percepción Remota	18
3.3. LiDAR aerotransportado	19
3.3.1. Estimación de la distancia por Tiempo de Vuelo	20
3.3.2. Estimación de la distancia por Diferencia de Fase	21
3.4. Percepción Remota Hiperespectral en la Vegetación	22
3.5. Características de los Sensores Hiperespectrales (HSS)	25
3.6. ¿Qué es un sensor Hiperespectral (HSS)	28
3.7. Principios Físicos y diseño de los sensores HS	30
3.8. Modos de Operación de los Sensores Hiperespectrales (HSS)	33
3.8.1. Imágenes HS con base en la tierra	33
3.8.2. Sensores HS Aerotransportados	34
3.8.3. Sensores HS montados en Satélites Artificiales	36
3.9. Mínimos Cuadrados Parciales (PLS)	39
3.10. Interpretación del modelo de regresión PLS	43
3.10.1. Interpretación Geométrica del modelo PLS	43
3.10.2. Selección del Número de Componentes	44
3.11. Importancia predictiva de la variable (VIP)	45
3.11.1. Bootstrap y selección de variables	46
4. Procesamiento de Datos y Resultados	48
4.0.1. Fuente y Tratamiento de los Datos Empleados	48
4.1. Construcción del Modelo	52
4.1.1. Selección de Componentes	52
4.1.2. Selección de las Variables	55
4.2. Validación del Modelo	59

5. Discusión y Conclusiones	62
Bibliografía	66
Glosario	72

Índice de figuras

1.1. Área de estudio, correspondiente al Suelo de Conservación de la Ciudad de México	III
3.1. Método de medición por tiempo de vuelo.	20
3.2. Medición por diferencia de fase.	21
3.3. Relación entre técnicas radiométricas, espectrométricas y de imagen. (Ortenberg et al., 2011, p. 43)	29
3.4. Concepto de imagen espectralradiométrica - cubo de datos compuesto por imágenes individuales grabadas sobre m bandas espectrales. Figura adaptada de (Ortenberg et al., 2011, p. 43)	30
3.5. Principio de Sensor de empuje, (Ortenberg et al., 2011).	32
3.6. Principio de Sensor de barrido,(Ortenberg et al., 2011).	32
3.7. Representación geométrica de la regresión PLS	44
4.1. Ejemplo de las huellas.	49
4.2. Huellas con puntos LiDAR en 2D.	49
4.3. Huellas con puntos LiDAR en 3D.	50
4.4. Diagrama del procesamiento de los datos.	51
4.5. Imagen LiDAR normalizada de la zona de estudio	51
4.6. Imagen LiDAR normalizada de la zona de estudio	51

4.7. Diagrama del Proceso para la construcción del Modelo. Cabe destacar que al momento de obtener la Y -estimada se repitió la regresión PLS omitiendo las bandas donde se presenta adsorción de agua para mejorar la correlación.	53
4.8. Selección de Componentes, para la variable Y	54
4.9. Estimación de la altura, para la variable Y	54
4.10. Varianza de las componentes del 1 al 30	55
4.11. Varianza acumulada de las componentes del 1 al 30	55
4.12. Índices del proceso de VIP	56
4.13. Contribución relativa	56
4.14. Diagrama de dispersión de altura LiDAR vs altura PLS.	57
4.15. Resultados del VIP y Coeficientes β de la regresión.	57
4.16. Gráfica de la Y -estimada.	58
4.17. Gráfica del RMSE de la segunda corrida de la regresión	58
4.18. Huellas y conglomerado con los sitios de la PAOT, para el caso de la huella i , se seleccionó la distancia d_1 y para la huella j se eligió la distancia d_5 , de cada sitio seleccionado se eligió la altura máxima de los árboles la cual se correlacionó con la altura estimada por la regresión en inglés Partial Least Squares, Mínimos Cuadrados Parciales (PLS).	60
4.19. Correlaciones de la altura máxima en los sitios de la PAOT y la altura máxima estimada por la regresión PLS, se tiene la mayor correlación para la distancia de 10.06 m. con un valor de $R^2 = 0.383$ tomando en cuenta las bandas de absorción de agua y $R^2 = 0.4766$ una vez que se descartaron las bandas de absorción de agua.	61

Capítulo 2

Antecedentes: Estado del Arte

2.1. Gases de efecto invernadero

El principal gas de efecto invernadero es el CO_2 , pero no es el único, también se considera al metano (CH_4), Oxido Nitroso (N_2O), como los principales, sin embargo se hace tanto énfasis en el CO_2 , ya que los efectos que este ha ocasionado al ambiente debido a la quema de combustibles fósiles entre otras actividades humanas han causado un gran impacto en la atmósfera, por si solo el CO_2 , contribuye aproximadamente con el 65% del forzamiento radiativo (es decir cualquier variación en la radiación entrante o saliente en un sistema climático, debido a la radiación solar incidente), es decir su contribución con el calentamiento o cambio climático es considerable, también se estima que contribuyo con el 81% aproximadamente del forzamiento radiativo en la década pasada (Organization, 2016). Antes de la industrialización se tenían márgenes aproximadamente de 278 ppm, sin embargo una vez que surgieron las industrias globales el CO_2 se ha incrementado un 144% en la atmósfera en 2015, esto a causa de la quema de combustibles fósiles, la producción de cemento, la deforestación y el cambio de uso de suelo también han contribuido a este aumento. Solamente del 2005 al 2014 se presentó un aumento de aproximadamente el 44%, debido a la actividad humana, el restante 56% aproximado fue eliminado por los océanos y la biosfera terrestre. Es preocupante que el promedio anual de CO_2 fue de 400.0 ± 0.1 ppm, este aumento se ha asociado también con la presencia de fenómenos como “El niño” (Organization, 2016).

2.2. Servicios Ecosistémicos

La captación de carbono es una de las problemáticas que más se menciona cuando se habla de cambio climático y el efecto invernadero, recientemente se han hecho diversas propuestas para poder lograr el objetivo que es la captación sostenible de tal manera que se vean beneficiados los diferentes actores de la sociedad que están involucrados, uno de los esfuerzos más grandes es el de los Servicios Ecosistémicos (SE), estos son el vínculo que media entre el sistema social y el ecológico (Rincón-Ruíz, 2014), es evidente cada día que el sistema social no es independiente del sistema ecológico, ya que claramente su bienestar está relacionado en gran parte de la biodiversidad y los servicios que los ecosistemas proveen. En este sentido existen iniciativas globales como el Plan Estratégico para la Diversidad Biológica 2011-2020 y las Metas de Aichi, el plan pretende implementar acciones a escala global, nacional y regional que lleven a disminuir y detener la pérdida de biodiversidad con el propósito de conservar la diversidad de la vida y el bienestar humano, las Metas de Aichi son las siguientes:

- Meta 1: Para 2020, a más tardar, las personas tendrán conciencia del valor de la diversidad biológica y de los pasos que pueden seguir para su conservación y utilización sostenible. (Ruiz, 2014)
- Meta 2: Para 2020, a más tardar, los valores de la diversidad biológica habrán sido integrados en las estrategias y los procesos de planificación de desarrollo y reducción de la pobreza nacionales y locales y se estarán integrando en los sistemas nacionales de contabilidad, según proceda, y de presentación de informes. (Ruiz, 2014)

Otra iniciativa importante es la Evaluación de los Ecosistemas del Milenio (EEM 2003), esta expuso las consecuencias para la humanidad en caso de continuar deteriorando los ecosistemas, así mismo sentó las bases conceptuales y metodológicas para el estudio de los SE y la iniciativa “La Economía de los Ecosistemas y la Biodiversidad”, conocida como el TEEB por sus siglas en inglés, la cual recalcó la necesidad de involucrar tanto los valores ecológicos, sociales y monetarios y manejarlos como un conjunto.

Una definición concreta para lo que se trata anteriormente sería, “La Gestión Integral de la Biodiversidad y sus Servicios Ecosistémicos se define como el proceso por el cual se planifican, ejecutan y monitorean las acciones para la conservación de

la biodiversidad y sus servicios ecosistémicos, en un escenario social y territorial definido y en diferentes estados de conservación, con el fin de maximizar el bienestar humano, a través del mantenimiento de la resiliencia de los sistemas socioecológicos a escalas nacional, regional, local y transfronteriza” (Rincón-Ruíz, 2014).

La definición anterior hace referencia a los sistemas socioecológicos, esto implica no solo tomar en cuenta los parámetros biofísicos, si no que ir más allá e involucrar temas socioculturales y económicos que generen una visión integral del territorio, esto generará un ambiente en el cual se puedan llegar a acuerdos entre los diferentes intereses de la sociedad promoviendo su participación, generando políticas públicas que mejoren las condiciones del bienestar humano, de esta manera los involucrados pueden reconocer el valor de la biodiversidad y sus servicios ecosistémicos como un valor público, siguiendo los principios definidos por el enfoque ecosistémico propuesto por el Convenio de Diversidad Biológica (CDB) (Rincón-Ruíz, 2014).

2.3. Biomasa

La biomasa se define como “la cantidad total de materia orgánica por encima del suelo en los árboles, expresada como toneladas por unidad de área secadas al horno”(Brown, 1997). La biomasa en los bosques provee una estimación de los depósitos de carbón en la vegetación del bosque, se calcula que es de cerca del 50 % del carbón. Por lo que la biomasa representa potencialmente el 50 % del carbón que podría ser emitido a la atmósfera en forma de CO_2 si el bosque fuera talado o quemado. Existen diversos trabajos en los que se ha tratado de estimar la biomasa en los bosques tropicales y sus efectos en el aumento del CO_2 en la atmósfera (Brown y Lugo, 1989), a continuación se describirán algunas de las metodologías que se han empleado en el cálculo de biomasa.

En las dos secciones siguientes se describen metodologías para el cálculo de biomasa. La primera propuesta se basa en datos de inventarios existentes, esta aproximación utiliza mediciones existentes de estimaciones de volumen (VCC, volumen con corteza (o existencias) (m^3/ha)) convertidas a densidad de biomasa (t/ha) (Brown, 1997). Cabe destacar que una de las desventajas de éste método y el de tablas de la frecuencia de ocurrencia de árboles requieren que se hayan hecho las mediciones referentes a los árboles a todas las especies en la zona de estudio, no se podrán aplicar si el inventario no incluye mediciones de todas las especies, ya que no hay manera de

que los resultados se extrapolen a las especies que se excluyen del inventario (Brown, 1997). Al utilizar datos de inventarios forestales generalmente se pueden superar muchas de las dificultades que se presentan en estudios ecológicos, en general existe una gran cantidad de datos y se recopilan de espacios territoriales amplios, ya sea a nivel nacional y áreas locales del país (Brown, 1997). Sin embargo los inventarios forestales no se libran de tener ciertos problemas, como los siguientes:

- Estos tienden a hacerse en bosques los cuales hasta cierto punto tienen una motivación económica o comercial, es decir bosques densos y no siempre toman en cuenta bosques que no son tan densos, así como también aquellos que no son de importancia debido a que la madera no se utiliza en el ámbito industrial (Brown, 1997).
- Generalmente se excluyen los árboles que tienen un diámetro menor a 10cm y por arriba de los 50 cm, al no incluir estos árboles pequeños se puede omitir una aportación del 30 % de la biomasa (Brown, 1997).
- El diámetro máximo en las tablas es ilimitado e incluye árboles de más de 80 cm de diámetro y estos están agrupados en una sola clase, afectando así la distribución por diámetro de estos árboles tan grandes, lo que altera significativamente la densidad de biomasa por encima del suelo (Brown, 1997).
- No se incluyen a todas las especies de árboles, se incluyen generalmente los que tienen mayor valor comercial (Brown, 1997).
- Se excluyen datos que son relevantes en la mayoría de los casos, las mediciones de campo no se almacenan correctamente y por lo tanto se pierden (Brown, 1997).
- No se tiene una definición consistente de inventario volumétrico (Brown, 1997).
- No se describe precisamente la situación actual del bosque, generalmente se describe como primario, pero de acuerdo a la distribución por diámetro y volumen se puede percibir de otra manera. (Brown y Lugo., 1991; Brown y Lugo, 1994)
- Muchos de los inventarios son viejos, lo que puede implicar que el bosque haya desaparecido o cambiado (Brown, 1997).

2.4. Estimación de Biomasa por volumen de existencias

La estimación de biomasa por medio de esta técnica funciona de mejor manera para bosques secundarios y densos maduros, que se encuentren en climas húmedos y secos (Brown, 1997). Se utiliza en bosques densos únicamente debido a que la base de datos esta basada en este tipo de bosques inicialmente. El dato que se utilizará para el cálculo sera el VCC/ha que es el volumen del árbol desde la base hasta la corona o a la rama principal, el inventario debe contener todos los árboles, comerciales o no, diámetro a la altura del pecho (DAP) de 10 cm o por encima del contrafuerte (Brown, 1997).

2.4.1. Ecuación General

Inicialmente se calcula la biomasa a partir del VCC, calculando la biomasa del volumen que ha sido inventariado, después se puede expandir este valor para hacer el cálculo de la demás vegetación que esta por encima del suelo (Brown y Lugo., 1992).

Densidad de biomasa por encima del suelo (ton/ha)

$$= VCC * DM * FEB \quad (2.1)$$

Donde

VCC = volumen con corteza (o existencias) (m^3/ha).

DM = densidad media de madera ponderada por el volumen (toneladas de biomasa secada al horno por volumen de existencias verdes en m^3).

FEB = factor de expansión de la biomasa (proporción entre la biomasa de árboles secada al horno existente por encima del nivel del suelo y la biomasa secada al horno del volumen inventariado).

La densidad media de madera (DM), es la relación entre la masa de madera seca al horno por unidad de volumen verde, expresado en $tons/m^3$ o grs/cm^3 ; después de hacer una regresión, la cual involucra la densidad media por especie y volumen por especie, se obtiene una densidad media ponderada por el volumen para especies dependiendo de la localización geográfica considerando también que la masa de madera

contiene un aproximado del 12% de humedad, (Reyes, 1992).

Tabla 2.1 Media aritmética y valores más comunes de densidad de madera (*tons/m³*) para las especies arbóreas tropicales por regiones, según (Brown, 1997).

Densidad de Madera		
Región Tropical	Media	Valores habituales
África	0.56	0.50-0.79
América	0.60	0.50-0.69
Asia	0.57	0.40-0.69

2.5. Estimación de Biomasa por tablas de frecuencias de ocurrencia

En esta sección se describe brevemente otra técnica que se puede implementar para la estimación de biomasa, la cual emplea ecuaciones de regresión y está basada en tablas de frecuencias de ocurrencia de árboles, estas contienen información del diámetro a la altura del pecho (dap); básicamente lo que se hace es estimar un promedio de biomasa por árbol de cada diámetro de clase, (cabe mencionar que los diámetros están agrupados en clases), posteriormente se multiplica el promedio de biomasa por el número de árboles que hay en cada clase, al final se sumará a través de todas las clases. Al momento de hacer el cálculo se debe seleccionar el diámetro promedio de cada clase del dap, de igual manera se hace para las clases que son de $dap \leq 10$ cm (Brown y Lugo, 1989). Una mejor opción para las clases con diámetros más grandes es utilizar el diámetro medio cuadrático (DMQ) de cada clase de dap, de igual manera se puede utilizar el área basal para cada clase de dap siempre y cuando este dato sea conocido, con esto se obtiene el dap de un árbol de área basal para cada clase de diámetro. El cálculo de DMQ se realiza de la siguiente manera (Brown, 1997):

$$DMQ = \frac{AB_{clase}}{N_{clase}}$$

Donde AB_{clase} = Área basal por clase y N_{clase} es el número de árboles por clase. La ecuación anterior determinará el área basal del árbol promedio. Por lo tanto para las clases con diámetros grandes, se podrá utilizar la siguiente relación para calcular el

dap:

$$dap = 2 * \sqrt{\frac{AB_{clase}}{3.142}}$$

Una vez que se tienen estos factores, se podrán implementar en las ecuaciones de regresión para cálculo de biomasa esto para bosques de hoja ancha, estas ecuaciones se basan en bases de datos que contienen datos de tres regiones tropicales y se pueden utilizar para bosques de clima seco, húmedo y lluvioso. En resumen se tiene la siguiente tabla la cual muestra la ecuación para cada una de las tres zonas climáticas con sus respectivas limitaciones:

Tabla 2.2 Ecuaciones de regresión para estimación de biomasa para árboles tropicales: Y=biomasa por árbol en kg, D =dap y AB = área basal en cm^2

Zona Climática	Ecuación	Rango de dap (cm)	Número de arboles	r^2
Seco	a) $Y = \exp(-1.996+2.32*\ln(D))$	5 - 40	28	0.89
	b) $Y = 10^{(-.0535+\log_{10}(AB))}$	3 - 30	191	0.94
Humedo	c) $Y = 42.69 - 12.800(D) + 1.242(D^2)$	5 - 148	170	0.84
	d) $Y = \exp(-2.134 + 2.530 * \ln(D))$			0.97
Lluvioso	e) $Y = 21.297 - 6.953(D) + 0.740(D^2)$	4 - 112	169	0.92

Algunas observaciones adicionales que se deben tener en cuenta es que estas ecuaciones de regresión solo se deberán aplicar a árboles que cumplan con los límites de diámetro, para especímenes con diámetros más grandes no se aconseja utilizarlas. La ecuación a) fue revisada por (Brown y Lugo, 1989) para bosques secos en la India y la ecuación b) por (Martinez Yrizar y Cervantes., 1992) para bosques secos en México (la ecuación original se basa en AB), para zonas con lluvias menores que 900 mm/año se deberá utilizar la ecuación a). Las ecuaciones c) y d) se basaron en la misma base de datos, A. J. R. Gillespie, pers. comm. basada en una revisión de una ecuación de (Brown y Lugo, 1989). La ecuación c) se basa en (Brown y Iverson, 1992). En caso de que se quiera conocer más acerca de estas ecuaciones se puede consultar (Brown, 1997).

2.6. Estimación de Biomasa a partir de Ecuaciones Alométricas

La alometría es un término acuñado por Julian Huxley y Georges Teissier en 1936, y hace referencia al crecimiento relativo que tiene una especie, ellos lo utilizaron para describir el crecimiento por ejemplo de las personas, esto se puede observar en el crecimiento que tienen los brazos y piernas de un niño, las cuales no crecen con la misma proporción, Julian Huxley y Georges Teissier (Gayon, 2000) observaron que la tasa de crecimiento de estas partes del cuerpo es mayor que las de la cabeza y torso, por lo que las proporciones de un niño y un adulto son totalmente diferentes. Otro caso particular se da en el cangrejo violinista (*Uca pugnax*), el cual desarrolla una pinza mucho más que la otra, llegando a constituir el 38 % de la masa corporal del organismo, lo que representa un crecimiento alométrico, y se puede establecer una tasa de crecimiento de la pinza de 6 veces más que el resto de su cuerpo.

Una definición formal de alometría dada por (S.J., 1966), es la siguiente: “Se define como el estudio de los cambios de proporción correlacionados con la variación de tamaño del total del organismo o parte que se tenga en consideración. Las variaciones pueden ser morfológicas, psicológicas o químicas”. Aunque la definición no se refiere explícitamente a una expresión matemática, las relaciones alométricas en general son de la forma:

$$Y = ax^b \rightarrow \log(Y) = \log(a) + b\log(x)$$

Donde Y es la magnitud del crecimiento diferencial de un órgano; x tamaño del cuerpo; k constante diferencial de la relación de crecimiento; b constante (índice de origen)

Las ecuaciones alométricas formalizan de forma cuantitativa una relación entre una medida de un árbol (por ejemplo la biomasa) y otra medida (su diámetro). Si se observa el diámetro D y biomasa B existe un coeficiente a que corresponde a:

$$\frac{dB}{B} = a \frac{dD}{D} \quad (2.2)$$

Al integrar la ecuación (2.2) se obtiene la relación de potencia $B = b \times D^a$, el parámetro a se conoce como el coeficiente de alometría (proporcionalidad entre los aumentos relativos), el parámetro b por su parte es la proporcionalidad entre las mag-

nitudes acumuladas. En ocasiones se añade un parámetro de intersección c , de tal manera que se tiene $B = c + b \times D^a$, esta precisión se realiza para representar la biomasa del individuo antes de que alcance la altura a la que se mide el diámetro (por ejemplo 1.3 m si D se midió a 1.3 m), (Picard et al., 2012).

Basandonos en lo anterior, las plantas en general, presentan un crecimiento alométrico (Niklas, 1994), es decir cada parte de la planta crece hasta alcanzar su etapa madura y funcional, lo que les permite crecer indeterminadamente y cambiar su estructura de acuerdo a las condiciones ambientales, esto lo logran añadiendo nuevos órganos y tejidos constantemente. Debido a lo anterior no todos los componentes de una planta contienen la misma cantidad de materia ni crecen en la misma proporción, por lo que haciendo una comparación, las partes con mayor materia o peso tendrán mayor capacidad de retener el carbono que por ejemplo las hojas con menor masa.

La estimación de biomasa por parte de científicos se remonta al uso de ecuaciones que relacionan el diámetro normal (a la altura del pecho), y a técnicas las cuales implican talar algunos arboles hasta varias decenas, para después separarlos por ramas, follaje y hacer cortes transversales del tronco denominadas rodajas de distintos espesores, posteriormente éstas se pesan en el mismo lugar donde se talaron para evitar la pérdida de humedad, esto se realiza con los árboles los cuales hayan sido cortados. Posteriormente son secados en microtuneles los cuales se revisten de plástico y en su interior tienen una rejilla la cual permite que circule el aire, esto a manera de estufa solar, esto hasta alcanzar un peso constante, el cual se denomina peso seco. Los árboles con diámetros normales por debajo de los 10 cm, se cortan y son secados de igual manera y pesados en seco. Debido a que no se puede eliminar toda la humedad al secarlos a temperatura ambiente algunos especímenes son secados en hornos, cabe destacar que este proceso puede durar varias semanas, y después pesados para determinar el porcentaje de humedad que pierden, al final se les resta este porcentaje a las muestras que se secaron en el microtúnel. Una vez que se han hecho estas mediciones del peso seco, se crea una relación entre peso seco y fresco de las rodajas, las ramas y follaje, lo cual genera unos coeficientes para cada parte del árbol, los cuales al multiplicarse por el respectivo peso en seco genera el cálculo de biomasa del fuste (esto es trozas y tocón) y de igual manera de la copa del árbol (ramas y follaje), lo que genera la biomasa aérea total de cada árbol (Dulce Ma. Avendaño Hernandez y Barra, 2009).

Como se ha explicado anteriormente la estimación de biomasa en material orgánico a partir de ecuaciones alométricas, resulta un proceso hasta cierto punto laborioso

y costoso en tiempo y esfuerzo, adicionalmente conlleva la destrucción de una parte del bosque, sin embargo este proceso solo será necesario realizarlo una ocasión, ya que las ecuaciones que se generen se podrán utilizar posteriormente, sin necesidad de realizar este proceso nuevamente; en este trabajo de tesis más adelante se propone una manera alternativa mediante técnicas de percepción remota para la estimación de biomasa, la cual resultará en una disminución de tiempo y recursos, además de que no implica ningún daño en la estructura física del bosque.

2.7. Estimación de Biomasa con RADAR

Dentro de los sensores activos es decir que este emite y recibe una porción de energía que interactúa con la superficie de algún objeto y objetivo y esta es retrodispersada, la cual permite medir la intensidad y tiempo de retraso de la señal reflejada, se encuentra el en inglés *R*Adio *d*etection and *r*anging (RADAR), este tipo de sensores trabajan en el rango de las microondas dentro del espectro electromagnético, una de sus ventajas es que cuentan con la capacidad operacional de recopilar información tanto de día como de noche, otra ventaja es que no son susceptibles a ruido por parte de la atmósfera (Pizaña y Romero, 2016).

Dicha energía cuando presenta un haz de iluminación con la misma longitud de onda y fase se le conoce como coherente, para generar imágenes se recomiendan diferentes esquemas de polarización (orientación de los campos electromagnéticos, estos campos son perpendiculares a la dirección de propagación de la señal del RADAR, los radares utilizan polarizaciones horizontales, verticales, lineales o circulares, en función de la aplicación). La resolución espacial de las imágenes de radar dependerá de la longitud de la antena (apertura) del receptor y del ángulo de inclinación del sensor, los radares de apertura sintética (SAR por sus siglas en inglés *S*ynthetic *A*perture *R*adar), aprovechan este hecho al ser montados en aeronaves o satélites, aprovechando el movimiento de una ruta establecida que a su vez simula una antena virtual de longitud muy amplia, los cuales reciben y transmiten señales de radar por medio de antenas, esto comprende señales de radar largas coherentes las cuales son muy efectivas, permitiendo la adquisición de imágenes de alta resolución (Pizaña y Romero, 2016).

Se han planteado diversas aplicaciones para el SAR algunas son: determinar el estado de los bosques, clasificación de bosques, densidad de biomasa, altura del do-

sel del bosque, degradación de los bosques por incendios, deforestación y para medir la humedad del suelo en el bosque, todo esto gracias a que la energía de las microondas penetran el dosel del bosque permitiendo que la energía retrodispersada la cual es influenciada por los parámetros estructurales de los árboles pueda ser cuantificada por el SAR. Se ha encontrado que las bandas *X* y *C* (con longitud de onda 2.4 y 7.5 cm, respectivamente), están asociadas con la retrodispersión de las hojas, también se asocian las bandas *L* y *P* (con longitud de onda 15 y 100 cm, respectivamente), con retrodispersión de ramas y tallos (Pizaña y Romero, 2016). A continuación se describirán tres de los principales enfoques 1. Coeficientes de retrodispersión, 2. datos interferométricos para SAR y 3. datos polarimétricos para SAR los cuales implican el uso de SAR, para la adquisición de información que sea relevante para el cálculo de la biomasa.

2.7.1. Coeficientes de retrodispersión para SAR

Para la construcción de este coeficiente se realiza una proporción entre la densidad de energía dispersada y la densidad de energía transmitida desde la superficie de la superficie objetivo por unidad de área. Los coeficientes de retrodispersión (σ^0) es la cantidad de la sección transversal del radar, ésta magnitud se expresa como una expresión logarítmica con unidades de decibel mediante la siguiente ecuación (Pizaña y Romero, 2016):

$$\sigma_{(db)}^0 = 10 * \log_{10} \sigma^0 \quad (2.3)$$

Estos coeficientes se relacionan con dos variables del sensor y con los parámetros de la superficie objetivo. Las características del sensor están en función de la longitud de onda, polarización y del ángulo incidente, mientras que las características del objetivo están asociadas con la rugosidad, geometría y propiedades dieléctricas. Los modelos de biomasa que utilizan este enfoque usualmente se aplican utilizando un modelo de regresión simple, asumiendo la suposición de correlación entre los coeficientes de retrodispersión y biomasa/carbón sobre el suelo. Los resultados dependerán de la saturación de la señal, ya que esta esta en función de la longitud de onda, polarización y de las características de la cobertura vegetal, además de las propiedades del terreno como la pendiente y entorno (Pizaña y Romero, 2016).

2.7.2. Datos Interferométricos para SAR

Los SAR contienen diversos métodos los cuales utilizan información derivada de la diferencia de fase o estado de vibración de la onda al momento que es recibida por el radar entre dos SAR los cuales son conocidos como maestro y esclavo, estos se localizan en dos posiciones diferentes, esta información es almacenada y se conoce como fase interferométrica ($\Delta\Phi_{Int}$) (Pizaña y Romero, 2016), se puede escribir como una ecuación de la siguiente manera:

$$\Delta\Phi_{Int} = \Phi_S - \Phi_M = \Phi_{Topo} + \Phi_{Mov} + \Phi_{Atm} + \Phi_{Noise} \quad (2.4)$$

donde Φ_S y Φ_M son la fase de observación de la imagen esclava y la fase de observación de la imagen maestra, respectivamente, Φ_{Topo} es la componente topográfica, Φ_{Mov} es la componente de cambio, Φ_{Atm} es la componente atmosférica y Φ_{Noise} es la fase de ruido (Pizaña y Romero, 2016).

En modelos de biomasa la coherencia interferométrica se utiliza bajo la suposición de que para áreas forestales la coherencia disminuye con el incremento de la densidad de vegetación, como la dispersión volumétrica aumenta con el movimiento provocado por el viento y el crecimiento de la vegetación. Los modelos de biomasa se aplican bajo este enfoque utilizando regresión lineal simple, bajo la suposición de correlación entre la coherencia interferométrica y la biomasa/carbón sobre el suelo. Otro enfoque utiliza métodos de regresión tales como Kriging y algoritmos de clasificación. Los resultados dependerán temporalidad de los datos, tipo de bosque, polarización y longitud de onda del sensor (Pizaña y Romero, 2016).

2.7.3. Datos Polarimétricos para SAR

Los radares pueden transmitir y recibir radiación electromagnética de diferente polarización, esta puede ser horizontal lineal H o vertical lineal V (las más comunes), esto dependerá en gran medida de la aplicación. Cuando la energía transmitida tienen la misma dirección que la energía emitida a esta se le denomina polarizada; cuando la dirección del campo de la energía transmitida es ortogonal a la del sistema receptor se le conoce como polarización cruzada. Los esquemas de polarización son HH para transmisión horizontal y recepción horizontal y HV para transmisión horizontal y recepción vertical.

Los sistemas de radar se pueden configurar en diferentes niveles de polarización, como se presenta a continuación:

- Polarización simple: HH, VV, HV Y VH
- Polarización dual: HH y HV, VV y VH o HH y VV
- tetra polarizado: HH, VV, HV y VH

Una polarización en cuadratura o radar polarimétrico utiliza estas cuatro polarizaciones con el fin de cuantificar las magnitudes y las diferencias de fases relativas entre el esquema de polarización o canales a través de una forma elíptica. Estos sistemas de radar dan origen a lo que se denomina polarimetría de radares de apertura sintética, los cuales describen la superficie mediante la combinación de polarizaciones suponiendo que la interacción de la energía electromagnética y los elementos de la superficie puedan cambiar la polarización de una porción de la longitud de onda transmitida por el sensor, por lo tanto recibe información de la amplitud y la fase relativa de el mismo objetivo en cuatro canales de información, esto se considera la base para la descripción de la dispersión polarimétrica de la superficie objetivo. Esto se puede resumir en una matriz de dispersión, la cual describe las diferentes formas de polarización de los campos eléctricos polarizados entre la onda incidente y la onda dispersa con el fin de ser la base para diversas maneras de analizar las propiedades de dispersión del objetivo, se puede utilizar como matriz de covarianza y coherencia, y se pueden aplicar descomposiciones y transformaciones a esta a través de la sinergia entre la polarimetría y la interferometría (Pizaña y Romero, 2016), esta matriz es de la siguiente forma:

$$S = \begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix}.$$

Se puede hacer uso de la polarimetría en modelos de biomasa bajo el supuesto de la correlación entre las propiedades estructurales del bosque y comportamiento polarimétrico. Se construye con mecanismos de análisis de dispersión, realizando descomposición polarimétrica, tal como Freeman Durden, eigenvalores y Cloude and Pottier. Usualmente se utilizan modelos de regresión simple y múltiple y modelos de regresión aleatorios no paramétricos, los resultados de este enfoque están relacionados con tipo de bosque, resolución espacial y longitud de onda del sensor (Pizaña y Romero, 2016).

2.8. Estimación de Biomasa a partir de Datos LiDAR

Existe otra técnica de percepción remota la cual permite monitorear zonas boscosas, esta implementa el uso de LiDAR aerotrasnportado, el cual es utilizado para la estimación de variables forestales, se pudo avanzar en el cálculo de perfiles de áreas boscosas y alturas en las décadas de los ochenta y noventa, con lo que se pudieron relacionar con variables previamente medidas como la altura media por parcela, densidad de copa, volumen y biomasa para distintos cálculos (Hyypä J., 2008).

Las técnicas de percepción remota aportan a los inventarios forestales, información valiosa ya que por por definición este evalúa y cuantifica uno o varios elementos de los recursos forestales, de esta manera el inventario forestal estima la cantidad y calidad de los árboles presentes en un área. Otro aporte es que al no ser una técnica invasiva y a distancia, se eliminan requerimientos presenciales en la zona de estudio, evitando con esto tener un contacto directo el cual en ciertos casos es destructivo, además de ahorrar recursos financieros y humanos, así se evitan también visitas programadas en ocasiones que se requiere tener un seguimiento en las mediciones de altura o diámetro a la altura del pecho (dap), para utilizarlos en ecuaciones alométricas para estimación de biomasa. Se puede decir que el camino esta trazado para estimar variables forestales con técnicas de percepción remota, las cuales siguen en pleno desarrollo.

El LiDAR aerotrasnportado permite la generación de nubes de puntos que pueden georeferenciarse tanto para el dosel, como para el área de estudio, además abre nuevas alternativas para calcular de manera precisa alturas y generar modelos del terreno, esto en su conjunto ha ayudado a la estimación de variables para inventarios, divididos en dos casos: 1) estimación de variables por árbol, en el que se segmentan los puntos de la nube para separarlos por árbol o se pueden generar modelos de las copas por árbol, permitiendo determinar la altura y los parámetros de la copa tales como el diámetro y área, mismos que se utilizan en ecuaciones alométricas tal como se vio en la sección anterior, para el cálculo de la biomasa. En el segundo caso, 2) la estimación es por parcela o área forestal y se ha estudiado generar variables estadísticas de la nube, las cuales puedan ser utilizados como parámetros para generar regresiones, las cuales junto con áreas previamente censadas en campo puedan generar variables forestales que sean de interes en el área de estudio (Shan, 2009a).

2.9. Estimación de Biomasa con datos HS

Para la estimación de biomasa por medio de datos Hiperespectrales (HS) se utilizan muestras recolectadas en campo por medio de un espectroradiómetro de campo, así como también se utilizan datos de sensores como el Hyperion. Algunos estudios realizados anteriormente (Psomas y Zimmermann, 2011), realizan un muestreo en campo el cual implica recolección de pastizales, para estimar su biomasa verde en laboratorio y su biomasa en seco después de que se ha colocado en un horno por 72 horas a una temperatura de aproximadamente 65°C. La diferencia de pesos es la que da una estimación del contenido de biomasa en el pastizal, de esta manera al realizar varios muestreos se cuenta con una muestra de biomasa que se utilizará para calibrar modelos de regresión con el objetivo de determinar si existe una relación entre la biomasa y los datos hiperespectrales (Thenkabail et al., 2000). Cabe señalar que las reflectancias recolectadas con el espectroradiómetro de campo, se transformaran en datos las cuales simulen ser bandas del sensor Hyperion para tener una homogeneidad de los datos.

Se ha podido relacionar a la biomasa con la productividad de las plantas en diversos estudios (Mittelbach et al., 2001), por lo que parece natural que esta tenga una dependencia directa con índices de vegetación de diferencias normalizadas (por ejemplo NDVI (Gould, 2000; Oindo et al., 2003; Oindo y Skidmore, 2002)) y con la reflectancia hiperespectral pura (Carter et al., 2005).

En el caso de los datos HS obtenidos con Hyperion, se menciona que algunas bandas se fueron fijadas en cero en el rango de los 1350-1415nm y entre 1800-1950 nm, ya que en esta región se ve afectada por la absorción atmosférica de vapor de agua. Posteriormente se realizan algunas correcciones atmosféricas, se puede consultar a detalle este procedimiento en (Psomas y Zimmermann, 2011).

Hablando sobre los índices que se han considerado dentro de algunos estudios destacan los siguientes: relacionados con la estructura de la vegetación (NDVI, RDVI, SR, SAVI, TSAVI, OSAVI and $MTVI_1$), al estado del agua en la vegetación (NDWI, SRWI, PWI and WDVI), a la clorofila y la pared roja (VOG_1 , VOG_2 , VOG_3 y MCARI), como se menciono antes, estos están relacionados con la productividad de la vegetación y por lo tanto con la biomasa (para más detalle sobre estos índices se puede consultar la tabla (2.3)).

Estos índices se utilizan de manera individual y en conjunto principalmente para

Tabla 2.3 Tabla de índices relacionados con la estructura, el estado de agua y clorofila de la vegetación

Índice	Ecuación	Referencia
NDVI	$\frac{(NIR-Red)}{(NIR+Red)}$	(Broge y Leblanc, 2001)
RDVI	$\sqrt{NDVI * DVI}$	(Broge y Leblanc, 2001)
SAVI	$\frac{(NIR-Red)}{(NIR+Red+L)} * (1 + L)$	(Broge y Leblanc, 2001)
WDVI	$NIR - a * Red$	(Broge y Leblanc, 2001)
TSAVI	$\frac{a(NIR-a*Red-b)}{a*NIR+Red-a*b}$	(Broge y Leblanc, 2001)
OSAVI	$\frac{(1+0.16)*(NIR-Red)}{(NIR+Red+0.16)}$	(Daughtry et al., 2000)
MCARI	$(R_{700} - R_{670}) - 0.2(R_{700} - R_{550}) * \frac{R_{700}}{R_{670}}$	(Daughtry et al., 2000)
MTV ₁	$1.2 * [1.2 * (R_{800} - R_{550}) - 2.5 * (R_{670} - R_{550})]$	(Zarco-Tejada et al., 2005)
SR	$\frac{(NIR)}{(Red)}$	(Zarco-Tejada et al., 2005)
VOG ₁	$\frac{(R_{740})}{(R_{720})}$	(Zarco-Tejada et al., 2005)
VOG ₂	$\frac{(R_{734}-R_{747})}{(R_{715}+R_{726})}$	(Zarco-Tejada et al., 2005)
VOG ₃	$\frac{(R_{734}-R_{747})}{(R_{715}+R_{720})}$	(Zarco-Tejada et al., 2005)
SRWI	$\frac{(R_{858})}{(R_{1240})}$	(Zarco-Tejada et al., 2003)
PWI	$\frac{(R_{970})}{(R_{900})}$	(Zarco-Tejada et al., 2003)
NDWI	$\frac{(R_{860}-R_{1240})}{(R_{860}+R_{1240})}$	(Zarco-Tejada et al., 2003)

realizar la calibración del modelo de regresión lineal, tomando en cuenta solo aquellos que tengan una correlación baja, con el objetivo de evitar la colinealidad es decir $r \leq 0.5$.

Se sabe que estos índices no abarcan en su totalidad todo el espectro, si no que se sitúan principalmente en regiones donde se absorbe la clorofila (680 nm), la reflectancia del en inglés Near Infrared, Infrarrojo cercano (NIR) (800 nm) o en el pico de la reflectancia del verde (550 nm). Esto representa una dificultad ya que no están tomándose en cuenta un gran número de bandas de datos hiperespectrales, en respuesta a esto se construyen otro tipo de índices denominados tipo-NDVI de banda estrecha (be_NDVI_{tipo}) (Thenkabail et al., 2000), y son de la siguiente forma:

$$be_NDVI_{tipo}[b_1, b_2] = \frac{b_1 - b_2}{b_1 + b_2}. \tag{2.5}$$

Donde b_1 y b_2 , son las bandas simulados tipo Hyperion provenientes de las mediciones de campo de la reflectancia, con esto se calcularon un total de 27889 índices de banda estrecha. Estos posteriormente serán utilizados en los modelos de regresión con el objetivo de determinar su eficacia para explicar la biomasa medida en laboratorio (Psomas y Zimmermann, 2011). La desventaja principal es que estos índices no

consideran todas las bandas hiperespectrales, además de que mucha de la información es redundante entre bandas que son vecinas (Thenkabail et al., 2004), es por esto que puede ser que la información que utilizan los índices no sea utilizada de manera óptima.

Los índices se calcularon para los datos adquiridos con el sensor Hyperion y con el espectroradiómetro de campo, posteriormente se calculó la diferencia entre ellos (medido como el en inglés root mean square error, error cuadrático medio (RMSE)), seleccionando solo aquellos con las diferencias más pequeñas.

En general no se encontró que los índices explicaran satisfactoriamente la biomasa, incluso los modelos que se basaron en los índices de vegetación existentes. Se encontró una relación alta con los índices que se relacionan con el contenido de agua en el dosel, esto podría explicar la fuerte relación entre el contenido de agua en el dosel y biomasa descritos por (Anderson et al., 2004; Mutanga et al., 2003). Esto corresponde a lo que propone (Asner, 1998), al mostrar que con un incremento en el contenido de agua en el dosel se obtendrá un incremento de biomasa.

Uno de los índices que describieron mejor esta relación fue el be_NDVI_{tipo} con las longitudes de onda 720, 1200, 1700 y 2280 nm. Esto hasta cierto punto es de esperarse ya que en los 720 nm se localiza la pared roja del espectro, donde se produce el cambio máximo en la pendiente del espectro de reflectancia de la vegetación (Thenkabail et al., 2000).

Las regiones del 1200, 1700 y 2800 nm están fuertemente relacionadas con el contenido de agua en la hoja, esto a su vez se relaciona con la biomasa del dosel y Índice de área foliar (IAF) (LAI por sus siglas en inglés) y a celulosa, almidón, lignina y a concentración de Nitrógeno. Con anterioridad se ha demostrado (Asner, 1998) que la absorción de agua entre los 1200 nm esta relacionada con una dependencia del LAI y por consiguiente con el incremento de la biomasa, aunque esta relación no es directa para el cálculo de la misma.

Existen diversos estudios en laboratorio para estimación de biomasa con datos hiperespectrales, los cuales pueden ser validados con muestras que se obtuvieron en estudios de campo, esto demuestra el potencial de la percepción remota hiperespectral, aunque debe mencionarse que este tipo de estudios relacionan la productividad de la vegetación, esto es una evidencias de la presencia de biomasa por lo tanto mediante modelos de regresiones puede ser estimada.

Capítulo 3

Fundamentos Teóricos de la Metodología

3.1. Aspectos Técnicos

En ésta sección se explican la tecnología LiDAR así como los sistemas HS, se ilustra su funcionamiento y principios básicos sin llegar a ser un documento técnico sobre su funcionamiento, además se presentará como resumen al final del capítulo una tabla con las bandas y canales HS relevantes para el estudio de cultivos y vegetación así como para determinar el estado de salud de las plantas. Se describen también aspectos técnicos que tienen que ver con el proceso y tratamiento que se hicieron a los datos, además de describir las técnicas metodológicas, que se implementaron, de igual manera se argumenta el por que se decidió utilizar la técnica PLS la cual permite extraer las variables latentes que explican la mayor cantidad de varianza del problema, además de utilizar la técnica Variable Importance in Pojection (VIP), con la que seleccionaremos las bandas importantes para la estimación de altura del bosque.

3.2. Enfoque de la Percepción Remota

A continuación se enlistan algunas de las nuevas tecnologías que se han utilizado en diversas aplicaciones y son fundamentales para detección y monitoreo en percepción remota, estas son de gran ayuda a la hora de calcular variables biofísicas y bioquímicas, un ejemplo de ello es la biomasa, sin embargo el propósito de esta tesis

es la estimación de la estructura vertical del bosque.

3.3. LiDAR aerotransportado

Este sistema está basado en la tecnología óptica denominada *láser* (light amplification by stimulated emission of radiation, que significa amplificación de luz por emisión estimulada de radiación), estos dispositivos se basan en la emisión inducida o estimulada, para generar un haz de luz coherente tanto espacial como temporal. La coherencia espacial es la capacidad de un haz para no variar su tamaño y permanecer pequeño al ser transmitido a distancias muy largas. La coherencia temporal es la capacidad del haz para concentrar la emisión en un rango espectral muy estrecho. El láser se desarrolló gracias a la teoría de la mecánica cuántica. El desarrollo del LiDAR aprovecha estas características para utilizarlo en aplicaciones tales como medir distancias, además el láser para poder ser utilizado en LiDAR se necesita que este emita pulsos cortos de alta potencia y con poca divergencia, es decir que abarque un ancho de banda angosto dentro del espectro electromagnético (esta característica se utiliza para las mediciones denominadas por tiempo de vuelo), o en todo caso poder modular su frecuencia, esto para las mediciones denominadas por diferencia de fase (Lohani, 2008; Wehr y Lohr, 1999).

Ahora ya que se tiene una noción básica del principio del LiDAR, se procede a dar una explicación de su funcionamiento. Se considera que es un sensor remoto activo (ya que este internamente genera la radiación que emite), y utiliza señales de láser para la detección de distancias entre el sensor y el objeto de interacción que se está analizando, lo que genera una nube de puntos tridimensional que se asemeja a la superficie del objeto y a los elementos que están sobre esta que reciben las señales del láser, además se cuenta con información de la ubicación del sensor ya que se suele equipar por un receptor GPS (del inglés Global Positioning System, Sistema de Posicionamiento Global) (González, 2013).

En la década de los sesenta y setenta del siglo pasado ocurrieron los primeros esfuerzos por realizar mediciones de la superficie terrestre, montándose en aeronaves. Para los noventa se utilizaron de manera más activa, gracias a la aparición del GPS y las IMU (Inertial Measurement Unit), otro factor relevante fue el avance mismo de los sensores láser, ya que con el avance de las computadoras hubo un mayor almacenamiento de datos y procesamiento de los mismos, así como también el interés por

parte de instituciones e investigadores se vio en aumento (Wagner et al., 2004). Los sistemas LiDAR se clasifican en dos grupos: los que se usan para fines topográficos, montados en la superficie terrestre, y operan en la banda del Infrarrojo (IR), el otro grupo son los utilizados en batimetría que sirven para realizar mediciones del fondo en cuerpos de agua, y operan en las bandas del azul y verde. Adicionalmente se pueden agregar los que operan en la región del ultravioleta visible y NIR (González, 2013).

Para realizar el cálculo de la distancia con LiDAR, existen dos métodos que se pueden aplicar:

3.3.1. Estimación de la distancia por Tiempo de Vuelo

Este es el método más utilizado, en el cual el sensor emite un pulso láser corto de gran intensidad, se mide el tiempo que transcurre desde que este sale del sensor hasta su detección una vez que es reflejado por el objeto o superficie que esta siendo monitoreado. Dado que la velocidad de la luz c se conoce y se conoce el tiempo que tarda el pulso en ir del sensor y regresar, se puede conocer la distancia s recorrida, ya que su trayectoria sera el doble de esta distancia se tiene la siguiente fórmula:

$$s = \frac{v \times t}{2} = \frac{c \times t}{2} \quad (3.1)$$

Donde s es la distancia entre el dispositivo y el objetivo, v es la velocidad del haz es decir la de la luz, y t el tiempo transcurrido entre la emisión y la detección. La figura (3.1) ilustra de manera más clara este proceso:

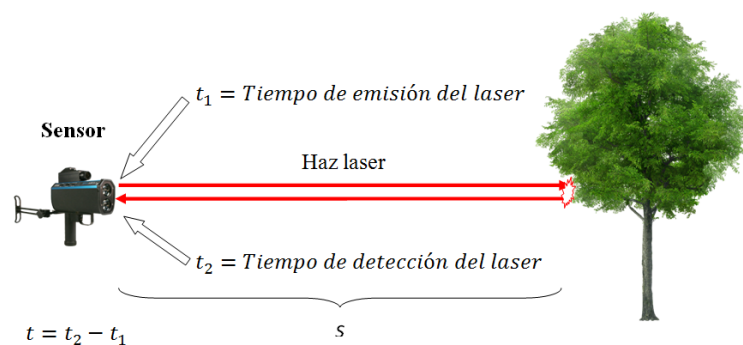


Figura 3.1 Método de medición por tiempo de vuelo.

3.3.2. Estimación de la distancia por Diferencia de Fase

En este caso el sensor debe emitir un haz de energía continuo con frecuencia constante (modulado), con una intensidad inicial al momento de su emisión, y al momento que es detectado al regreso tendrá otra intensidad diferente a la inicial, la cual se amplifica. Gracias a que se puede determinar la diferencia entre las fases inicial en la salida y final a su llegada, esta es medida y transformada a distancia. Se debe contabilizar 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 una fracción de la longitud de onda, y todo dividido entre dos, ya que se considera toda la trayectoria de ida y vuelta (?):

$$s = \frac{n\lambda + \Delta\lambda}{2} \quad (3.2)$$

Donde s es la distancia entre el sensor y el objeto, λ es la longitud de onda de la señal, n es el número de longitudes de onda completas en todo el recorrido del haz y $\Delta\lambda$ es una fracción de la longitud de onda (Shan, 2009b), a continuación la figura siguiente ilustra el método:

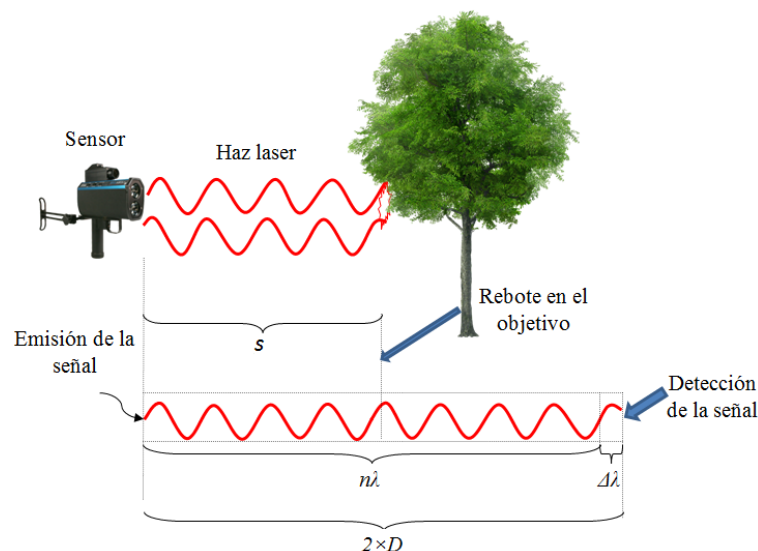


Figura 3.2 Medición por diferencia de fase.

Donde $\Delta\lambda$ se calcula a partir de la diferencia de fase y esta dado en radianes φ , por lo que se tiene:

$$\Delta\lambda = \frac{\varphi}{2\phi} \cdot \lambda \quad (3.3)$$

Para conocer el número de longitudes de onda completas n , se tienen que realizar varias mediciones al mismo objetivo, variando la longitud de onda, ya que no es posible determinarlo en una sola medición, a esto se le conoce como resolución de la ambigüedad, a final de cuentas se generan una serie de ecuaciones como las siguientes:

$$\begin{aligned} s &= \frac{n_1\lambda_1 + (\varphi_1/2\phi)\cdot\lambda_1}{2} \\ s &= \frac{n_2\lambda_2 + (\varphi_2/2\phi)\cdot\lambda_2}{2} \\ &\vdots \\ s &= \frac{n_m\lambda_m + (\varphi_m/2\phi)\cdot\lambda_m}{2} \end{aligned} \quad (3.4)$$

Donde las incógnitas son s y n_m y se encuentran resolviendo el sistema, (Shan, 2009b).

3.4. Percepción Remota Hiperespectral en la Vegetación

En unas cuantas palabras se puede definir a la percepción remota como fotos o una grabación electrónica de la distribución espacial y espectral de radiación electromagnética, la cual es emitida por algún objeto que esta siendo monitoreado ya sean en la superficie de la tierra o bajo de ella, (Thenkabail y Huete, 2016).

Gracias al avance de la percepción remota hiperespectral, se han podido desarrollar nuevas aplicaciones en el monitoreo de la tierra, ahora es posible detectar especies y comunidades de plantas, esto permite monitorearlas directamente utilizando herramientas modernas así como su análisis. Se puede afirmar que el análisis de datos hiperespectrales supera por mucho a los análisis tradicionales de banda ancha, ya que se obtiene una cantidad de información mayor con sensores hiperespectrales.

Algunas de las aplicaciones que se han implementado son a) detección del estrés en las plantas, b) determinar el contenido de clorofila c) identificar pequeñas diferencias en la variación del porcentaje de cobertura verde d) extracción de variables bioquímicas tales como nitrógeno y lignina, e) discriminación de tipos de cobertura,

f) detección de la humedad en los cultivos, g) determinar variaciones ligeras en la concentración de pigmento de las hojas h) modelado biofísico y características del rendimiento de cultivos agrícolas i) mejora en la detección de cambios de vegetación escasa y j) evaluar el contenido absoluto de agua en las hojas de las plantas, (Thenkabail y Huete, 2016). Todas estas aplicaciones dan una idea del potencial que tiene el uso de la percepción remota hiperespectral, por lo que se espera que existan muchos análisis e investigaciones novedosos con toda la información que arrojen estas aplicaciones que resultan muy ambiciosas.

Se sabe que los atributos biofísicos y bioquímicos de la vegetación determinan sus propiedades biofísicas, tal es el caso del índice de área foliar (LAI, por sus siglas en inglés), la cantidad de biomasa viva y biomasa senescente, contenido de humedad, pigmento (clorofila) y el acomodo de su estructura espacial. Ahora es posible medir estos fenómenos y procesos, lo que permite probar hipótesis y crear nuevas aplicaciones en una variedad de ecosistemas y así poder evaluar propiedades biofísicas y bioquímicas de pastizales, cultivos agrícolas y distintas hierbas, con el objetivo de evaluar su productividad, lo que generará información valiosa para agricultores e instituciones, así como también para determinar si el pasto con el que se cuenta es suficiente para alimentar cierta cantidad de ganado (Thenkabail y Huete, 2016).

Con los años se ha probado que los sensores tradicionales de banda ancha suministran información limitada para estimación de variables biofísicas y bioquímicas, tales como crecimiento e identificación de especies forestales, la limitación de estos sensores se hace evidente cuando se trabaja con índices de vegetación (IV), los cuales se saturan a partir de cierta cantidad de biomasa y LAI (Thenkabail et al., 2000), esto nos indica que estos sensores no son aptos para las aplicaciones que se mencionaron, existen también problemas en la saturación cuando se trata de plantas planófilas, además de que las plantas erectófilas se ven más influenciadas por el brillo del suelo.

Por el contrario con conjunto de datos hiperespectrales no se han reportado estas complicaciones, permitiendo la identificación de características medibles de las variables del dosel, tales como contenido bioquímico (clorofila, nitrógeno, lignina), tipo de bosque, distribución química, volumen de madera, agua, etc, y biofísicos por ejemplo biomasa y características de rendimiento (Thenkabail y Huete, 2016).

El desarrollo de los sensores hiperespectrales ha propiciado que surjan nuevas aplicaciones enfocadas en los recursos de la tierra, es una tecnología con mucho potencial, tanto así que se han montado sensores en satélites tales como el Hyperion a bordo del Earth Observing-1 (EO-1) y del Hyperspectral Imagin Spectrometer an

Infrared Imager (HypSIIRI). Este sensor recolecta datos entre las 210 - 220 bandas estrechas que van de los 380 a los 2500 nm con una resolución espacial de 60 m. La banda del infrarrojo térmico (TIR) del HypSIIRI tiene 7 bandas entre los 7500 - 12000 nm y se satura a los 400k y 1 banda en 3000 - 5000 nm que se satura a los 400k y adquiere datos con una resolución espacial de 60 m (Thenkabail y Huete, 2016).

Sin embargo, esta tecnología enfrenta desafíos que debe superar; con la aparición de sensores hiperespectrales, se tienen nuevos retos para el manejo del volumen de información, el cual demanda mucho espacio en disco duro; la tasa de transferencia de datos o ancho de banda para su transmisión también es un problema, esto implica el desarrollo de nuevos algoritmos para el manejo de datos. Las próximas generaciones de satélites, podrán llevar consigo sensores, que solo se enfoquen en captar ciertas bandas estrechas, las cuales provean de información útil para ciertas aplicaciones. Actualmente el sensor Hyperion y HypSIIRI cuentan con esta capacidad, de esta manera los usuarios extraen solamente las bandas estrechas (óptimas) que necesitan. No siempre tener mucha información es mejor, ya que existen datos redundantes, por ejemplo en bandas que son muy cercanas este fenómeno suele darse, esto implica que el usuario tenga que dedicar tiempo esencial para hacer la minería de datos y poder eliminar información que no es óptima, esto dará como resultado un proceso complejo y tardado (Thenkabail y Huete, 2016).

De este tipo de problemas surge uno de los objetivos generales de esta tesis, el cual es poder llegar a implementar un filtro para seleccionar las bandas estrechas que pueden servir para determinar la estructura vertical del bosque de ciertas regiones a partir de datos hiperespectrales, y poder seleccionar solo aquellas que son óptimas y dedicadas a una aplicación específica. Esto implica un ahorro eficiente de tiempo. Los sensores hiperespectrales óptimos que solo detecten las bandas seleccionadas (utilizando algún filtro en el sensor), reducirán el volumen de datos recopilando solo la información relevante para el cálculo de la estructura vertical, esto eliminará el problema de conjuntos de alta dimensión hiperespectral, haciendo posible la implementación de métodos de clasificación tradicionales sobre las bandas óptimas las cuales capturan la mayoría de la información característica de los bosques (y/o cultivos) para reducir los costos de análisis de datos y recursos computacionales.

Existen varios estudios recientes que indican las ventajas de utilizar conjuntos de datos discretos de bandas estrechas de una región específica del espectro cuando se comparan con datos de banda ancha para llegar a ciertas características óptimas en la cosecha o cultivos.

Un obstáculo que se presenta cuando se tienen datos hiperespectrales es el conocido fenómeno de Hughes (Thenkabail y Huete, 2016, p. 11), el cual se describe a continuación: si el número de la dimensión aumenta entonces también crece el número de datos, haciendo que los datos disponibles se vuelvan dispersos. Cuando se tienen datos hiperespectrales y se cuentan con cientos o miles de longitudes de onda, el número de "training pixels" necesita crecer exponencialmente, haciendo difícil direccionar esta diversidad espectral.

El proceso de selección de bandas angostas hiperespectrales que se ajusten mejor para el estudio de vegetación y cultivos agrícolas es complicado, sin embargo se han podido identificar bandas estrechas (véase la tabla 3.1) donde es posible realizar esta tarea, eliminando bandas que son redundantes. Se llegó a determinar bandas hiperespectrales óptimas para el estudio de la vegetación las cuales están comprendidas entre los 400 - 2500 nm, (Thenkabail y Huete, 2016, p. 11).

3.5. Características de los Sensores Hiperespectrales (HSS)

Ahora se describirán algunas de las características de los HSS, que son de importancia, principalmente se sabe que este tipo de sensores son de los más importantes en la percepción remota, ya que es una herramienta muy poderosa para la observación y monitoreo de áreas muy diversas que sean de utilidad en materias económicas, científicas y militares. Se utilizan para monitorear la actividad de la superficie de la tierra o por debajo de ella, se pueden colocar en la superficie de la tierra misma, ser aerotransportados o incluso en satélites que sobrevuelan la órbita terrestre.

Hoy en día se pueden realizar monitoreos y grabaciones electrónicas en bandas del espectro electromagnético que van desde las ondas de radio, hasta los rayos X y ondas Gamma. Sin embargo existen regiones que se utilizan principalmente en percepción remota, estas son la luz visible con el infrarrojo muy cercano (NIR por sus siglas en inglés), térmico, microondas y radio (Ortenberg et al., 2011). En estas regiones los objetos que se monitorean presentan imágenes únicas dentro de cada banda, se caracterizan específicamente por su emisiones, reflectancia, absorción y distribución de ondas electromagnéticas. Los objetos que tienen la capacidad de reflejar la energía solar se encuentran dentro de la región de la luz visible y IR muy cercano, esta capacidad esta determinada por la composición química de la superficie de los

objetos (Ortenberg et al., 2011). Todos los objetos radian energía a consecuencia de su temperatura en consecuencia se localizan en la región del IR térmica (TIR), esta radiación indica la temperatura de la superficie de los objetos. Si se tiene una superficie se puede determinar cual región es la más fría y la más caliente; aunque esta radiación no es visible para el ojo humano, existen técnicas que hacen posible visualizar este comportamiento, una de ellas es el falso color, por medio de la cual se puede asignarse una escala de colores, tradicionalmente se asocia el frío con el azul o violeta y el calor con el rojo o café, existen diversas aplicaciones dentro del TIR, desde la detección de pérdidas de calor en complejos industriales o en zonas residenciales; militarmente se utiliza para localizar puntos de calor invisibles para el ojo humano en la oscuridad; y en la salud humana puede proporcionar información valiosa que determine el estado de salud de las personas. La región de radio de Ultra Alta Frecuencia (UHF), tiene como uso principal el radar, la reflexión de los objetos está determinada por la suavidad de su superficie y su textura, a diferencia de sensores que trabajan dentro de las regiones del visible y del IR, el radar es capaz de atravesar las nubes sin interferencia y ser utilizado de día y de noche, pertenece a la clase de sensores activos, a consecuencia de esto puede obtener patrones de relieve con una precisión de decenas de centímetros (Ortenberg et al., 2011).

Existen imágenes denominadas pancromáticas las cuales están asociadas a sensores que operan en la banda de la luz visible, su principal característica es la resolución. Una cámara localizada en un satélite Low Earth Orbit, Órbita Terrestre Baja (OTB), permite obtener imágenes de la superficie de la tierra con una resolución espacial de 1 m. Además existen sensores que pueden descomponer la región visible del espectro para localizarse en bandas específicas de algún color (p.e. azul, verde, etc), para obtener imágenes multizona con el objetivo de sintetizar las variaciones de color, para determinar detalles que son de interés, tales como caminos, edificios, cuerpos de agua y vegetación (Ortenberg et al., 2011).

Hay que tener en cuenta que la resolución espacial no es la única cualidad que se debe tomar en cuenta, también se debe tomar en cuenta en el desarrollo de estos sensores el número de zonas espectrales es decir el número de canales espectrales, entre mayor sea este número de canales y se obtenga una resolución espectral razonable, entonces el sensor se puede denominar HS (Thenkabail y Huete, 2016).

La región del ultravioleta (UV), puede ser útil en cuanto a su medición para determinar la vegetación que se encuentra saludable de la que se ha degradado, otra aplicación es para diferenciar construcciones hechas por el hombre de formaciones

naturales (Ortenberg et al., 2011).

Es por eso que para la percepción remota las imágenes HS son el método más prometedor, ya que se adquieren cientos o miles de imágenes con bandas espectrales angostas. Cada material tienen una absorción espectral específica, lo que hace relativamente sencillo identificar vegetación, minerales, formaciones geológicas, suelos, estructuras artificiales, pavimento esto gracias a la composición física y química de los objetos. Debido a la alta resolución, comparable a la del orden que se alcanza en un laboratorio con espectrómetros el volumen de información obtenida se incrementa por un factor de miles (Ortenberg et al., 2011).

Las imágenes HS evolucionaron de los sistemas multispectrales (MS), esto debido al desarrollo de nuevas tecnologías, lo que ha implicado que el número de canales de información se incremente de unos 3 - 10 a tener de 100 - 1000, con una alta resolución espectral de 1 - 10 nm. Esto da como resultado una imagen espacial-espectral multidimensional en la cual cada pixel se caracteriza por tener su espectro individual, a estas imágenes se les denomina cubos de información, donde dos dimensiones representan la proyección del área de la imagen sobre el plano, y la tercera es la frecuencia de la radiación recibida (Ortenberg et al., 2011).

EL dispositivo de imagen HS es un sistema óptico-electrónico multicanal, diseñado para la generación independiente simultánea de una imagen y la señal de video correspondiente en una secuencia discreta o continua de intervalos del espectro, complementado por parámetros radiométricos, espectrales y espaciales pertenecientes a esos rangos (Ortenberg et al., 2011).

En resumen los dispositivos actuales Multiespectral (MS) e HS son espectrómetros multicanal calibrados radiométricamente. La asociación de los valores de brillo registrados por el sistema de imágenes para algunos objetos, en zonas espectrales distintas, junto con sus curvas espectrales, permiten identificar plenamente al objeto y percibirlo en la imagen. Como un dato interesante una de las razones por las cuales se ha tenido interés en estos dispositivos óptico-electrónicos no solo es por que existen muchas aplicaciones, si no también por el hecho de que se puede imitar la percepción de muchos animales incluyendo al hombre, es de todos conocido que entre el 80 % y 90 % de toda la información de nuestro alrededor entra por la vista (Ortenberg et al., 2011, p. 41).

Los HSS proveen información vital para diversas áreas quizá por eso su interés por parte de distintas áreas, resolviendo problemas tanto en militares como económicos. Algunos expertos consideran que los HSS son capaces de resolver arriba del 70 % de

los problemas referentes a la observación de la tierra, mientras que la información visual con alta resolución espacial solamente resolverían un 30 %, (Thenkabail y Huete, 2016).

3.6. ¿Qué es un sensor Hiperespectral (HSS)

Este tipo de sensores son la combinación de distintas tecnologías: imágenes convencionales, espectroscopia y radiometría, que al unirse dan como resultado imágenes las cuales tienen asociadas una firma espectral, y cualquier resolución espacial (pixel) (Ortenberg et al., 2011). A continuación en la figura 3.3 se explica el origen de una imagen espectralradiométrica, la cual esta conformada por una imagen radiométrica y otra espectral la primera es adquirida por un radiómetro, este es un instrumento capaz de detectar y medir la intensidad de la energía térmica radiante es decir rayos infrarrojos y luz ultravioleta, esta medición esta dada en Watts (W); el segundo tipo de imágenes es adquirida por un espectrómetro este también mide la longitud de onda en un rango específico, a diferencia del radiómetro este cuenta con una graduación óptica y multiples sensores para dividir la energía entrante en diferentes longitudes de onda o componentes, al ser utilizados con cámaras se pueden medir Watts por metro cuadrado, adicionalmente se puede medir irradiación y dar mediciones de forma muy precisa (Ortenberg et al., 2011).

Los datos producidos por imágenes espectrales generan un cubo, que como ya se había mencionado anteriormente dos de sus dimensiones determinan la posición y la tercera su longitud de onda esto se observa en la figura 3.4, con una adecuada calibración los valores que se obtienen del cubo se pueden convertir a cantidades radiométricas, que se relacionan con escena del fenómeno, por ejemplo la radiancia, reflectancia, emisividad, y otros, además de que dado este hecho se pueden generar modelos espectrales analíticos, librerías espectrales, que dan origen a varias aplicaciones (Ortenberg et al., 2011).

Si el sensor se encuentra en un satélite el movimiento del mismo barre la imagen sobre la superficie de la tierra, mientras el sensor continuamente recopila información, proporcionando un conjunto completo de imágenes HS por linea, el resultado final es un cubo de datos de toda el área, donde cada imagen contiene información de una longitud de onda. La combinación de análisis espacial-espectral, puede ser muy útil tanto así que permite la detección en una imagen de objetos que ópticamente son

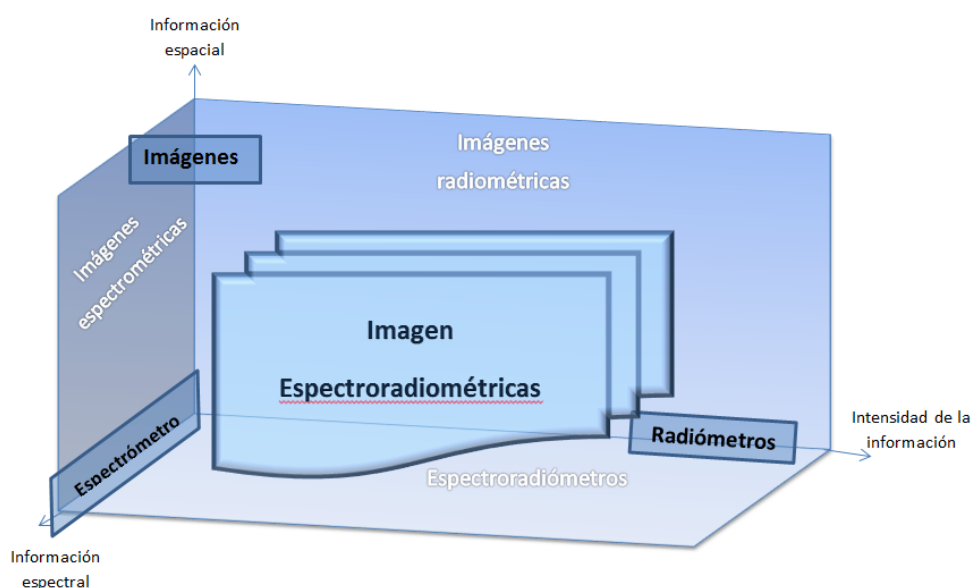


Figura 3.3 Relación entre técnicas radiométricas, espectrométricas y de imagen. (Ortenberg et al., 2011, p. 43)

desconocidos (tamaño subpixel). A esta combinación de imágenes y espectroscopia y su conjunción de análisis se le denomina *fusión de datos* (Ortenberg et al., 2011), la cual es muy útil para mejorar la percepción de imágenes.

Aunque los sistemas de imágenes hiperespectrales manejan datos con alta resolución, sobre un gran número de bandas, y se pueden aplicar en muchas áreas, no significa que desplacen a su contraparte MS aunque estas operan con menos de 15 bandas, tienen sus ventajas, son más baratas, el volumen de datos es menor, por lo que se utilizan en aplicaciones específicas, incluso algunas de sus bandas no son óptimas, por lo que después de pasar por una depuración para obtener las bandas óptimas, pueden proveer soluciones superiores, (Thenkabail y Huete, 2016).

Para que se pueda considerar a una imagen HS, se debe tener un ancho de banda angosto, se puede hacer una comparación entre dos sensores con el mismo número de bandas (canales), por decir 15 las cuales no excedan los 10 nm, si el sensor mide entre los 500 y 700 nm, con bandas estrechas se puede considerar que es HS, a diferencia de otro que con el mismo número de canales abarque las bandas de onda corta, la región visible y los tres rangos de IR, a este último se le considerará MS. Los sistemas de imágenes HS, tienen la capacidad de descomponer la radiación óptica en cientos de bandas, las cuales caracterizaran de manera única a los objetos que están siendo monitoreados, haciendo posible su identificación, además de poder penetrar entre el denso follaje de la vegetación, haciendo posible la identificación de

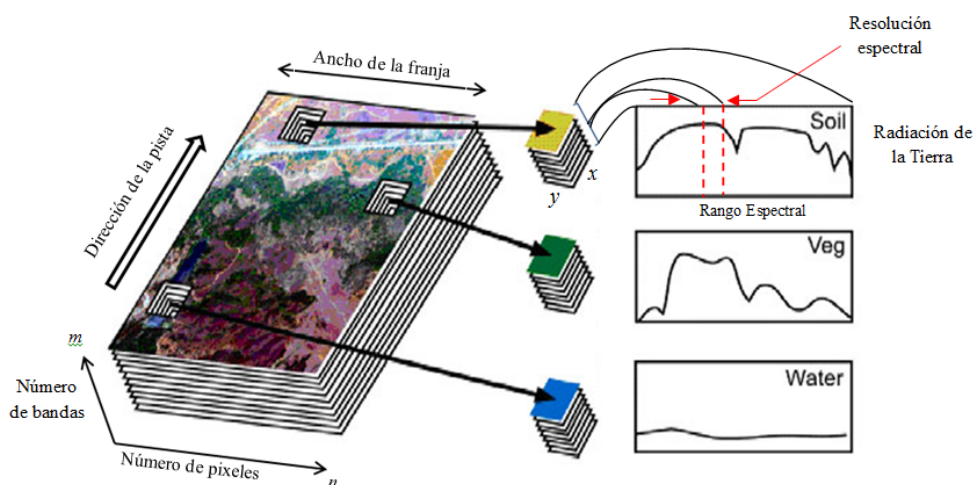


Figura 3.4 Concepto de imagen espectroradiométrica - cubo de datos compuesto por imágenes individuales grabadas sobre m bandas espectrales. Figura adaptada de (Ortenberg et al., 2011, p. 43)

instalaciones bajo tierra, túneles, tuberías y otras (Ortenberg et al., 2011, p. 14). De esta apreciación surgió el prefijo “hiperespectral”, debido a que puede ser interpretado como una hipersensibilidad a objetos enterrados o subterráneos, sin embargo esto es erróneo, lo correcto sería llamarlas imágenes de hiperresolución espectral, lo que se interpreta como una representación visual de datos de un espectroradiómetro. Las imágenes HS son útiles en términos de la cobertura de la vegetación para observar su situación actual, controlar su dinámica actual, pronósticos para el futuro, prevenir catástrofes, recomendaciones de su mitigación o prevención de daños (Ortenberg et al., 2011).

3.7. Principios Físicos y diseño de los sensores HS

Los espectroradiómetros utilizan una matriz en 2D almacenados en circuitos integrados llamados en inglés charge-coupled device (CCD), los cuales producen cubos de datos almacenados progresivamente uno junto al otro, cada uno con una imagen espacial completa y cada uno con diferente longitud de onda, cada píxel contiene su correspondiente firma espectral. Los espectrómetros separan la luz en bandas estrechas mediante un elemento óptico, acomodando adyacentemente bandas de longitudes de onda, la energía en cada banda es medida por un detector diferente, de esta manera se pueden acomodar cientos o miles de bandas estrechas adyacentes de 1

nm, en una región que va de los 400 - 2400 nm (del visible al IR medio). Actualmente se encuentran en desarrollo sensores capaces de cubrir miles de bandas estrechas llamados ultraespectrales (Ortenberg et al., 2011).

Existen dos métodos de adquisición de imágenes HS conocidos como de empuje (electrónico) y de barrido (electromecánico), estos son los más comunes y se describen a continuación:

- Escáner de empuje (Figura 3.5): Este utiliza una línea de sensores en una escena en 2D. El número de píxeles es igual al número de celdas para una franja determinada. La trayectoria del satélite o de la aeronave determina la dirección del escaneo, para calcular el tiempo de escaneo sobre cada píxel se calcula el inverso de la frecuencia de línea (número de líneas escaneadas por segundo). En los sensores de 2D, una de las dimensiones determina el ancho de la franja (dimensión espacial y), y la otra el rango espectral. Algunas de sus ventajas con respecto a los escáner de barrido es que estos son más ligeros, más pequeños y menos complejos, debido a que tienen menos partes mecánicas. También su resolución radiométrica y espacial son mejores. La adquisición de espectrómetros de imagen se clasifican en Imágenes de Campo Amplias (en inglés medium resolution imagin spectrometer, Imágenes espectrométricas de mediana resolución (ISMER), Sistema de Imágenes espectrográficas óptico relectivas (SIEOR)) e Imágenes de Campo Estrecho (Imágenes Hiperespectrales (IHS)). Con estos sensores se pueden obtener coberturas globales, siempre y cuando los HSS sean instalados en un satélite OTB (Ortenberg et al., 2011).
- Escáner de barrido (Figura 3.6): Tienen la óptica en el eje, y hacen el escaneo por medio de espejos desde un borde de la franja hasta el final de la misma. El campo de visión, field of view (FOV) del escáner puede ser cubierto cubierto por un único detector en una sola línea de detección, la franja esta determinada por el movimiento del satélite o aeronave que lo transporte. El tiempo de permanencia en cada píxel es muy corto, en un campo de visión instantáneo (en inglés Instantaneous field of view, Campo de visión instantaneo (IFOV)), determinado, ya que cada línea escaneada esta comprendida por multiples celdas en la tierra que serán cubiertas. Su desventaja es que pueden ser grandes y complejos, el movimiento de los espejos genera distorsión espacial, necesitando un preprocesamiento para ser corregida. Sus ventajas es que no se necesita mucha calibración ya que el número de sensores es reducido. Ejemplos de esos son en inglés Advanced very high resolution radiometer, radiómetro avanzado de muy alta re-

solución (AVHRR), Land remote sensing satellite program, Programa satélite de teledetección (Landsat) y en inglés sea-viewing wide field-of-view sensor, Sensor de campo de visión de gran amplitud (SeaWiFS) (Ortenberg et al., 2011).

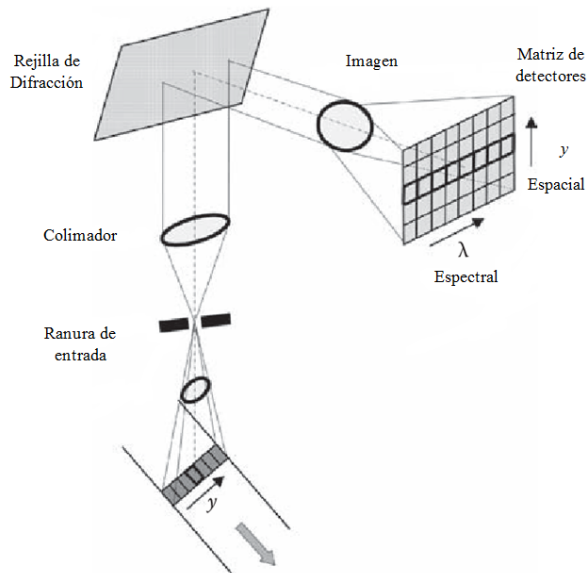


Figura 3.5 Principio de Sensor de empuje, (Ortenberg et al., 2011).

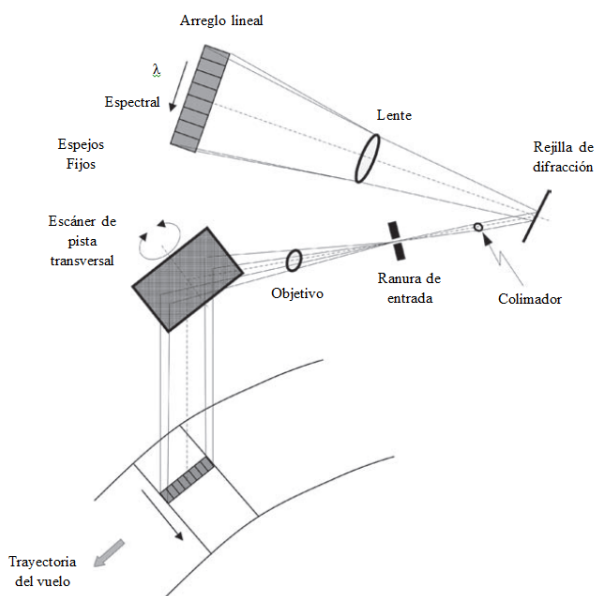


Figura 3.6 Principio de Sensor de barrido,(Ortenberg et al., 2011).

Existen dos enfoques en los que se basan los sensores HS, por dispersión, que se utilizan en ambos sistemas el de barrido, empuje y por transformada de Fourier este se

utiliza más en sensores con espejos fijos. Cuando se realizan estudios de la superficie de la tierra que son muy detallados, en las bandas del 400-2400 nm, el equipo de imágenes HS debe proveer simultáneamente una alta resolución espacial entre 1-5 m, alta resolución espectral de 5-10 nm, en decenas o cientos de canales, y anchos de bandas (swath en inglés) por encima de los 100 km y relación de señal-ruido ≥ 100 , (Thenkabail y Huete, 2016).

3.8. Modos de Operación de los HSS

Actualmente con el avance de las técnicas de adquisición de imágenes HS, se han desarrollado diversos métodos para este fin, principalmente existen tres, los cuales permiten apreciar excelentes imágenes de la superficie de la tierra con alta resolución espectral: En tierra, aerotransportado e imágenes espaciales, todos tienen sus ventajas y desventajas particulares, se puede trabajar con los tres simultáneamente para obtener los resultados deseados de ser posible. A continuación se da una explicación de estas técnicas:

3.8.1. Imágenes HS con base en la tierra

Esta técnica inicialmente se puso en práctica en laboratorios detectando cualidades químicas de la materia, esto gracias a la existencia y avance de los espectroradiómetros y por ende a la espectroscopia, la cual estudia la emisión, absorción y reflectancia de la luz que incide en la materia esto como una función de la longitud de onda de la radiación. En aplicaciones de teledetección óptica, la espectroscopia trata con el espectro de la luz solar que es reflejada difusamente (dispersa) por los distintos materiales que se encuentran en la superficie de la tierra. Ya sea en el laboratorio o en bases terrestres las mediciones de la luz reflejada por los distintos materiales se considera una de las primeras aplicaciones exitosas de los espectrómetros o espectroradiómetros (Ortenberg et al., 2011).

Algunas de las aplicaciones de las bases terrestres de obtención de imágenes HS son: detección de objetivos oscuros o de poca resolución, detección de camuflaje, detección de químicos sólidos o gaseosos. Utilizando datos HS y MS diversas industrias manufactureras han podido realizar análisis para inspeccionar el color, es decir determinar la calidad de la pintura, detección de moho, así como detección de fallas

en el revestimiento de películas, entre otras. Existen sensores con base terrestres, con herramientas de software que incluyen calibraciones radiométricas completas, cálculo de reflectancia, filtros adaptados y corrección atmosférica. Algunos investigadores utilizan estos sistemas adaptados a sistemas de riego colocando sensores, los cuales realizan muestreos por encima de la cosecha, permitiendo obtener datos HS, los cuales proveen mucha mayor resolución que los sistemas aerotransportados o satelitales. Otros sistemas con base terrestre se adaptan a camionetas y realizan mediciones en tiempo real, estos monitorean los cultivos y sensores que controlan atomizadores aumentando o disminuyendo la cantidad de herbicida suministrado al cultivo sobre la marcha en tiempo real. A esta aplicación tecnológica de la percepción remota se le conoce como visión de máquina y se emplea actualmente, (Thenkabail y Huete, 2016).

3.8.2. Sensores HS Aerotransportados

Durante las últimas décadas del siglo pasado las imágenes HS se consideraba un área en investigación y desarrollo, lo que hacía que las imágenes adquiridas estuvieran al alcance solo de investigadores, pero recientemente han aparecido sistemas aerotransportados comerciales lo que ha promovido su uso en diversas áreas particularmente en agricultura, permitiendo su monitoreo, exploración entre otras aplicaciones (Ortenberg et al., 2011).

Al igual que los equipos terrestres, los sistemas HS aerotransportados miden la intensidad de la radiación dispersa por la superficie de la Tierra en ciertas bandas espectrales, las imágenes adquiridas con sus píxeles tendrán simultáneamente información espectral y espacial del terreno inspeccionado. El flujo de radiación de salida, se divide en varias componentes, de acuerdo a longitudes de onda. Para cada banda de longitud de onda, se reservan ciertas líneas de la matrix. Esta información es digitalizada y escrita directamente en una unidad de disco duro (Ortenberg et al., 2011).

La colección de imágenes se vincula espacialmente gracias a el en inglés Inertial Navigation System, Sistema de Navegación Inercial (INS), la información de este se utiliza en el procesamiento post vuelo, junto con los datos del escaneo de las imágenes, con el fin de agregar la información adecuada de las coordenadas geográficas para cada pixel. Existen sistemas avanzados muy complejos, los cuales controlan el proceso de adquisición de imágenes con software especializado, haciendo posible controlar la aeronave y los sensores durante el vuelo (Ortenberg et al., 2011).

Habitualmente los escáneres HS se conforman por un sensor (unidad optoelectrónica), fuente de poder, unidad de control y acumulación de datos, plataforma de amortiguación de vibraciones, pantalla de operación, receptor GPS, unidad inercial de medición y un dispositivo de procesamiento de datos (Ortenberg et al., 2011).

Los sistemas de imágenes HS aerotransportados se utilizan para reconocer objetos por sus características fisicoquímicas, haciendo posible identificar especies de plantas y el estado de la vegetación, permitiendo mantener controlado el ambiente natural de las diferentes plantas, incluyendo hierbas, plantas medicinales, manchas de vegetación y anomalías en el suelo, encontrar humedales y suelos salados, entre otros (Ortenberg et al., 2011).

Estos sistemas se pueden montar, en aviones, drones, helicóptero y zepelín, cada uno tiene sus ventajas, los aviones pueden volar a relativamente bajas alturas, admitiendo imágenes con resolución espacial submétrica. Los sistemas aerotransportados son flexibles, debido a la variación de la altura de vuelo y a al cambio de la velocidad de vuelo. Además la resolución espacial y el ancho de la franja pueden adaptarse fácilmente de acuerdo a los requerimientos de la tarea asignada, durante el vuelo (Ortenberg et al., 2011).

Los aviones pueden cambiar rápidamente su horario de monitoreo, para evitar problemas climatológicos, tales como las nubes, que pueden bloquear los sensores pasivos, permitiendo dirigirse a una zona con mayor iluminación solar. El mantenimiento a plataformas que se encuentran en aviones es relativamente sencillo, prácticamente no existen fronteras para estos sistemas. Al mismo tiempo si se hace un vuelo a baja altura, y se cuenta con un sensor con FOV estrecho se tendrán que realizar varios vuelos para cubrir áreas muy grandes. La información se obtendrá una vez que el avión haya aterrizado. Además puede ser integrado con sistemas LiDAR que operan en otras bandas espectrales.

Los escáneres HS son los sistemas más avanzados actualmente en percepción remota de la Tierra, algunos de estos son AVIRIS, HYDICE, AISA, HyMAP, ARES, CASI 1500 y AisaEAGLET (este último se utiliza para monitoreo de bosques, agricultura y defensa) (Thenkabail y Huete, 2016).

3.8.3. Sensores HS montados en Satélites Artificiales

Actualmente los satélites artificiales permiten que se instalen sensores que monitorean áreas de interés durante los trayectos del satélite, permitiendo una cobertura global de la superficie de la Tierra periódica. La resolución de la imagen se define por los parámetros de la órbita del satélite; principalmente su altitud. Una de sus ventajas es que al estar en el espacio no obedecen fronteras políticas, por lo que pueden barrer cualquier parte de el globo, independientemente de la altura de vuelo. Su principal debilidad es el costo multimillonario de estos sistemas para su desarrollo, y su soporte en tierra, además de su corto tiempo de durabilidad, cerca de 5 años o menos. Lo que ha llevado a desarrollar minisatélites (entre 100 - 500 kg), los cuales implican menor gasto para su lanzamiento para ponerlos en órbita (Ortenberg et al., 2011).

Se espera que en los próximos años el proyecto Vegetation an Environment New Micro Satellite (VENuS) sea lanzado, este se utilizará principalmente en actividades de monitoreo de vías agrícolas como control ecológico, basado en fotografías de alta definición. Este tendrá un peso de 260 kg, y se espera que su vida útil sea de más de 4 años. Se podrán obtener franjas de 27.5 km, y podrá identificar objetos de menos de 5.4 m desde una altitud de 720 km. Las fotografías contendrán 12 canales espectrales estrechos entre los 415 - 910 nm, cada banda varia entre los 16 - 40 nm. Una desventaja de estos sistemas es que son costosos y llegan a fallar incluso algunos días después de ser puestos en órbita, es por esto que muchos de estos satélites comerciales están en constante riesgo. Aún se presentan muchos retos en la industria para satélites miniatura, las restricciones de peso y dimensión, a la vez restringen la posibilidad de colocar otros sensores los cuales permitan realizar correcciones atmosféricas por ejemplo, por otra parte a menor peso se tiene un número menor de bandas espectrales, y al contrario si se aumenta la masa del satélite la resolución espacial es mucho mejor y el número de bandas espectrales también aumenta, por lo que se pretende llegar a un equilibrio en el que se pueda aprovechar eficientemente el tamaño del sistema HS espacial con cualidades optoelectrónicas ideales. Es por esto que se siguen realizando esfuerzos para mejorar la tecnología existente en un futuro cercano. La tabla 3.1 que aparece a más adelante presenta un resumen de las bandas y canales que son relevantes para nuestro estudio, detallando la importancia de distintas franjas dentro del espectro electromagnético, que van de los 405 nm a los 2295 nm. De esta se desprenden distintas aplicaciones y usos (Ortenberg et al., 2011).

Tabla 3.1 (Thenkabail y Huete, 2016, p. 665-666)

Bandas Estrechadas (o Angostas) Hiperespectrales Óptimas entre 400-2500 nm para el estudio de Vegetación y Cultivos Agrícolas, Propiedades Biofísicas y Bioquímicas y para Clasificación de Tipos y Especies		
Número serial	Centro de la banda	Nombre de la variable de la planta
<i>A. Bandas Azules</i>		
1-2	405, 450	<i>Senescencia, nitrógeno</i> : Sensibilidad al cambio de nitrógeno en la hoja. Absorción significativa debido a la clorofila y carotenoides: cambios de la reflectancia debido a los pigmentos de moderada a baja, sensible a la detección
3	490	<i>Eficiencia del uso de luz (EUL), carotenoides, estrés en vegetación</i> : Sensible a la senescencia y pérdida de clorofila, maduración, rendimiento del cultivo, y efectos en el fondo del suelo.
<i>B. Bandas Verdes</i>		
4	515	<i>Pigmentos(carotenoides, clorofila, antocianinas), nitrógeno, vigor</i> : Los cambios positivos por unidad de área en el espectro visible y es máximo al rededor de esta banda verde.
5	531	<i>EUL, ciclo de la xantofila, estrés en vegetación, parásitos y enfermedades</i> : Sensible a la senescencia y pérdida de clorofila, maduración, rendimiento del cultivo, y efectos en el fondo del suelo.
6	550	<i>Antocianinas, clorofila, IAF, nitrógeno, EUL</i> : Sensible a numerosas variables de vegetación
7	570	<i>Pigmentos (carotenoides, clorofila), nitrógeno</i> : El cambio negativo en la reflectancia por unidad de cambio en la longitud de onda es máximo como resultado de la sensibilidad al vigor de la vegetación, pigmento y N.
<i>C. Bandas Rojas</i>		
8	650	Pigmentos, nitrógeno: Sensibilidad moderada hacia alta a cambios en pigmentos (clorofila, antocianinas) y Nitrógeno.
9	687	<i>Cantidades biofísicas, clorofila, fluorescencia de la clorofila inducida por el sol</i> : IAF, biomasa, rendimiento, discriminación de tipo de cultivo. Mejor contraste suelo-cosecha. Picos de emisión inducidos activamente en rojo / rojo lejano 687 y 740 nm.
<i>D. Bandas de Borde Rojas</i>		
10 - 11	705,720	<i>Detección de estrés en la vegetación en el borde rojo, estrés y sequía</i> : estrés de nitrógeno, estrés en el cultivo, estudios de etapas de crecimiento en los cultivos. Cambio del rojo al azul en caso de estrés. Cambio hacia el NIR en vegetación saludable.
12	700 - 740	<i>Clorofila, senescencia, estrés, sequía</i> : El índice de derivados de primer orden sobre 700-400 nm, tiene aplicaciones en estudios de vegetación (p.e. cambio al azul en caso de estrés y cambio al rojo en caso de crecimiento saludable).

Tabla 3.1 (Thenkabail y Huete, 2016, p. 666-667).*(Continuación)*

Bandas Estrechas (o Angostas) Hiperespectrales Óptimas entre 400-2500 nm para el estudio de Vegetación y Cultivos Agrícolas, Propiedades Biofísicas y Bioquímicas y para Clasificación de Tipos y Especies. (Continuación)		
Número serial	Centro de la banda	Nombre de la variable de la planta
<i>E. Bandas NIR</i>		
13	760	<i>Biomasa, IAF, emisiones pasivas inducidas por energía solar:</i> Banda de referencia NIR, para diversos índices. Emisiones pasivas inducidas por energía solar con las recuperaciones realizadas en las características atmosféricas de CO ₂ en 687 y 760 nm.
14	855	<i>Cantidades biofísicas/bioquímicas, estrés relacionado a metales pesados:</i> IAF, biomasa, rendimiento, discriminación de cultivos, clorofila, antocianinas, carotenoides. Sensible a metales pesados relacionados con el estrés debido a la reducción de clorofila. Alta estabilidad en bandas del NIR para desarrollo de índices
15	970	<i>Banda relacionada con absorción de agua:</i> Canal más prominente en la absorción de agua. Útil también para cálculo de propiedades biofísicas y bioquímicas.
16	1045	<i>Cantidades biofísicas, bioquímicas:</i> IAF, biomasa húmeda y seca, altura de las plantas, rendimiento de granos, tipo de cultivos, discriminación de cultivos, clorofila total, antocianinas, carotenoides.
<i>E. Bandas del Far Near Infrared, infrarrojo muy cercano (FNIR)</i>		
17	1100	<i>Cantidades biofísicas:</i> Sensible a biomasa e IAF.
18	1180	<i>Banda de absorción de agua</i>
19	1245	<i>Sensible a Agua:</i> Índice de banda de agua, agua en las hojas, biomasa. Pico de la reflectancia en 1050-1300 nm.
<i>F. Bandas del Early short-wave infrared, Infrarrojo de onda corta temprana (ESWIR)</i>		
20	1450	<i>Banda de absorción de agua:</i> Muy alta absorción de agua a través de infrarrojos de onda corta temprana ESWIR. Se utiliza como índice con 1548 o 1620 o 1690 nm.

Tabla 3.1 (Thenkabail y Huete, 2016, p. 667-668). (Continuación)

Bandas Estrechas (o Angostas) Hiperespectrales Óptimas entre 400-2500 nm para el estudio de Vegetación y Cultivos Agrícolas, Propiedades Biofísicas y Bioquímicas y para Clasificación de Tipos y Especies. (Continuación)		
Número serial	Centro de la banda	Nombre de la variable de la planta
G. Bandas Bandas Far short-wave in- frared, Infrarrojo de onda corta lejana (FSWIR)		
26	1950	<i>Banda de absorción de agua:</i> Canal de mayor absorción de humedad en FSWIR. Se utiliza como un índice en cualquiera de los siguientes canales 2025, 2133 y 2213 nm. Se ve afectado en ocasiones por ruido.
27	2025	Hojas de árboles caídas en el suelo, lignina, celulosa, diferenciación de hojas caídas en el suelo.
33	2295	<i>Estrés y contenido de hierro en el suelo:</i> Sensible al fondo del suelo y al estrés de las plantas.

3.9. Mínimos Cuadrados Parciales (PLS)

La regresión de mínimos cuadrados parciales (PLS, Partial Least Square), es un método que tiene aplicaciones en el campo de la bioestadística para el procesamiento de cadenas de ADN y ARN, de igual manera también en Quimiometría se pueden encontrar algunas aplicaciones para determinar la relación de diversas sustancias y su respuesta a ciertas longitudes de onda, sus orígenes fueron en el área de las ciencias sociales. El origen del método se les atribuye al sueco Herman Wold alrededor de 1975 y a su hijo Svante Wold (Svante Wold, 2001).

La razón por la cual se decidió utilizar este método es debido a que se realizó una revisión bibliográfica y se observó que existen registros de estudios previos en los que se ha implementado con buenos resultados (Svante Wold, 2001), ya que este método relaciona dos matrices X e Y , mediante un modelo lineal multivariado, a la matriz X , se le conoce como matriz de factores esta tiene una dimensión de $N \times K$ y a Y matriz de respuesta, su dimensión es de $(N \times M)$, adicionalmente es de ayuda cuando existe la presencia de datos con mucho ruido como en el caso de los datos que estamos analizando o incluso cuando las variables están incompletas en

X o Y, evitando el sobreajuste de los datos, lo cual es común cuando se tiene que el número de factores es mucho más grande que el número de observaciones, en estos casos es cuando es útil utilizar el método de PLS. Cuando se tiene una gran cantidad de factores, es evidente que no todos serán relevantes y solo unos pocos serán considerados fundamentales, ya que estos aportan la mayor parte de variación de la respuesta, a estos factores fundamentales se les denominará factores latentes, estos últimos son los que se encarga el PLS de extraerlos siempre que sea posible y modelen bien la respuesta, es por esto que a PLS se le denomina como "proyección de la estructura latente"(Svante Wold, 2001).

Estos factores latentes se denominan $X - scores$ (con una dimensión de $N \times A$) y serán utilizados para predecir \mathbf{Y} y de igual manera modelan a \mathbf{X} por lo que se puede suponer que \mathbf{Y} y \mathbf{X} serán modelados al menos parcialmente por los mismos coeficientes lineales (se muestra en las ecuaciones (3.6) y (3.8)), a continuación se muestran las ecuaciones que dan origen al método de PLS (Svante Wold, 2001), del lado derecho entre paréntesis se aprecian en forma matricial y al lado izquierdo en forma de elementos, se tiene que los $X - scores$ se denotan como t_a ($a = 1, 2, 3, \dots, A$) (donde A representa el número de componentes del modelo), y son el producto de la variable original x_k con los coeficientes denominados "pesos" w_{ka}^* (donde $a = 1, 2, \dots, A$; $k = 1, 2, \dots, K$ e i es el número de observaciones, tal que $i = 1, 2, \dots, N$). Adicionalmente los t_a son ortogonales y se estiman a partir de una combinación lineal de las variables x_k originales, con los coeficientes w_{ka}^* "pesos" ($a = 1, 2, 3, \dots, A$), estos a su vez generan la matriz de pesos W^* .

$$t_{ia} = \sum_k w_{ka}^* x_{ik}; \quad (T = XW^*) \quad (3.5)$$

Las propiedades que poseen los t_a , son: (a) Son una buena manera de "explicar" a \mathbf{X} cuando son multiplicados por los $X-loadings$ (en adelante se denominan p_{ak}), lo que dará como resultado que los residuales de \mathbf{X} e_{ik} en la ecuación 3.6 sean "pequeños":

$$x_{ik} = \sum_a t_{ia} p_{ak} + e_{ik}; \quad (X = TP' + E) \quad (3.6)$$

Cuando se tiene el caso que \mathbf{Y} es multivariada, (es decir $M > 1$), entonces se tiene que al multiplicar "Y-score" (u_a), por su respectivo peso c_{am} , se puede explicar adecuadamente el comportamiento de la matriz de respuesta \mathbf{Y} , esto implica que los

residuos g_{im} tienda a ser muy “pequeños”:

$$y_{im} = \sum_a u_{ia} c_{am} + g_{im}; \quad (Y = UC' + G) \quad (3.7)$$

Donde $m = 1, 2, \dots, M$ es el índice de la variable Y .

(b) por lo tanto los X -scores son buenos prediciendo a Y , o lo que es igual:

$$y_{im} = \sum_a c_{ma} t_{ia} + f_{im}; \quad (Y = TC' + F) \quad (3.8)$$

Con base a lo anterior se tiene que los residuos f_{im} , expresan la desviación entre lo observado y la respuesta que se está modelando, y esta representada por la matriz de residuos F .

Se puede observar que la ecuación (3.5) se puede sustituir en (3.8), por lo que se puede expresar para que parezca un modelo de regresión múltiple:

$$y_{im} \sum_a c_{ma} \sum_k w_{ka}^* x_{ik} + f_{im} = \sum_k b_{mk} x_{ik} + f_{im}$$

$$Y = XW^*C' + F = XB + F. \quad (3.9)$$

Por lo tanto los coeficientes de la regresión b_{mk} (B), del modelo PLS se pueden escribir como:

$$b_{mk} = \sum_a c_{ma} w_{ka}^* \quad (B = W^*C') \quad (3.10)$$

Note que los b 's no son independientes a menos que el número de componentes de la regresión PLS sea igual al número de variables del vector X . Por lo tanto su intervalo de confianza es infinito.

Cuando se da el caso de que solo existe una única variable y ($M = 1$), entonces $X'X$ es una matriz diagonal, proveniente de una base ortogonal. Para este caso se reduce a una regresión multilínea (RML) de una sola componente (Manne, 1987), por lo que la solución para regresión PLS coincide con la de la RML, además de que los coeficientes también coinciden y son de la forma $w_1 c'_1$.

Al aplicar una técnica de Algebra Lineal denominada “deflation” (“desinflar” en español), la cual elimina la influencia de un eigenvector dado (Mackey, 2009), la cual

se aplica a cada componente a , de la matriz \mathbf{X} , es decir restando $t_{ik}p_{ka}$ de x_{ik} . Con esto se puede expresar el modelo alternativo de la regresión PLS con pesos w_a (estos generan la matriz W_a), los cuales hacen referencia al error residual E_{a-1} , en lugar de relacionarse a la variable X por si misma. Dicho lo anterior se puede reescribir la (3.5) de la siguiente forma:

$$t_{ia} = \sum_k w_{ka} e_{ik,a-1}; \quad (t_a = E_{a-1} W_a) \quad (3.7a)$$

$$e_{ik,a-1} = e_{ik,a-2} - t_{i,a-1} p_{a-1,k}$$

$$(E_{a-1} = E_{a-2} - t_{a-1} p'_{a-1}) \quad (3.7b)$$

$$e_{ik,0} = X_{ik} \quad (E_0 = X). \quad (3.7c)$$

Para relacionar directamente este resultado con la matrix \mathbf{X} de la ecuación 3.5, los pesos \mathbf{w} , se pueden transformar a w^* , esta relación estará expresada de la siguiente manera (Manne, 1987):

$$W^* = W(P' - W)^{-1} \quad (3.12)$$

No es necesario realizar la técnica “deflated”(“desinflado”) para la matriz \mathbf{Y} , ya que se obtiene prácticamente el mismo resultado que sin aplicar la técnica (Svante Wold, 2001).

Como se menciona en (Svante Wold, 2001), el primer vector de peso w_1 corresponde al primer eigenvector de la matriz combinada de varianza y covarianza $Z'_a Y Y' Z'_a$, donde $Z_a = Z_{a-1} - T_{a-1} P'_{a-1}$, así mismo el primer vector de scores es un eigenvector de $X X' Y Y'$, además de que los los vectores t_a son eigenvectores de $Z_a Z'_a Y Y'$.

Esto demuestra que los vectores w_a generan un conjunto ortonormal, y que los vectores t_a , son ortogonales unos de otros. También que los vectores p_a no son ortogonales unos con otros, y tampoco lo son los u_a . Respectivamente los vectores \mathbf{u} y \mathbf{p} , son ortogonales a \mathbf{t} y a \mathbf{w} , i.e. $u'_b t_a = 0$ y $p'_b w_a = 0$, si $b > a$. Además de que $w'_a p_a = 1$.

3.10. Interpretación del modelo de regresión PLS

La regresión PLS se puede explicar de tal manera que esta crea un conjunto de x -variables nuevas, t_a , las cuales son combinación lineal de las variables x originales, de tal manera que estas nuevas variables sirven para estimar la matriz Y . Por lo tanto en inglés PLS regression, regresión PLS (PLSR) se basa en un modelo lineal. Solo se crearan las t_a necesarias mientras sean significativas para la predicción (Svante Wold, 2001).

Todos los parámetros t , u , w , w^* , p y c son determinados por el algoritmo de la PLSR descrito anteriormente. Se pueden describir los scores t y u como variables que contienen información sobre los objetos y sus similitudes y diferencias con respecto a un problema y modelo dado (Svante Wold, 2001).

Los pesos w_a los cuales son muy similares a los pesos w_a^* , y c_a , proporcionan información acerca de como las variables se combinan para formar la relación cuantitativa entre X e Y , proporcionando así una interpretación de los scores t_a y u_a . En esencia los pesos nos indican cuales variables de X son importantes (es decir valores numéricos grandes para w_a), y cuales variables de X proporcionan la misma información. Ambos pesos w_a expresan la correlación positiva entre X e Y , y la correlación que se tiene que compensar necesaria para predecir Y a partir de X , (Svante Wold, 2001).

3.10.1. Interpretación Geométrica del modelo PLS

La regresión del PLS se puede interpretar geoméricamente como una proyección de la matriz X , la cual contiene una nube de N puntos en un espacio de dimensión K , donde cada columna de X (x_k) define un eje coordenado. La regresión PLS genera un hiperplano de dimensión A , que a su vez es definido por una línea y una dirección por componente. Los coeficientes de esta dirección están dados por p_{ak} . Las coordenadas de cada objeto, i , cuando sus datos (fila i en X) son proyectados en este plano son t_{ia} ($a = 1, 2, \dots, A$), esto indica que se pueden utilizar como buenos estimadores de Y , el caso para dos componentes se puede visualizar en la figura (3.7), (Svante Wold, 2001).

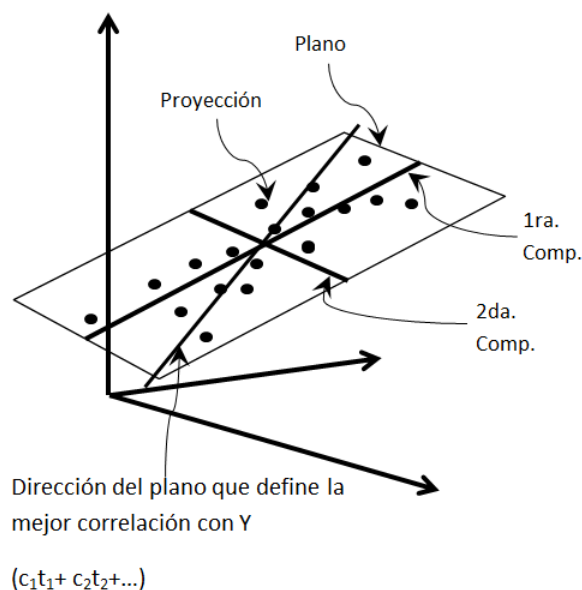


Figura 3.7 Representación geométrica de la regresión PLS

3.10.2. Selección del Número de Componentes

Uno de los problemas que se presentan con datos que están altamente correlacionados, tal es el caso de este estudio, es el sobreajuste de los datos, para esto se debe tener cuidado de seleccionar un número adecuado de componentes (A), esto se logra utilizando alguna técnica que detenga el número de componentes cuando estas comiencen a ser no significativas (Svante Wold, 2001).

La Validación Cruzada (VC), es un técnica que es de confianza y útil para probar cuando el número de componentes pierde significancia (Svante Wold, 2001).

La VC lo que hace es dividir los datos en un número de grupos G , para después generar un grupo de modelos paralelos con los datos reducidos con uno de los grupos eliminado. No se recomienda utilizar esta técnica con el total de los grupos es decir cuando el número de grupos sea igual al número de puntos ($G = N$) (Svante Wold, 2001).

Una vez que se genera un modelo, las diferencias entre los valores de la Y estimada y la Y original se calculan para los datos eliminados. Posteriormente la suma de cuadrados de estas diferencias es calculada y almacenada junto con la de los demás modelos que se calcularon en paralelo, para formar la suma de cuadrados residuales predictivos (PRESS por sus siglas en inglés, predictive residual sum of squares), la cual estima la capacidad predictiva del modelo (Svante Wold, 2001).

Se puede recurrir a la implementación de la VC de manera *secuencial*, es decir en una componente y posteriormente seguir con la siguiente, se debe mencionar que el cálculo de la ecuación 3.7b solo se realiza una vez para las matrix completas, posteriormente las matrices residuales **E** y **F** se dividen en grupos para la VC de la siguiente componente. Se calcula el cociente $PRESS_A/SS_{a-1}$ después de que se calculan las componentes y esta es calificada como significativa si el cociente es menor a 0.9 aproximadamente, esto se hace al menos para una de las variables de la matrix Y. SS_{a-1} denota la suma residual de cuadrados de la componente anterior que se esta analizando. Este proceso continua hasta que una componente no sea significativa. El número de componentes se puede establecer manualmente desde el principio del algoritmo para posteriormente obtener un número de componentes óptimo de acuerdo a las necesidades del problema (Svante Wold, 2001).

Otro alternativa es utilizar la VC *total*, donde primero se dividen los datos en grupos, y se calcula el PRESS para cada componente, de tal manera que el grupo con el menor cociente $PRESS/(N - A - 1)$ se utiliza. Este método es más costoso computacionalmente, además de que arroja resultados similares. Este tipo de técnicas son útiles para validar el modelos cuando no se cuenta con un conjunto de datos independiente para realizar una validación (Svante Wold, 2001).

3.11. Importancia predictiva de la variable (VIP)

VIP en inglés variable importance in projection, es una metodología que se utilizó en la selección de las bandas de mayor importancia, las cuales a fin de cuentas contribuyen mayormente a la estimación de la variable Y, éste método genera un índice el cual define un límite, si este es mayor que uno, se tomarán estas bandas como relevantes, pero si el índice es menor a uno, se descartaran estas bandas para el análisis posterior. Cabe señalar que existen diversos métodos para la selección de las bandas relevantes, sin embargo se optó por el VIP, debido a su facilidad para implementarlo y al menor costo computacional (Mehmood et al., 2012).

Este método trata de medir la influencia de cada variable j reflejada por los pesos w de cada componente. El índice VIP esta dado por la fórmula:

$$VIP_j = \sqrt{p \sum_{a=1}^A [SS_a(w_{aj}/\|w_A\|^2)] / \sum_{a=1}^A (SS_a)} \quad (3.13)$$

Donde como se vio anteriormente SS_a es la suma de cuadrados explicada por la componente a -ésima componente. Cada uno de los índices VIP_j , explicaran la contribución de cada variable esto de acuerdo a la varianza explicada por cada componente de la regresión PLS, de este modo $w_{aj}/\|w_A\|^2$ servirá para medir la importancia de cada j -ésima variable. La varianza explicada para cada componente se calcula con el término $q^2 t'_a t_a$, por lo que se puede reescribir la expresión 3.13 de la siguiente forma:

$$VIP_j = \sqrt{p \sum_{a=1}^A [q^2 t'_a t_a (w_{aj}/\|w_A\|^2)] / \sum_{a=1}^A (q^2 t'_a t_a)} \quad (3.14)$$

Para nuestro caso esta expresión se reduce ya que los X -scores son ortogonales y el producto $t'_a t_a = 1$, debe recalcar que si el número de componentes del PLS se incrementa, entonces la varianza explicada por componente disminuye y la importancia de las variables disminuye progresivamente (Gosselin y Duchesne, 2010).

Al calcular los índices VIP_j , generalmente se seleccionan los que son mayores a 1, (Chong y Jun, 2005; Eriksson et al., 2001; Gosselin y Duchesne, 2010), aunque también en (Chong y Jun, 2005) se menciona que seleccionando las variables que estén entre $0.83 < VIP < 1.21$ se pueden obtener resultados mejores, los valores de $VIP < 0.8$ representan variables menos importantes. La contribución de cada variable se puede representar mediante un porcentaje si se desea. Se puede aplicar el método probabilístico bootstrapping si se requiere, para mejorar la estabilidad de los resultados comparado con los coeficientes originales de la regresión, existen diversos métodos para la selección de variables sin embargo se utilizó el VIP gracias a su sencillez para implementación y el bajo costo computacional.

3.11.1. Bootstrap y selección de variables

Este es un método estadístico de remuestreo, que permite estimar errores estándar y sesgo en la estimación, la forma de trabajar de este método es remuestreando aleatoriamente el conjunto de datos un número b determinado de veces con reemplazo, esto genera b estimaciones de los parámetros de interés, a la vez se generan modelos de regresión para cada uno de los b conjuntos de datos, de tal forma que las submuestras como los datos originales son representativos de la población, una condición que debe de cumplirse es que las submuestras sean recopiladas uniformemente en todo el conjunto de datos. Además entre mayor número de submuestras b se tengan se

podrán manejar más eficientemente datos no lineales y no tan suaves, a expensas de aumentar el costo computacional, (Gosselin y Duchesne, 2010). Este proceso también es conocido como validación cruzada y fue útil para la selección de las componentes ya que tomar una muestra de la variable Y , esta se trata de estimar con los datos sobrantes y así n veces, de tal manera que cada componente va contribuyendo a explicar la varianza de la variable Y , este proceso se detiene en el momento en que ya no hay contribución de las componentes, evitando así el sobre ajuste de los datos (Gosselin y Duchesne, 2010).

En este trabajo se emplea el método de VIP para seleccionar bandas con las longitudes de onda que sean representativas para la estimación de la altura máxima del dosel y se trató de mejorar esta estimación utilizando el bootstrap, estos resultados se presentaran posteriormente.

Capítulo 4

Procesamiento de Datos y Resultados

A continuación se da información respecto al área de estudio, se presentan los resultados que se obtuvieron del procesamiento de los datos con la regresión PLS y aplicando el método de VIP para la selección de las bandas con mayor importancia, esto a partir de la elección de la selección de componentes

4.0.1. Fuente y Tratamiento de los Datos Empleados

Los datos LiDAR con los que se trabajó de origen fueron proporcionados por INEGI, el vuelo se realizó en los meses de noviembre y diciembre de 2007, se obtuvieron a una altura de vuelo 2000 a 3600 m, con una velocidad promedio de vuelo de 165 nudos y ángulo de captura de 25 y 45 grados, dentro del Sistema Geodésico de referencia 1980 (GRS 1980, en inglés Geodetic Reference System 1980), la densidad de la nube es de aproximadamente $0.5 \text{ pts}/\text{m}^2$; adicionalmente los datos con la información de las firmas espectrales se obtuvieron gracias al apoyo del proyecto CB-2011-01-169069 (Cárdenas, 2016), estos últimos contenían información de la localización geográfica de las huellas (Comission., 1991), así como la huella misma (*footprint*) que se generó debido a la medición realizada en un sobrevuelo en la zona de conservación de la Ciudad de México el 7 de Junio de 2013. Para más información sobre como se recolectaron los datos referirse a (Silván-Cárdenas y Corona-Romero, 2017). El radio de la huella se aumentó un 10% para absorber el movimiento del helicóptero en el vuelo, con lo que se obtuvo un buffer de la firma espectral de la huella con un 10% más de radio que el original, a continuación se realizó la intersección de las huellas (se localizaron 766) y la nube de puntos lidar (se tenían 27,382,651 puntos en la nube inicialmente),

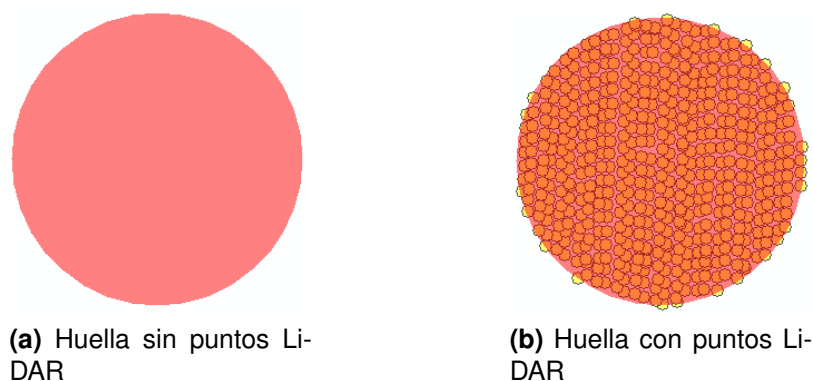


Figura 4.1 Ejemplo de las huellas.

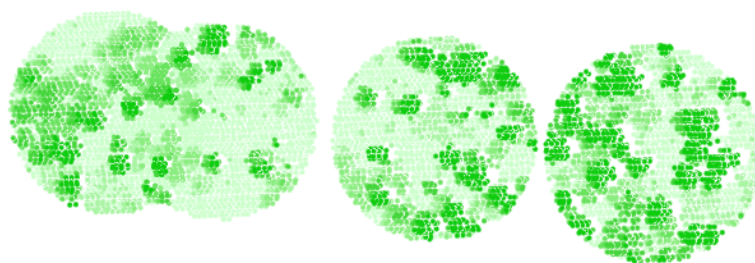


Figura 4.2 Huellas con puntos LiDAR en 2D.

ésta intersección se realizó con el propósito de determinar el número de huellas que contienen datos LiDAR, por lo que se obtuvieron finalmente 675 huellas con datos lidar de los 766 iniciales y se utilizaron 345,168 puntos LiDAR contenidos en las huellas antes mencionados, estos últimos datos son los que se analizaron y con los que se obtuvieron los resultados de ésta tesis, cabe mencionar que el procesamiento de los datos se realizó en ArcMap 10.2 [®] y el software R. En la figura (4.1), se puede apreciar un ejemplo de dos huellas, uno sin datos LiDAR los cuales se descartaron del análisis y otro con datos LiDAR, estas huellas son los que se utilizaron para el análisis posterior. También se puede apreciar en la figura (4.2) una vista en 2D de las huellas, estas mismos huellas se pueden apreciar en 3D en la figura (4.3), se puede percibir el perfil del dosel de los árboles de la zona.

De las 675 huellas seleccionados se conoce la firma espectral de la vegetación, adicionalmente se cuenta con la nube de puntos LiDAR con la elevación de terreno sustraída, de tal forma que representan las alturas del dosel en cada huella, estas alturas inicialmente por conveniencia se tomaron en percentiles de 10 %, teniendo como altura mínima 0 metros y altura máxima 46.52 metros. Sin embargo al realizar el análisis

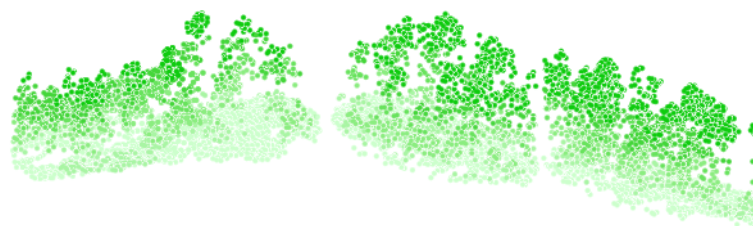


Figura 4.3 Huellas con puntos LiDAR en 3D.

sis este enfoque se descartó debido a que los resultados no mostraron los mejores valores de correlación esperada, aunque se puede mencionar que este proceso fue útil ya que se pudo determinar que solo el percentil 100 se puede explicar con una mayor precisión. Debido a lo anterior se optó por seleccionar de cada huella la altura máxima y así utilizar estos datos para alimentar al modelo de la regresión.

Para llegar a nuestro objetivo se utilizó la técnica de regresión mínimos cuadrados parciales (PLS por sus siglas en inglés) con el propósito de determinar si existe dicha relación entre altura y ciertas bandas características dentro del espectro electromagnético, las cuales se determinarán a partir del método VIP (variable importance in projection). Después se realizó una limpieza en la matriz de firmas espectrales, descartando algunos campos que no eran relevantes para el análisis, además de reordenarla para tener en las filas la información de cada huella y en las columnas las longitudes de onda que van de los 350 a los 2500 nanómetros, entonces se tiene una matriz de 675 filas por 2151 columnas, a esta matriz en adelante se le denominará matriz $X_{675,2151}$, la cual es la matriz de respuesta. Los datos LiDAR de igual manera se sometieron a una limpieza, eliminando columnas irrelevantes para el análisis, y seleccionando las huellas que contenían puntos de la nube LiDAR con la información de alturas máximas del dosel por huella, esto generó un vector de alturas con 675 filas, una por huella, a este vector se le denominará $Y_{675,1}$ o matriz de factores.

La figura (4.4) muestra un diagrama en el cual se ilustra un resumen de el proceso descrito anteriormente para el procesamiento de los datos en las 675 huellas, para su posterior análisis con el PLS y VIP, estos métodos se describirán más adelante.

A continuación a manera de ilustrar se presentan las imágenes (4.5) la cual corresponde a una porción del terreno con alturas de hasta 47.23 m y (4.6) con alturas de hasta 27.98 m, se aprecian los datos LiDAR generados con la ayuda del software FUSION:

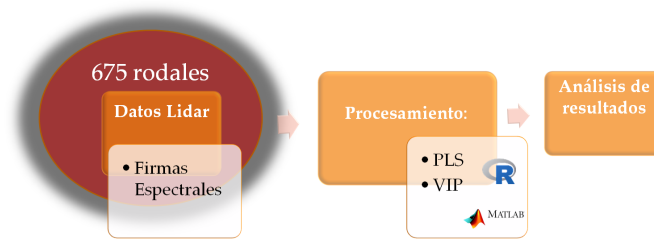


Figura 4.4 Diagrama del procesamiento de los datos.

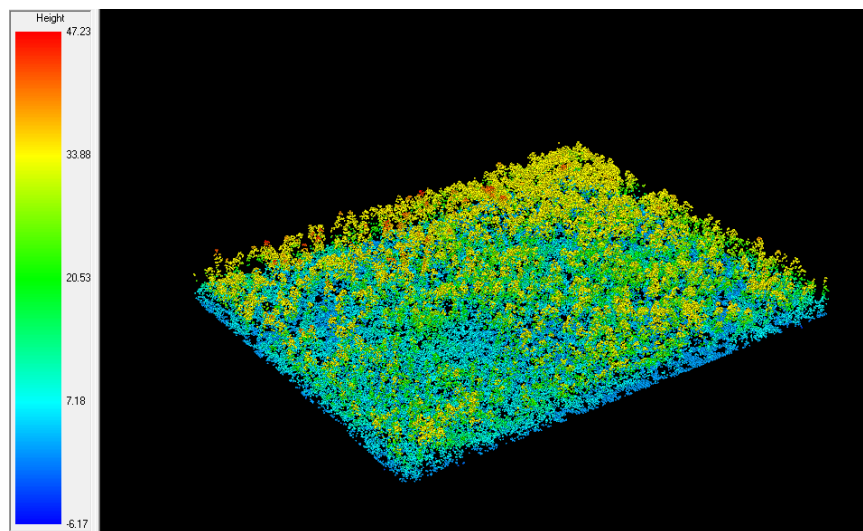


Figura 4.5 Imagen LiDAR normalizada de la zona de estudio

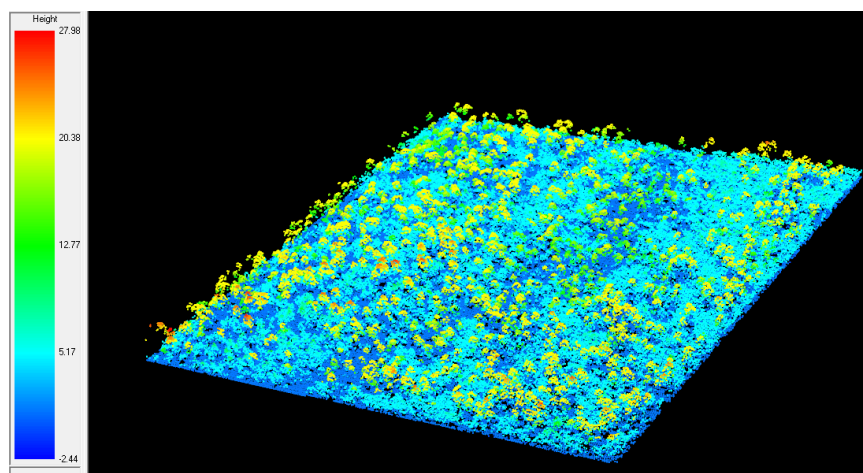


Figura 4.6 Imagen LiDAR normalizada de la zona de estudio

4.1. Construcción del Modelo

La regresión utilizando PLS se implementó a las matrices de frecuencias X (la cual consta de 675 filas correspondientes a las huellas y 2151 columnas que representan las longitudes de onda de la firma espectral estas van de los 350 nm a los 2500 nm), y al vector de altura máxima por huella Y (con 675 filas de igual manera una por cada huella y 1 columna correspondiente a la altura máxima), todo esto con el objetivo de determinar si existe alguna relación entre las alturas del dosel de los árboles y ciertas longitudes de onda dentro de la firma espectral de las huellas como ya se mencionó en el capítulo 3, y saber si la reflectancia de la luz que incide puede manifestar un comportamiento característico en el crecimiento de la vegetación. La figura (4.7), muestra un diagrama que plasma la secuencia de pasos para la construcción del modelo, donde inicialmente se procesan los datos con la regresión PLS utilizando un número de componentes inicial el cual es propuesto por el usuario, para el siguiente paso se grafica el RMSE y se obtiene de manera visual el número de componentes ($NCOM_{OBS}$) donde sea mínimo el error, posteriormente se recalcula la regresión PLS y se implementa el método VIP a la par de utilizar el método de Bootstrap para seleccionar un número de componentes óptimo ($NCOMP_{OPT}$), y así finalmente poder obtener la estimación de la variable Y y las bandas de longitudes de onda que sean importantes para realizar la estimación, omitiendo las que no resultaron importantes.

4.1.1. Selección de Componentes

Lo primero que se debe considerar es el número de componentes (variables latentes) que utiliza la regresión, esto se explicará a continuación. El número óptimo de componentes generalmente se obtiene mediante alguna técnica heurística la cual se basa en la variación residual (Tobias et al., 1995). Otra técnica que se sugiere es partir de un número de componentes y realizar algunas pruebas de tal manera que el error de la predicción sea mínimo (van der Voet, 1994), de esta última forma es como se procedió. En la figura (4.8) se ilustra la gráfica del RMSE; para este caso se tomaron inicialmente 30 componentes y se puede observar que el error decrece entre las 20 y 30 componentes, lo que indica que este es un número suficiente de componentes para continuar el proceso, ya que cabe mencionar que se realizaron otras pruebas aumentando las componentes hasta 500 pero no se observaron cambios significativos además de que el tiempo de procesamiento se incrementa sustancialmente, siendo así, se determinó que con este número de componentes es suficiente para poder pro-

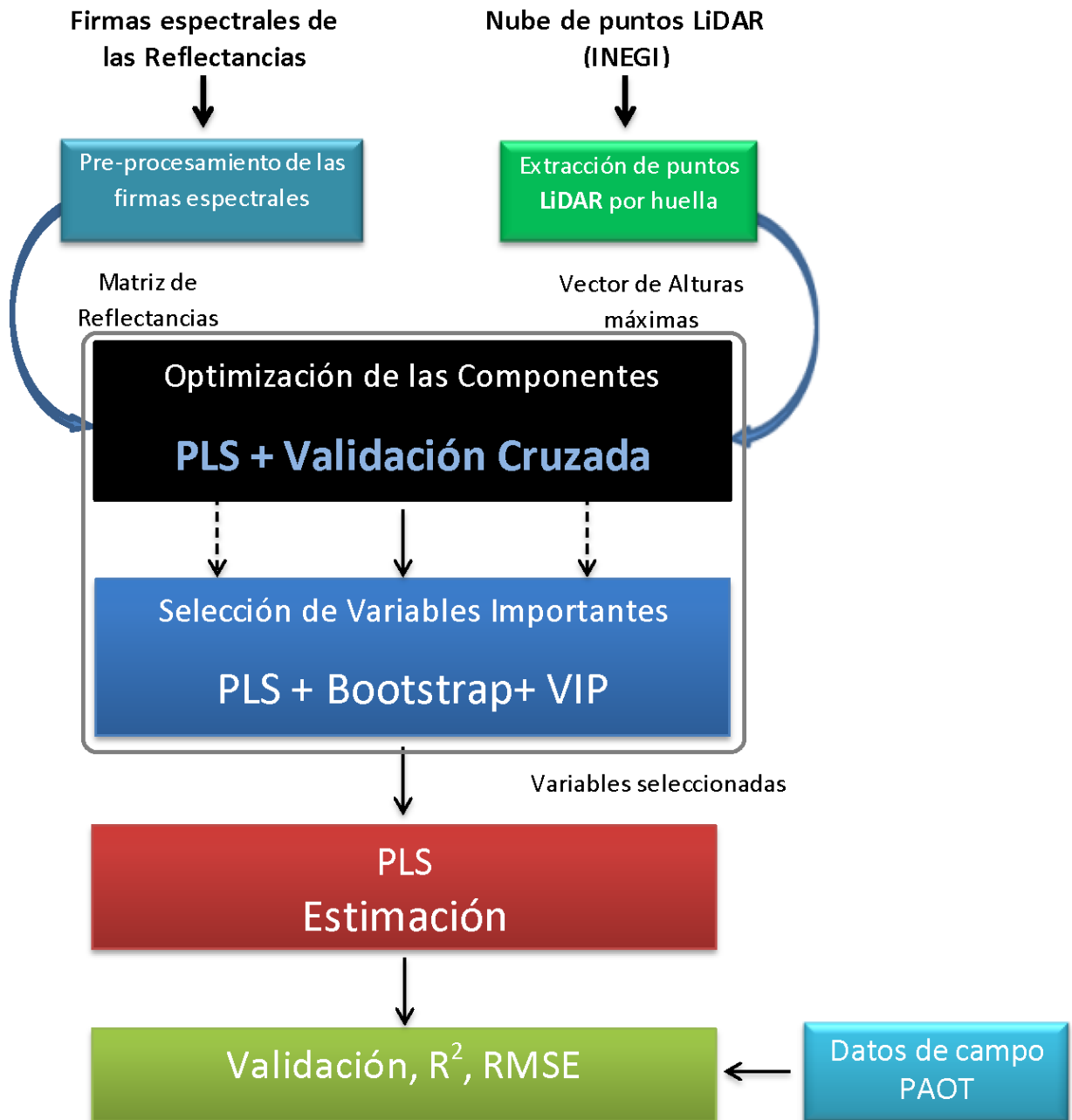


Figura 4.7 Diagrama del Proceso para la construcción del Modelo. Cabe destacar que al momento de obtener la Y-estimada se repitió la regresión PLS omitiendo las bandas donde se presenta adsorción de agua para mejorar la correlación.

seguir con el método del VIP.

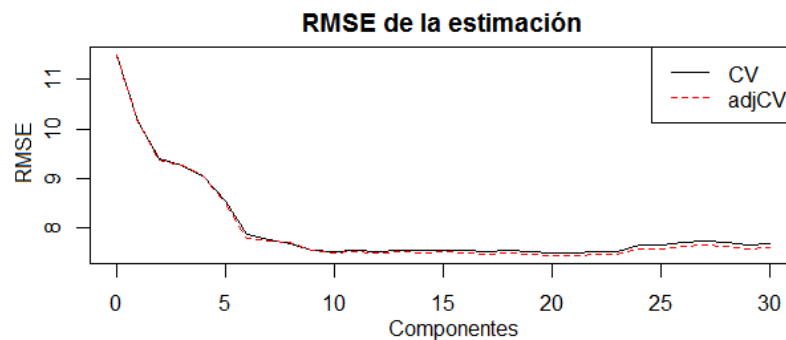


Figura 4.8 Selección de Componentes, para la variable Y

Una vez que se ha implementado la regresión PLS se obtiene una estimación de la variable Y para el vector de altura máxima Y la cual posteriormente solo denominaré Y -estimada, a continuación se muestra la gráfica correspondiente a la estimación de la variable Y obtenidos por la regresión, estos valores se encuentran entre los -4 y 47 metros, en caso de obtener valores negativos estos deberán descartarse, sin embargo más adelante se podrá observar que la estimación mejora.

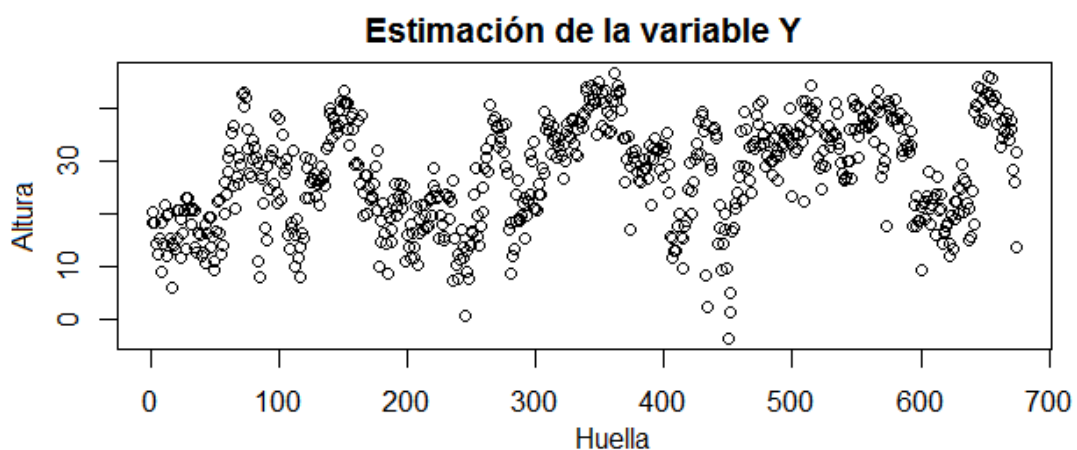


Figura 4.9 Estimación de la altura, para la variable Y

La regresión PLS adicionalmente calcula la matriz de variables latentes X scores que sirven para aproximar a la variable Y como ya se mencionó en el capítulo 3, esta tiene una dimensión de 675 filas (huellas) por 30 columnas correspondientes al número de componentes calculadas. Adicionalmente se observó que la varianza de cada columna es decir para cada componente es la misma para todas (4.10), este

valor fue de 0.00148368. Esto era hasta cierto punto esperado ya que la regresión utiliza vectores ortogonales para el cálculo de la matriz de Xscores.

Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6
0.00148368	0.00148368	0.00148368	0.00148368	0.00148368	0.00148368
Comp 7	Comp 8	Comp 9	Comp 10	Comp 11	Comp 12
0.00148368	0.00148368	0.00148368	0.00148368	0.00148368	0.00148368
Comp 13	Comp 14	Comp 15	Comp 16	Comp 17	Comp 18
0.00148368	0.00148368	0.00148368	0.00148368	0.00148368	0.00148368
Comp 19	Comp 20	Comp 21	Comp 22	Comp 23	Comp 24
0.00148368	0.00148368	0.00148368	0.00148368	0.00148368	0.00148368
Comp 25	Comp 26	Comp 27	Comp 28	Comp 29	Comp 30
0.00148368	0.00148368	0.00148368	0.00148368	0.00148368	0.00148368

Figura 4.10 Varianza de las componentes del 1 al 30

Con el fin de determinar el porcentaje de la varianza explicada se realizó una suma acumulada para realizar un diagrama de Pareto el cual se presenta en la figura (4.11), se observa que se tiene un acumulado de la varianza explicada del 71.32% para la variable Y, adicionalmente se ha observa que las alturas máximas, son las que presentan una relación más fuerte con ciertas bandas del infrarrojo, esta es una razón más de por que solo se trabaja con la altura máxima por huella.

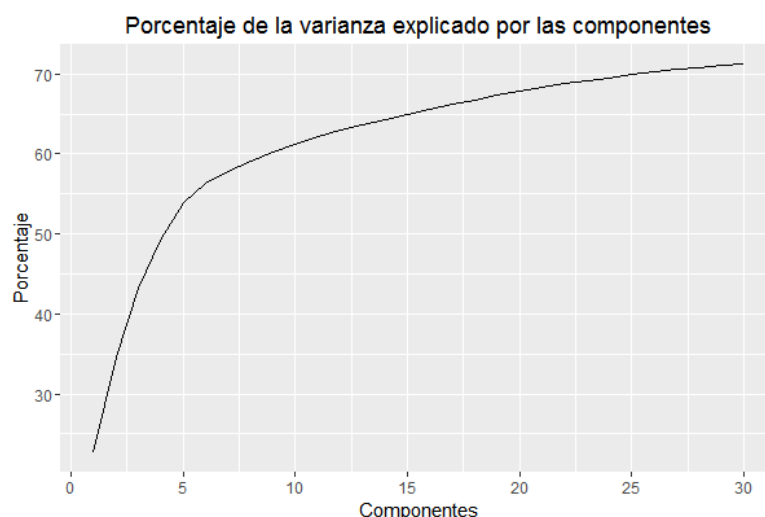


Figura 4.11 Varianza acumulada de las componentes del 1 al 30

4.1.2. Selección de las Variables

Posteriormente al implementar la regresión PLS y el método de selección de variables VIP se obtuvieron las longitudes de onda que son relevantes esto debido a que

tienen un valor de $VIP + STD > 1$, (en inglés Standard Deviation, Desviación Estándar (STD)), se puede apreciar en la gráfica (4.12) que todos los valores por encima de 1 serán las longitudes de onda elegidas, éstas son las que obtuvieron mayor importancia, el método VIP combinado con el bootstrap calculan el óptimo de componentes para el método, inicialmente se comenzó con 30 componentes como se menciona anteriormente y al final el bootstrap arroja un total de 22 componentes óptimas.

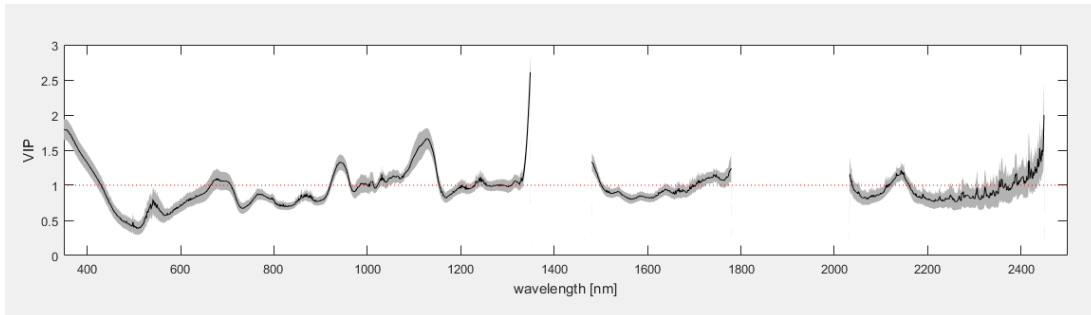


Figura 4.12 Índices del proceso de VIP

Adicionalmente se obtuvieron los coeficientes de la regresión los cuales sirven para estimar a la variable Y , estos se pueden apreciar en la figura (4.13), se puede observar un comportamiento en el cual cada cresta tiene su correspondiente valle a lo largo de todo el ancho de banda que se está ingresando.

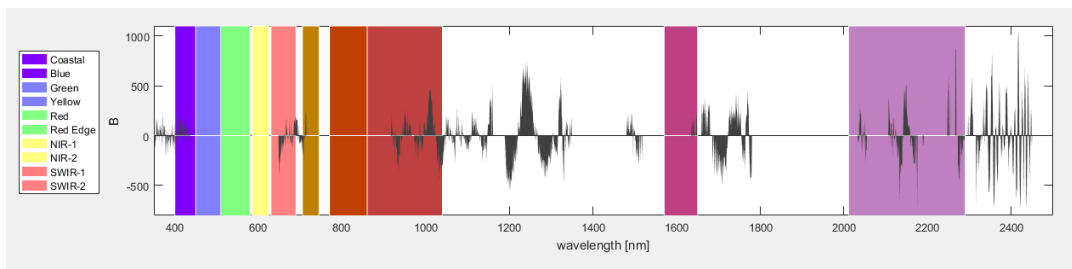


Figura 4.13 Contribución relativa

Para esta primera estimación se obtuvo un RMSE de 7.44 m; esto se puede apreciar en la figura (4.14). La gráfica de dispersión muestra la altura máxima medida con LiDAR y la altura estimada por el método PLS. Más adelante se trató de disminuir el error, posteriormente se presentaron resultados utilizando como base para la regresión PLS la Y -estimada.

Una vez que se obtuvo la Y -estimada, se realiza de nuevo la estimación con la regresión PLS con el objetivo de mejorar la estimación utilizando ahora la nueva matriz

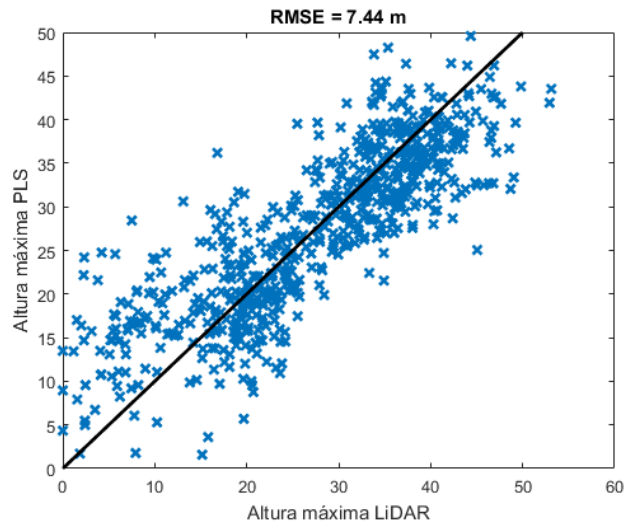


Figura 4.14 Diagrama de dispersion de altura LiDAR vs altura PLS.

X en la que se omiten las bandas que no resultaron importantes, es decir aquellas donde el valor de $VIP + VIP_{STD}$ es menor a 1, a esta se le denominará X_{vip} , en la figura (4.15) se muestra el resultado del método VIP y abajo los coeficientes de la regresión β , en apariencia los resultados son muy similares a la primera estimación, sin embargo en la figura (4.16), se aprecia que el valor mínimo es 1 m. y el valor máximo es 48, a diferencia de la primera estimación (4.9), donde aparecían valores negativos.

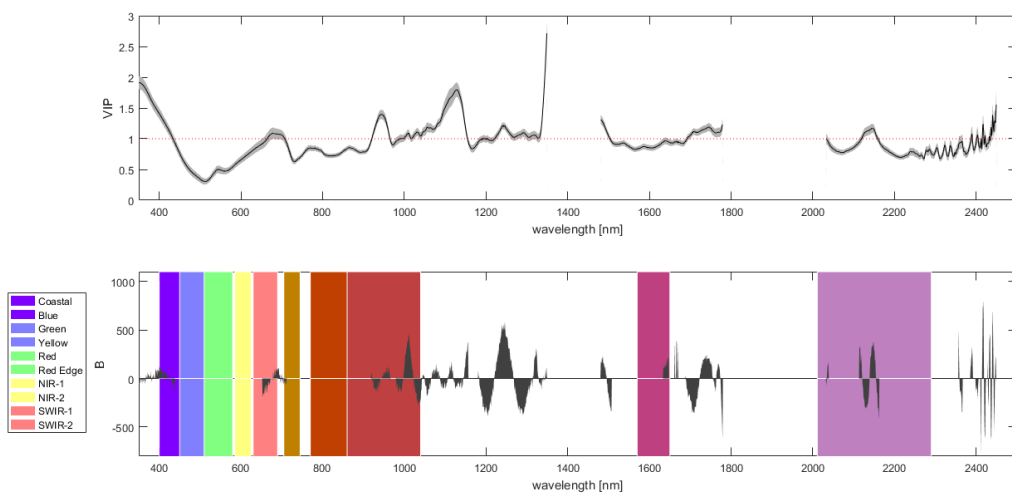


Figura 4.15 Resultados del VIP y Coeficientes β de la regresión.

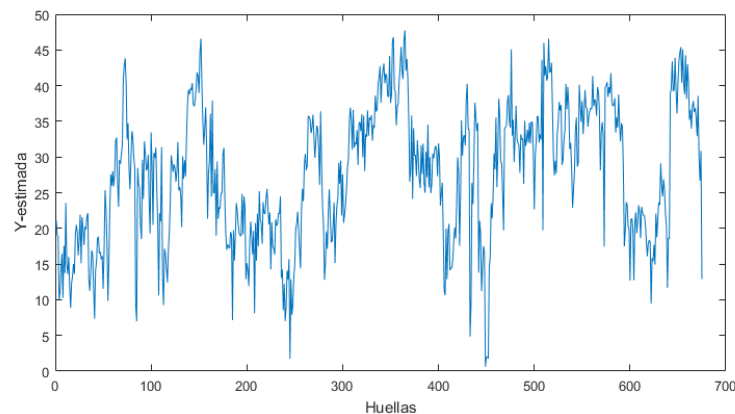


Figura 4.16 Gráfica de la *Y-estimada*.

Lo interesante de esta segunda estimación (con esta se trabajó posteriormente) utilizando la X_{vip} es que al obtener el RMSE este se redujo a 6.68 m; mejorando la estimación inicial. La gráfica de dispersión del RMSE de la *Y-estimada* mediante PLS y de la *Y* medida con LiDAR se muestra a continuación en la figura (4.17):

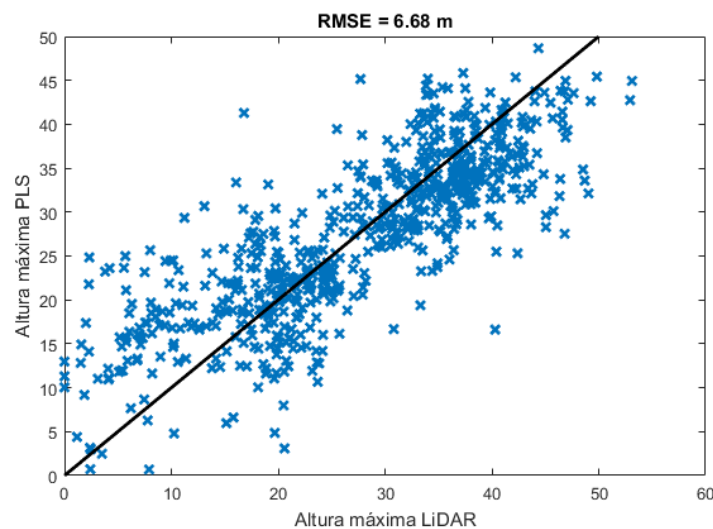


Figura 4.17 Gráfica del RMSE de la segunda corrida de la regresión

Al observar la primera imagen de la figura (4.15), se puede apreciar que la región de la luz visible no influye para nuestros resultados, sin embargo la región del infrarrojo si es relevante entre las bandas del 921 - 965 nm, 1019 - 1154 nm estas bandas están relacionadas con cantidades biofísicas tales como la altura de las plantas (ver tabla (3.1)), 1222 - 1349 nm esta banda se relaciona con biomasa (ver tabla (3.1)),

después evitando la región de absorción de vapor de agua, se tiene otra banda de importancia entre los 1481 - 1503 nm, 1694 - 1779 nm (estas están relacionadas con absorción de agua), de nuevo se evita la región de absorción de agua y se encuentra otra banda pequeña entre los 2120 - 2159 nm que también se relacionan con absorción de agua: humedad moderada a baja; y material de plantas muertas, finalmente entre los 2401 y 2449 nm, hay bandas muy pequeñas estas son del 2041 - 2404 nm, 2412 nm, 2416 - 2418 nm, 2421 - 2422 nm, 2432 - 2433 nm, 2436 - 2449 nm, estas bandas se relacionan con el estrés asociado a las plantas, sin embargo no se encontró mayor información sobre estas.

4.2. Validación del Modelo

Para validar el modelo óptimo se determinó la distancia entre la huella de los datos LiDAR y los sitios de los conglomerados (definidos por la Procuraduría Ambiental y del Ordenamiento Territorial (PAOT)), y se calculó la correlación entre la altura medida en cada sitio mediante trabajo de campo y la *Y-estimada* de la regresión PLS, la metodología que se implementó para seleccionar esta distancia es: al seleccionar una huella se determinó la distancia entre esta y los sitios (1, 2, 3 y 4), del conglomerado, tomando por conveniencia como rango distancias entre los 4.59 metros y los 40.29 metros, los que se encontraron fuera de este rango se descartaron ya que la correlación era cercana a cero, de tal manera que el sitio con la distancia mínima es elegido y se utiliza la altura máxima medida en campo de este sitio para correlacionarla con la estimación en la huella cercana con el fin de determinar el grado de aproximación del modelo a las observaciones en campo, este proceso se ejemplifica en la figura (4.18). Se debe destacar que lo importante para este estudio no es tanto el error entre la altura medida en campo y la estimación realizada, si no la correlación que hay entre estos dos. Se aprecian dos columnas, la primera columna de figuras es la correlación para la altura total contra la *Y-estimada*, la segunda columna es de igual manera altura total vs *Y-estimada* con la diferencia que se eliminaron las bandas de absorción de agua (1350 - 1480 nm, 1780 - 2032 nm y 2450 - 2500 nm), de la matriz de frecuencias para determinar si esto tenía un efecto en la estimación, se corrieron de nuevo los datos con el método del PLS y Bootstrap y se obtuvieron los resultados siguientes:

Posteriormente se obtuvo la correlación para los rangos de distancia de 4.59 - 40.29 m; 4.59 - 30.84 m; de 4.59 - 20.78 m; 4.59 - 10.06 m; 4.59 - 7.81 m, los resultados se puede apreciar en la figura (4.19), se observa que la mayor correlación 0.383 y

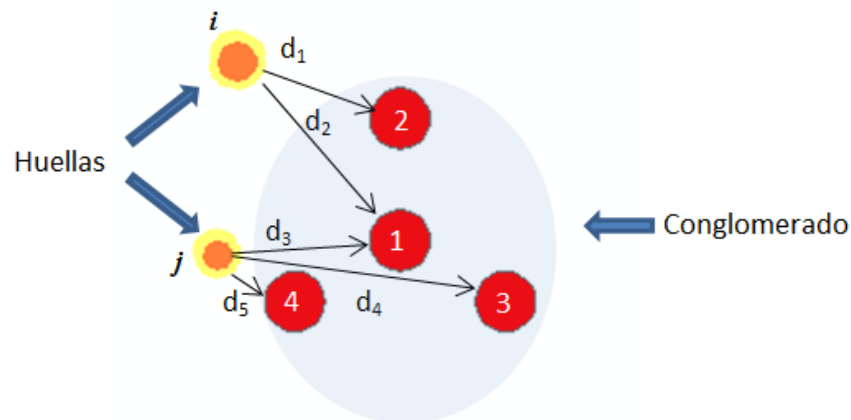


Figura 4.18 Huellas y conglomerado con los sitios de la PAOT, para el caso de la huella i , se seleccionó la distancia d_1 y para la huella j se eligió la distancia d_5 , de cada sitio seleccionado se eligió la altura máxima de los árboles la cual se correlacionó con la altura estimada por la regresión PLS.

0.4766 (para el caso en que se conservan las bandas de absorción de agua y el caso en que se eliminan, respectivamente), se encuentra a los 10.06 metros de distancia entre la huella y el sitio elegido, esto tiene sentido ya que al tener una mayor distancia entre estos la correlación tiende a disminuir y al ser menor a esta distancia se tienen pocos datos.

Es también importante mencionar que este tipo de metodología no explica con tanta precisión alturas menores a las del dosel, por lo que se descartaron inicialmente.

Los resultados se contrastan con los obtenidos en (González, 2013), donde se realizó estimación de alturas, se puede encontrar una diferencia principalmente en que se hace la estimación para altura promedio, en cambio en este trabajo se realizó la estimación para altura de dosel, otra diferencia importante es que para los resultados obtenidos en este estudio no importa la especie de árbol, (González, 2013) realizó la estimación por especie de árbol (Oyamel, Pino, Pastizal y Matorral, Bosque mixto e inducido), obteniendo distintas correlaciones entre cada especie, para fines prácticos en esta tesis se obtuvo una generalización para estudio de alturas independiente de la especie. Sin embargo para (González, 2013) su resultado principal es el de la estimación de biomasa a diferencia de este trabajo de tesis el cual tiene como objetivo principal determinar la estructura vertical del bosque.

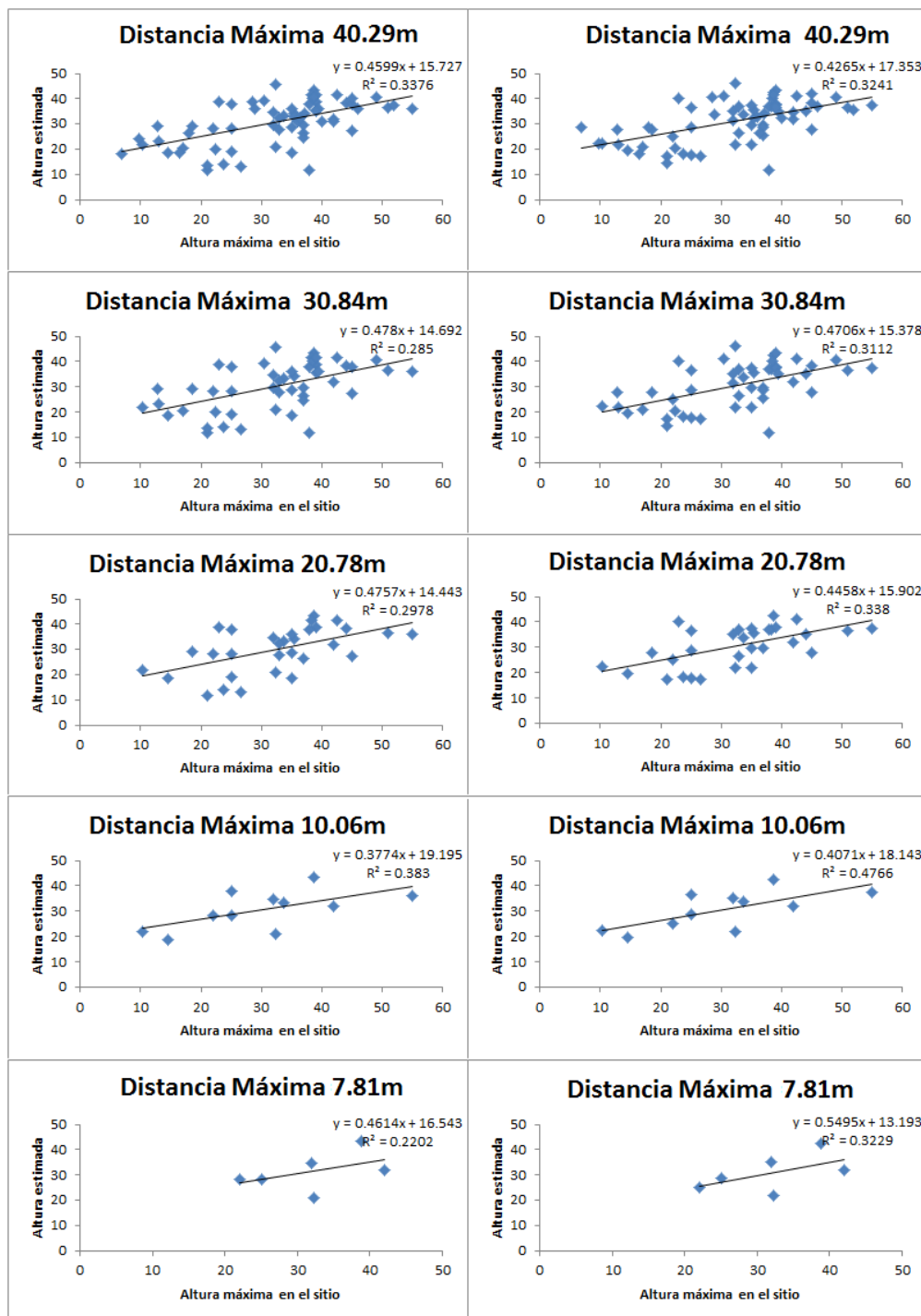


Figura 4.19 Correlaciones de la altura máxima en los sitios de la PAOT y la altura máxima estimada por la regresión PLS, se tiene la mayor correlación para la distancia de 10.06 m. con un valor de $R^2 = 0.383$ tomando en cuenta las bandas de absorción de agua y $R^2 = 0.4766$ una vez que se descartaron las bandas de absorción de agua.

Capítulo 5

Discusión y Conclusiones

Una vez que se obtuvieron los resultados y basándome en la hipótesis inicial la cual era: La luz que incide sobre la estructura de los árboles genera una respuesta denominada reflectancia, esta es el cociente entre la luz reflejada y la luz incidente, de tal manera que la reflectancia se puede utilizar para determinar la altura del bosque del suelo de conservación de la Ciudad de México.

Puedo concluir con base en los resultados obtenidos que si existe una correlación entre la altura máxima del dosel de los árboles medida en campo y la altura estimada mediante la regresión PLS de tal manera que existe una respuesta en la reflectancia de ciertas bandas dentro del espectro electromagnético las cuales sirvieron para explicar hasta un 71.32% (figura (4.11)) de esta altura máxima, demostrando así la hipótesis ya que se relacionó la estructura vertical del bosque con la reflectancia medida de la luz incidente sobre el dosel.

Un resultado relevante es que la región de la luz visible no genera una respuesta que sea importante para determinar la altura del dosel. Por otro lado las bandas dentro del NIR entre los 921 - 965 nm y 1019 - 1154, si generaron una respuesta para considerarlas importantes, estas a su vez están relacionadas con propiedades biofísicas tales como la altura de las plantas (ver tabla 3.1). También una región pequeña dentro del FSWIR en los 2120 -2159 nm, que esta relacionada con hojas de árboles caídas en el suelo, lignina, celulosa, diferenciación de hojas caídas en el suelo, además de relacionarse con estrés y contenido de hierro en el suelo.

La información que arroja este estudio se puede aprovechar para desarrollar un sensor con los rangos de longitud de onda específicos de tal manera que solo se midan las bandas que se consideraron importantes y así poder realizar estimaciones

de altura.

Con base en uno de los resultados obtenidos (ver figura (4.19)), se puede decir que la estimación de altura aumenta su correlación al tener a una distancia aproximada de 10.06 metros a algún sitio de un conglomerado proporcionado por la PAOT, siempre y cuando se hayan omitido las bandas de absorción de agua, obteniendo una $r^2 = 0.4766$ y un RMSE = 6.68 m.

En (Lin y Zhang, 2017) realizaron una estimación de altura para siete sitios los primeros dos eran bosques en Estados Unidos, uno localizado en Oregon (coníferas) y en Tennessee (bosque caducifolio), localizados en una zona con pendientes que van de 0% a 27% para Oregon y entre 1 y 24% para Tennessee, además se seleccionaron tres sitios en Brasil con temporadas de 10 meses sin lluvia, esto para tener un referente de zonas con poca lluvia y por último dos sitios el primero fue un bosque mixto de coníferas en el Bosque Nacional de Tahoe (California, EU, con pendientes que van de 0 a 60%), y el último sitio era un bosque caducifolio oriental en el Bosque Experimental Bartlett (New Hampshire, EU, con pendientes de 0 a 30%), reportando para el Amazonas una $r^2 = 0.83$ y RMSE = 4.3, para Bartlett $r^2 = 0.82$ y RMSE = 2.3, Oregon $r^2 = 0.90$ y RMSE = 5.9, Tahoe $r^2 = 0.75$ y RMSE = 5.5, Tennessee $r^2 = 0.40$ y RMSE = 4.8, todos los demás $r^2 = 0.83$ y RMSE = 4.9, la adquisición de los datos se realizó mediante el sistema de altimetría láser (GLAS, Geoscience Laser Altimeter System por sus siglas en inglés) montado en el satélite ICESat el cual tuvo un diámetro de las huellas de entre 52 y 90 metros.

Por otro lado (Kugler y Papathanassiou, 2014), utilizando datos adquiridos por los satélites TanDEM-X (lanzado en 2010) y TerraSAR-X (lanzado en 2007, ambos utilizan tecnología de interferometría de radar de apertura sintética polarimétrica Pol-InSAR), realizó una estimación de altura para tres tipos de bosques, boreal, templado y tropical, los tres con diferentes tipos de bosques y condiciones de terreno; el bosque boreal se localiza en Krycklan, Suecia, el templado en Traunstein, Alemania y el tropical Mawas, Indonesia. Los resultados obtenidos para el bosque boreal en condiciones de suficiente penetración el resultado para la correlación fue de $r^2 = 0.77$ se reporta que la adquisición de los datos se realizó en verano; para el templado se tuvo $r^2 = 0.77$ en invierno, para el caso del bosque tropical se reporta un $r^2 = 0.50$, se menciona que este resultado se puede deber a la poca penetración.

Las estimaciones realizadas con el TanDEM-X y TerraSAR-X tuvieron una huella de aproximadamente 3 ha, y para el sistema GLAS satélite ICESat el diámetro fue de entre 52 y 90 metros, en mi caso se tuvo diámetros de entre 10 y 70 metros, esto puede

influir en la correlación obtenida, además de que TAnDEM-X y TerraSAR-X tienen una resolución espacial de entre 12 y 30 metros, para el ICESat se tiene una resolución de 100 metros, en el caso de las imágenes hiperespectrales se tiene una resolución de 1 nm, esto puede explicar el valor de $r^2 = 0.47$ que obtuvimos ya que se porta mayor resolución que los sensores de radar.

Se concluye que utilizando la estimación se cuenta con un método que puede ser utilizado para cálculo de biomasa a partir de imágenes hiperespectrales, recordando según lo visto en el capítulo 2 que tradicionalmente se realiza esta estimación con ecuaciones alométricas, como trabajo a futuro se pueden realizar estimaciones de biomasa utilizando la estimación obtenida, esto se puede lograr utilizando ecuaciones de biomasa que solo relacionen la altura de los árboles. Sin embargo se tiene que mencionar que la estimación lleva un grado de error, debido a la temporalidad de los datos LiDAR recopilados en 2008 y de los datos multiespectrales medidos en 2013, una posible solución a esto sería trabajar idealmente con datos que sean tomados al mismo tiempo es decir LiDAR y datos medidos en campo, para así poder realizar una validación que arroje resultados de mayor precisión.

Hablando de servicios ecosistémicos se puede incrustar este trabajo en esta área para servir como una herramienta para implementar acciones a diferentes escalas, de tal manera que permita detener la pérdida de biodiversidad, además de beneficiar a diferentes actores de la sociedad, por un lado al preservar la zona de conservación de la Ciudad de México, se está garantizando la captación de carbono así como el suministro de oxígeno a la población de la zona metropolitana, por otro lado si se realiza el pago por servicios ambientales a los pobladores de la región estos se están beneficiando, generando un beneficio social y ambiental con el fin de maximizar el bienestar humano, con esto se cumple el fin de los servicios ecosistémicos el cual es vincular a la sociedad con la ecología y así poder alcanzar las metas de *Achie* (Rincón-Ruíz, 2014).

Una de las ventajas principales de utilizar una metodología como la que se propone en este trabajo es que al utilizar imágenes hiperespectrales se tiene una mayor cantidad de información a diferencia de solo utilizar datos LiDAR, los cuales solo proporcionan información de la posición, además de que se pueden analizar regiones mucho más grandes.

Otra opción para realizar la estimación de alturas es utilizar fotogrametría utilizando Remotely Piloted Aircraft System (RPAS) (comúnmente conocido como Drón), sin embargo el tiempo de vuelo de los RPAS es limitado lo que solo permite tener una

cobertura de algunas decenas o cientos de hectáreas; en el caso del suelo de conservación que se analizó se tiene un aproximado de 4,870 hectáreas además de que se necesita contar con habilidades especiales para realizar los vuelos, por esta razón es que utilizar este tipo de técnica por el momento no es viable, sin embargo en un futuro considero que será una de las principales técnicas, una vez que la autonomía de vuelo aumente considerablemente.

De acuerdo a lo mencionado anteriormente existe evidencia de que la hipótesis planteada inicialmente no es falsa, ya que se recopiló evidencia significativa que la prueba, se espera que para trabajos futuros se puedan incluir más variables para mejorar el RMSE.

Bibliografía

- Anderson, M., Neale, C., Li, F., Norman, J., Kustas, W., Jayanthi, H., and Chavez, J. (2004). Upscaling ground observations of vegetation water content, canopy height, and leaf area index during smex02 using aircraft and landsat imagery. *Remote sensing of environment*, 92(4):447–464.
- Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote sensing of Environment*, 64(3):234–253.
- Barnett, T.P., K. H. M. C. T. D. G. H. P. J. E. R. E. R. C. R. B. S. y. S. T. (1999). Detection and attribution of recent climate change: a status report. *Bull. Am. Met. Soc.*, . 80 (12):2631–2659.
- Bazzaz, F. J. C. and Morse, S. (1990). Growth responses of seven major cooccurring tree species of the northeastern united states to elevated CO_2 . *For. Res*, 20(9):1479–1484. Can.
- Broge, N. H. and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote sensing of environment*, 76(2):156–172.
- Brown, S., A. J. R. G. and Lugo., A. E. (1991). Biomass of tropical forest of south and southeast asia. *Canadian Journal of Forest Research*, 4:366–384.
- Brown, S., L. R. I. A. P. and Lugo, A. E. (1994). *Effects of land use change on atmospheric CO_2 concentration: Southeast Asia as a case of study.*, chapter 4. Land use and biomass changes of forest in Peninsular Malaysia during 1972-82: use of GIS analysis.
- Brown, S. (1997). Estimating biomass and biomass change of tropical forests: A primer. *FAO Forestry Paper*, 134.

- Brown, S. and Iverson, L. R. (1992). Biomass estimates for tropical forest. *World Resources Review*, 4:366 – 384.
- Brown, S. and Lugo, A. E. (1992). Aboveground biomass estimates for tropical moist forest of the Brazilian Amazon. *interciencia*, 17:8–18.
- Brown, S. A. G. and Lugo, A. (1989). Biomass estimation methods for tropical forest with applications to forest inventory data. *Forensic Science*, 35:881–902.
- Cárdenas, J. L. S. (2016). Análisis hiperespectral de la dispersión de la luz en doseles y su relación con la biomasa. *REPORTE TÉCNICO. Proyecto CB-2011-01-169069. SEP-CONACYT. Para consulta de los datos ponerse en contacto al e-mail: jsilvan@centrogeo.edu.mx.*
- Carter, G. A., Knapp, A. K., Anderson, J. E., Hoch, G. A., and Smith, M. D. (2005). Indicators of plant species richness in AVIRIS spectra of a mesic grassland. *Remote sensing of Environment*, 98(2):304–316.
- Chong, I.-G. and Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1):103–112.
- Comission., F. (1991). Forestry practice.
- Daughtry, C., Walthall, C., Kim, M., De Colstoun, E. B., and McMurtrey, J. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment*, 74(2):229–239.
- Dulce Ma. Avendaño Hernández, Miguel Acosta Mireles, F. C. A. and Barra, J. D. E. (2009). Estimación de biomasa y carbono en un bosque de abies religiosa. *Fitotec.*, 32(3):233–238.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., and Wold, S. (2001). Multi-and megavariate data analysis, principles and applications. *ometrics academy, Umeå, Sweden.* Technical report, ISBN 91-973730-1-X.
- Gayon, J. (2000). History of the concept of allometry. *American Zoologist*, 40:748–758.
- González, J. A. A. (2013). Estimación de almacenamiento de carbono en el suelo de conservación del Distrito Federal mediante el uso de datos Lidar aerotransportado. Master's thesis, CentroGeo.

- Gosselin, Ryan, R. D. and Duchesne, C. (2010). A bootstrap-vip approach for selecting wavelength intervals in spectral imaging applications. *Chemometrics and Intelligent Laboratory Systems*, 100(1):12–21.
- Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications*, 10(6):1861–1870.
- Hyypä J., Hyypä, H. L. D. G. F. Y. X. y. M. 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.
- Jones, R.G., J. M. M. N. . B. K. (1997). Simulation of climate change over europe using a nested regional-climate model. ii: Comparison of driving and regional model responses to a doubling of carbon dioxide. *Q.J.R. Meteorol.*, 123:265–292.
- Kugler, Florian, S. D. H. I. P. H. and Papathanassiou, K. P. (2014). Tandem-x pol-insar performance for forest height estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6404–6422.
- Labraga, J. (1997). The climate change in south america due to a doubling in the co2 concentration: intercomparison of general circulation model equilibrium experiments. *Int. J. Climatol.*, 17:377–398.
- Lin, Dongfang, Z. J. F. H. X. Q. and Zhang, B. (2017). A tsvd-based method for forest height inversion from single-baseline polinsar data. *Applied Sciences*, 7(5):435.
- Lohani, B. (2008). Airborne altimetric lidar: Principle, data collection, processing and applications. URL: http://home.iitk.ac.in/~blohani/LiDAR_Tutorial/Airborne_AltimetricLidar_Tutorial.htm, accessed Feb.
- Mackey, L. W. (2009). Deflation methods for sparse pca. In *Advances in neural information processing systems*, pages 1017–1024.
- Mann, M.E., R. S. B. and Hughes., M. K. (1999). Northern hemisphere temperatures during the past millennium; inferences, uncertainties, and limitations. *Geophys. Res. Lett.*, 26 (6):759–762.
- Manne, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):187–197.

- Martinez Yrizar, A. J. Sarukhan, A. P. J. E. R. J. M. M. A. S. M. and Cervantes., L. (1992). Aboveground phynomass of a tropical deciduous forest on the coast of jalisco, mexico. *Journal of tropical Ecology.*, 8:87–96.
- Mehmood, T., Liland, K. H., Snipen, L., and Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118:62–69.
- Miller, G. (1992). Living in the environment, an introduction to environmental science. *Wadsworth Publishing Co.*
- Mittelbach, G. G., Steiner, C. F., Scheiner, S. M., Gross, K. L., Reynolds, H. L., Waide, R. B., Willig, M. R., Dodson, S. I., and Gough, L. (2001). What is the observed relationship between species richness and productivity? *Ecology*, 82(9):2381–2396.
- Mutanga, O., Skidmore, A. K., and van Wieren, S. (2003). Discriminating tropical grass (cenchrus ciliaris) canopies grown under different nitrogen treatments using spectroradiometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(4):263–272.
- Niklas, K. L. (1994). *Plant Allometry: The scaling of form and process*. The University of Chicago Press.
- Oindo, B., Skidmore, A., and De Salvo, P. (2003). Mapping habitat and biological diversity in the maasai mara ecosystem. *International journal of remote sensing*, 24(5):1053–1069.
- Oindo, B. O. and Skidmore, A. K. (2002). Interannual variability of ndvi and species richness in kenya. *International journal of remote sensing*, 23(2):285–298.
- Organization, W. M. (2016). The state of greenhouse gases in the atmosphere based on global observations through 2015. Technical Report 12, Global Atmosphere Wath.
- Ortenberg, F., Thenkabail, P., Lyon, J., and Huete, A. (2011). *Hyperspectral sensor characteristics: airborne, spaceborne, hand-held, and truck-mounted; Integration of hyperspectral data with Lidar*. Boca Raton, FL, USA: CRC Press.
- Picard, N., Saint-André, L., and Henry, M. (2012). Manual de construcción de ecuaciones alométricas para estimar el volumen y la biomasa de los árboles: del trabajo de campo a la predicción.

- Pizaña, José Mauricio Galeana, H. J. M. N. and Romero, N. C. (2016). Remote sensing-based biomass estimation. In *Environmental Applications of Remote Sensing*. InTech.
- Psomas, A, K. M. H. S. I. K. and Zimmermann, N. (2011). Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing*, 32(24):9007–9031.
- Reyes, Gisel; Brown, S. C. J. L. A. E. (1992). Wood densities of tropical tree species. *USDA*, 88.
- Rincón-Ruiz, A., E.-D. M. P. A. M. T. C. H. D. A. A.-A. P. y. Z. P. A. (2014). Valoración integral de la biodiversidad y los servicios ecosistémicos: Aspectos conceptuales y metodológicos. *Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH)*, page 151.
- Ruiz, A. R. (2014). Valoración integral de la biodiversidad y los servicios ecosistémicos.
- Shan, J., y. T.-C. K. (2009a). Topographic laser ranging and scanning: principles and processing. *Taylor & Francis Group*.
- Shan, J., y. T.-C. K. (2009b). Topographic laser ranging and scanning: principles and processing. *Taylor & Francis Group*, page 590p.
- Silván-Cárdenas, J. and Corona-Romero, N. (2017). Radiation budget of vegetation canopies with reflective surface: A generalization using the markovian approach. *Remote Sensing of Environment*, 189:118–131.
- S.J., G. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews*, 41:587–640.
- SMA (2007). Agenda ambiental de la ciudad de México. *Programa de Medio Ambiente*, 2012:2007–2012.
- Svante Wold, Michael Sjöström, L. E. (2001). PLS-regression: a basic tool of chemometrics. *ELSEVIER. Chemometrics and Intelligent Laboratory Systems*, 58:109–130.
- Thenkabail, Prasad S, L. J. G. and Huete, A. (2016). *Hyperspectral remote sensing of vegetation*. CRC Press.
- Thenkabail, P. S., Enclona, E. A., Ashton, M. S., and Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote sensing of environment*, 91(3):354–376.

- Thenkabail, P. S., Smith, R. B., and De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote sensing of Environment*, 71(2):158–182.
- Tobias, R. D. et al. (1995). An introduction to partial least squares regression. In *Proceedings of the twentieth annual SAS users group international conference*, pages 1250–1257. Citeseer.
- van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomization test. *Chemometrics and Intelligent Laboratory Systems*, 25(2):313–323.
- Vitousek-P.M. (1994). Beyond global warming: ecology and global change. *Ecology*, 75: (7):1861–1876.
- Wagner, W., Ullrich, A., Melzer, T., Briese, C., and Kraus, K. (2004). From single-pulse to full-waveform airborne laser scanners: potential and practical challenges.
- Watson R.T., H. Rodhe, H. O. U. S. (1991). Green house gases and aerosols. in: Houghton j.t., g.j. jenkins, j.j. ephraums (eds.). *Climate change, the IPCC scientific assessment*, Cambridge University Press, Cambridge, pages 1–40.
- Wehr, A. and Lohr, U. (1999). Airborne laser scanning an introduction and overview. *ISPRS Journal of photogrammetry and remote sensing*, 54(2):68–82.
- Zarco-Tejada, P. J., Berjón, A., López-Lozano, R., Miller, J. R., Martín, P., Cachorro, V., González, M., and De Frutos, A. (2005). Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment*, 99(3):271–287.
- Zarco-Tejada, P. J., Rueda, C., and Ustin, S. (2003). Water content estimation in vegetation with modis reflectance data and model inversion methods. *Remote Sensing of Environment*, 85(1):109–124.

Glosario

AVHRR en inglés Advanced very high resolution radiometer, radiómetro avanzado de muy alta resolución. 31

CCD en inglés charge-coupled device. 30

DAP diámetro a la altura del pecho. 5

ESWIR Early short-wave infrared, Infrarrojo de onda corta temprana. 38

EUL Eficiencia del uso de luz. 37

FNIR Far Near Infrared, infrarrojo muy cercano. 38

FOV field of view. 31, 35

FSWIR Bandas Far short-wave infrared, Infrarrojo de onda corta lejana. 39, 62

HS Hiperespectrales. 15, 18, 26, 27, 31–36

HSS Sensores Hiperespectrales. 25, 31, 33

IAF Índice de área foliar. 17, 37, 38

IFOV en inglés Instantaneous field of view, Campo de visión instantáneo. 31

IHS Imágenes Hiperespectrales. 31

INS en inglés Inertial Navigation System, Sistema de Navegación Inercial. 34

IR Infrarrojo. 20, 31

ISMER en inglés medium resolution imagin spectrometer, Imágenes espectrométricas de mediana resolución. 31

Landsat Land remote sensing satellite program, Programa satélite de teledetección. 32

MS Multispectral. 27, 33

NIR en inglés Near Infrared, Infrarrojo cercano. 16, 20, 37, 38, 62

OTB Órbita Terrestre Baja. 26, 31

PAOT Procuraduría Ambiental y del Ordenamiento Territorial. 59, 63

PLS en inglés Partial Least Squares, Mínimos Cuadrados Parciales. 18, 41–43, 46, 52, 54–56, 58–60, 62

PLSR en inglés PLS regression, regresión PLS. 43

RADAR en inglés RAdio detection and ranging. 10

RMSE en inglés root mean square error, error cuadrático medio. 17, 52, 56, 58, 65

RPAS Remotely Piloted Aircraft System. 64

SE Servicios Ecosistémicos. 2

SeaWiFS en inglés sea-viewing wide field-of-view sensor, Sensor de campo de visión de gran amplitud. 32

SIEOR Sistema de Imágenes espectrográficas óptico relectivas. 31

STD en inglés Standard Deviation, Desviación Estandar. 56

VC Validación Cruzada. 44, 45

VENuS Vegetation an Environment New Micro Satellite. 36

VIP Variable Importance in Pojection. 18, 45, 46, 54–57