

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE CIENCIAS FORESTALES
SUBDIRECCIÓN DE POSTGRADO



MODELACIÓN GEOESPACIAL DE VARIABLES DE DENSIDAD FORESTAL

TESIS

QUE COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS

CON ESPECIALIDAD EN MANEJO DE RECURSOS NATURALES

PRESENTA

M.C. CARLOS ARTURO AGUIRRE SALADO

LINARES, N.L.

ENERO, 2013

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FORESTALES

SUBDIRECCIÓN DE POSTGRADO

MODELACIÓN GEOESPACIAL DE VARIABLES DE DENSIDAD FORESTAL

TESIS

QUE COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS
CON ESPECIALIDAD EN MANEJO DE RECURSOS NATURALES

PRESENTA

M.C. CARLOS ARTURO AGUIRRE SALADO

COMITÉ DE TESIS



DR. EDUARDO JAVIER TREVIÑO GARZA
DIRECTOR



DR. OSCAR ALBERTO AGUIRRE CALDERÓN
ASESOR



DR. JAVIER JIMÉNEZ PÉREZ
ASESOR



DR. MARCO AURELIO GONZÁLEZ TAGLE
ASESOR



DR. JOSÉ RENÉ VALDEZ LAZALDE
DIRECTOR EXTERNO

Manifiesto que la presente investigación es original y fue desarrollada para obtener el grado de Doctor en Ciencias con Especialidad en Manejo de Recursos Naturales. Donde se utiliza información de otros autores se otorgan los créditos correspondientes.

Carlos Arturo Aguirre Salado
Enero de 2013

CONTENIDO

	Página
ÍNDICE DE FIGURAS	v
ÍNDICE DE TABLAS.....	viii
RESUMEN	ix
SUMMARY.....	x
CAPÍTULO 1. INTRODUCCIÓN	1
1.1. Presentación.....	1
1.2 Hipótesis.....	3
1.3. Objetivos.....	4
1.3.1. Objetivo General.....	4
1.3.2. Objetivos Particulares.....	4
1.4. Antecedentes.....	5
1.4.1. Evaluación de los Bosques del Mundo.....	5
1.4.2. Inventarios Nacionales Forestales.....	6
1.4.2.1. Finlandia.....	6
1.4.2.2. Suecia.....	7
1.4.2.3. Alemania.....	9
1.4.2.4. Canadá.....	10
1.4.2.5. Estados Unidos.....	11
1.4.2.6. México.....	13
1.4.3. Procesamiento de Imágenes de Satélite en el Estudio de Recursos Forestales.....	15
1.4.3.1. Clasificación de los Métodos de Interpretación de Imágenes.....	15
1.4.3.2. Metodología General del Procesamiento de Imágenes.....	16
1.4.3.2.1. Muestreo en Campo.....	16
1.4.3.2.2. Preprocesamiento de la Imagen de Satélite.....	16
1.4.3.2.3. Relación Datos de Campo vs Datos	

	Espectrales.....	19
1.4.3.2.4.	Estimación de las Variables de Interés o Clasificación de la Imagen..	20
1.4.4.2.5.	Validación del Mapa Obtenido.....	23
1.5.	Literatura Citada.....	25

CAPÍTULO 2. CONSTRUCCIÓN DE MODELOS DE BIOMASA SUPERFICIAL CON TECNOLOGÍA DE PERCEPCIÓN REMOTA EN LA ZONA INTERTROPICAL EN MÉXICO

	Resumen.....	31
	Abstract.....	32
2.1.	Introducción.....	33
2.2.	Materiales y Métodos.....	35
2.2.1.	Área de Estudio.....	35
2.2.2.	Datos de Campo.....	36
2.2.3.	Datos Espectrales.....	37
2.2.4.	Procesamiento de la Información.....	39
2.2.5.	Validación.....	39
2.3.	Resultados y Discusión.....	40
2.3.1.	Relación entre las Variables de Densidad Forestal y los Indices de Vegetación.....	40
2.3.2.	Modelos de Regresión.....	41
2.3.3.	Mapas de Densidad.....	43
2.3.4.	Validación.....	45
2.4.	Conclusiones.....	46
2.5.	Literatura Citada.....	47

CAPÍTULO 3. MAPEO DE LA COBERTURA FORESTAL EN UNA PORCIÓN NORTE CENTRO DE MÉXICO: UNA COMPARACIÓN DE MÉTODOS DE PROCESAMIENTO DIGITAL DE IMÁGENES

Resumen.....	52
Abstract.....	53
3.1. Introducción.....	54
3.2 Materiales y Métodos.....	56
3.2.1. Área de Estudio.....	56
3.2.2. Materiales.....	57
3.2.3. Datos Espectrales.....	58
3.2.4. Procesamiento de la Información.....	59
3.2.5. Validación.....	61
3.3. Resultados y Discusión.....	62
3.3.1. Selección de Endmembers.....	63
3.3.2. Selección del Método.....	63
3.3.3. Mapas de Cobertura Fraccional.....	65
3.3.4. Distribución Espacial del Error.....	67
3.4. Conclusiones.....	68
3.5. Literatura Citada.....	69

CAPÍTULO 4. COMPARACIÓN DE ALGORITMOS NO PARAMÉTRICOS DEL VECINO MÁS CERCANO PARA LA ESTIMACIÓN DE BIOMASA FORESTAL

Resumen.....	73
Abstract.....	74
4.1. Introducción.....	75
4.2 Materiales y Métodos.....	78
4.2.1. Área de Estudio.....	78
4.2.2. Estimación de la Biomasa con Datos de Campo.....	78
4.2.3. Variables Predictoras.....	79
4.2.4. Procesamiento de la Información.....	83
4.3. Resultados y Discusión.....	85

4.3.1.	Selección de Variables.....	85
4.3.2.	Comparación de Métodos.....	88
4.3.3.	Estimación de la Biomasa.....	91
4.4.	Conclusiones.....	95
4.5.	Literatura Citada.....	96
CAPÍTULO 5. CONCLUSIONES FINALES		103

LISTA DE FIGURAS

	Página
Capítulo 1	
Figura 1.1. El cluster y la parcela empleada en el IFN sueco.....	8
Figura 1.2. Estimación de volumen maderable resultado de los Inventarios Nacionales Forestales apoyados en datos espectrales.....	9
Figura 1.3. El cluster y la parcela empleada en el IFN alemán.....	10
Figura 1.4. Esquema de una parcela en campo del FIA estadounidense	12
Figura 1.5. Esquema de las unidades de muestreo del Inventario Nacional Forestal de México.....	14
Figura 1.6. Marco conceptual de los métodos de clasificación de imágenes.....	15
Figura 1.7. Comportamiento teórico de la reflectividad vigorosa.....	19
Capítulo 2	
Figura 2.1. Localización del área de estudio. A) Contexto nacional y zona intertropical (hachurado) (Límite 23° 27' N) y B) Zonas geográficas del Estado.....	36
Figura 2.2. Distribución espacial de los sitios de muestreo del Inventario Forestal Nacional para el Estado de San Luis Potosí y transecto para comparación de estimaciones.....	37
Figura 2.3. Distribución de la biomasa arbórea aérea respecto a los índices de vegetación construidos y los VCF.....	41
Figura 2.4. Biomasa en función del NDVI62 y los VCF.....	43
Figura 2.5. Biomasa arbórea aérea estimada mediante los modelos seleccionados y las variables NDVI62 y VCF. A) Lineal, B) Exponencial.....	44
Figura 2.6. Comparación de estimaciones (píxel) en un transecto dirección oeste-este.....	44

Figura 2.7.	Biomasa arbórea aérea píxel a píxel para bosque mesófilo de montaña. A) Modelo Lineal. B) Modelo no lineal.....	45
Figura 2.8.	Valores de biomasa arbórea aérea observados vs predichos.....	46

Capítulo 3

Figura 3.1.	Localización del área de estudio. A) Contexto nacional y zona intertropical (hachurado) (Límite 23° 27' N) y B) Mapa topográfico.....	57
Figura 3.2.	Distribución espacial de los sitios de muestreo seleccionados del Inventario Forestal Nacional en el área de estudio sobre imagen en falso color compuesta con los tres primeros componentes principales.....	59
Figura 3.3.	Firmas espectrales puras basadas en datos MODIS MCD43A4. IRC: Infrarrojo Cercano, IRM: Infrarrojo Medio.....	62
Figura 3.4.	Figura 3.4. Izq. Localización de la zona de validación y Der. Clasificación con el algoritmo de máxima verosimilitud. Coordenadas UTM-14N (m) en Azul, Coordenadas Lat-Long en Negro.....	64
Figura 3.5.	Cobertura fraccional en el área de validación. A) Bosque y B) Matorral.....	64
Figura 3.6.	Cobertura fraccional estimada vs datos de referencia. Izq. Bosque. Der. Matorral.....	65
Figura 3.7.	Mapas de cobertura fraccional con el método de análisis de mezcla espectral y escalamiento de la cobertura fraccional con el mapa de biomasa arbórea aérea. A) y C) Bosque, B) y D) Matorral.....	67
Figura 3.8.	Distribución espacial del error para LMA.....	68

Capítulo 4

Figura 4.1.	Localización del área de estudio: A) Contexto nacional y B) Mapa topográfico y zona intertropical (hachurado).....	81
Figura 4.2.	Distribución espacial de los sitios de muestreo seleccionados del Inventario Forestal Nacional en el área de estudio.....	82
Figura 4.3.	Selección de variables con base en la validación cruzada y su efecto en el RMSD para el algoritmo kNN basado en la distancia euclideana sin estandarizar. A) variables espectrales, B) variables climáticas, C) variables topográficas y D) todas las variables.....	87
Figura 4.4.	Comparación de métodos no paramétricos entrenados con las variables satelitales: NDVI62 y VCF. A) RMSD absoluto y B) RMSD relativo.....	90
Figura 4.5.	Desempeño de los diferentes métodos no paramétricos: A) Raw, B) Euclideana, C) Mahalanobis y D) MSN.....	91
Figura 4.6.	Mapa de biomasa arbórea aérea obtenido con el algoritmo kNN empleando la distancia MSN (RMSD=30.56 Mg/ha, k=14, r = 0.90): San Luis Potosí, México.....	94

LISTA DE TABLAS

Página

Capítulo 2

Tabla 2.1.	Coefficientes de correlación de Pearson (r) obtenidos para biomasa y las variables satelitales.....	41
Tabla 2.2.	Modelos probados.....	42

Capítulo 4

Tabla 4.1.	Estimación de la matriz de ponderación para la estandarización de las distancias espectrales (Crookston y Finley, 2008).....	83
Tabla 4.2.	Estimaciones promedio de biomasa para cada tipo de vegetación obtenidas con el kNN (distancia MSN) (RMSD=30.56 Mg/ha, k=14, r = 0.90).....	94

RESUMEN

Carlos Arturo Aguirre Salado

Fecha de graduación: Enero de 2013

**Universidad Autónoma de Nuevo León
Facultad de Ciencias Forestales**

Título del estudio: Modelación geoespacial de variables de densidad forestal

Número de páginas: 105

**Candidato a Doctor en Ciencias con
Especialidad en Manejo de Recursos Naturales**

Área de estudio: Geomática Aplicada al Manejo de Recursos Naturales

Propósitos y Métodos de Estudio: Se presenta un marco metodológico para la modelación geoespacial de la densidad forestal mediante datos espectrales de resolución moderada para una porción del límite septentrional de la zona intertropical: San Luis Potosí, México. Se emplearon algoritmos de modelación estadística paramétrica y no paramétrica, análisis espacial y procesamiento de imágenes de satélite.

Contribución y Conclusiones: El Capítulo 1 presenta un estado general de las metodologías en otros países de referencia como Finlandia, Suecia, Alemania, Canadá, Estados Unidos y México e introduce a temas relacionados con inventarios forestales realizados con herramientas de análisis geoespacial. El Capítulo 2 investiga los índices espectrales de vegetación en la modelación de la densidad forestal concluyendo que los índices normalizados sensibles al contenido de humedad con una tendencia no lineal, modelan mejor la biomasa arbórea aérea. El Capítulo 3 explora métodos estadísticos para la estimación de cobertura fraccional a nivel subpixel y concluye que el más apropiado es el análisis mezcla espectral lineal para dos clases puras: bosque y matorral. En el Capítulo 4 se comparan variantes del algoritmo no paramétrico del vecino más cercano que relacionan la densidad forestal con variables espectrales y auxiliares, con la flexibilidad propia de retener la estructura de los datos de referencia en las estimaciones. Los métodos empleados en esta investigación representan un esfuerzo en la modelación píxel a píxel de variables de densidad forestal en México, sobre todo cuando se trata de áreas de estudio considerables. Este trabajo propone una metodología interesante para la modelación de la biomasa/carbono aéreo para satisfacer la necesidad de información de la iniciativa de la Organización de las Naciones Unidas (ONU) para la Reducción de Emisiones de la Deforestación y la Degradación Forestal (REDD, por sus siglas en inglés).

Asesor Principal: Dr. Eduardo Javier Treviño Garza.

SUMMARY

Carlos Arturo Aguirre Salado

Graduation date: January, 2013

**Universidad Autónoma de Nuevo León
Facultad de Ciencias Forestales**

Research Title: Geospatial modeling of forest density variables

Number of Pages: 105

**Doctoral Candidate in Sciences Specialized
in Natural Resources Management**

Field of study: Geomatics applied to Natural Resources Management

Purposes and Study Method: A methodological framework for geospatial modeling of forest density using moderate spatial resolution spectral data for a portion of the septentrional border of the intertropical zone: San Luis Potosí, México. Methods for parametric and nonparametric statistical modeling, spatial analysis and processing of satellite imagery were used.

Contribution and Conclusions: The Chapter one describes a brief overview of methodologies used in other countries of reference such as Finland, Switzerland, Germany, Canada, United States and Mexico and introduces the reader to topics related to forest inventory made with geospatial analysis tools. The Chapter two deals with spectral vegetation indexes in modeling of forest density and concludes that normalized indexes which are sensible to moisture content of vegetation explain better the aboveground tree biomass in a nonlinear trend. The Chapter three explores statistical methods for fractional cover estimation at subpixel level and concludes that the more suitable approach is the linear mixture analysis for two endmembers: forest and shrubland. The Chapter four compares variations of the nonparametric algorithm called k-nearest neighbor for estimating forest density using spectral and ancillary covariates and concludes that most similar neighbor estimator can be used for estimating forest density variables. The methods used in this research are applied for pixel-to-pixel modeling of forest density variables in Mexico, mainly when large study areas are considered. Therefore, it represents an interesting approach for modeling aboveground biomass/carbon to be applied for the information requirements under the United Nation's Reducing Emissions from Deforestation and Forest Degradation (REDD) initiative.

Major Advisor: Dr. Eduardo Javier Treviño Garza.

CAPÍTULO 1

INTRODUCCIÓN

1.1. Presentación

La realización de inventarios forestales es importante para manejar los recursos forestales y entender el valor de los ecosistemas. Tradicionalmente, los datos de tales inventarios han sido obtenidos mediante la realización de muestreos en campo que resultan costosos, imprácticos y tediosos. No obstante, el principal objetivo de estos inventarios tradicionales es proporcionar información confiable del recurso forestal, aunque de manera general se atribuye un bajo detalle espacial (Günlü *et al.*, 2008).

La Geomática incluye disciplinas como la Percepción Remota, los Sistemas de Información Geográfica, y los Sistemas de Posicionamiento Global, ofreciendo nuevas alternativas a la práctica del inventario forestal haciendo posible estimar variables forestales a diferentes resoluciones espaciales (Merem y Twumasi, 2008; DeFries, 2008).

Las imágenes captadas desde satélites por sensores como Aqua/Terra MODIS (Anaya *et al.*, 2009; Gallaun *et al.*, 2009; Nelson *et al.*, 2009) o incluso Landsat TM/ETM (Tomppo *et al.*, 2008), han sido empleadas satisfactoriamente para aplicaciones de gran escala, tales como inventarios forestales nacionales. Más en detalle, la información satelital ha servido de base para conocer estados de desarrollo, cobertura de copa, y dinámica de uso del suelo, siendo considerados como fuentes importantes de información, por su bajo costo, en el marco de la planeación del manejo del recurso forestal (Antonio-Némiga *et al.*, 2008; Fraser y Pouliot, 2009; Fisher, 2010). En este sentido, la Percepción Remota en el inventario de ecosistemas forestales puede ser empleada para reducir el trabajo de campo y transformar los resultados de las mediciones en

diversos mapas para ser usados en la toma de decisiones. Toda esta información es útil cuando se desea contar con un inventario forestal nacional espacialmente explícito que modele a detalle la situación dasométrica actual que guardan los recursos forestales (Köhl *et al.*, 2006).

Como parte de este capítulo, se presenta una revisión sobre generalidades de los inventarios forestales nacionales de algunos países líderes en la materia y se describe brevemente la metodología para la modelación geoespacial de variables forestales.

Posteriormente se presentan los resultados del uso de índices de vegetación en la modelación geoespacial de la biomasa arbórea aérea en una porción del límite septentrional de la zona intertropical donde se probó capacidad de los índices normalizados de vegetación construidos con el producto MODIS MCD43A4 con las bandas del infrarrojo cercano (clorofila) e infrarrojo medio (humedad), para estimar la densidad forestal mediante modelos lineales y no lineales.

En el tercer capítulo se compararon métodos para el mapeo de la cobertura forestal usando la probabilidad posterior bayesiana, el k vecino más cercano en su versión difusa y el análisis de mezcla espectral como métodos estadísticos para la estimación de cobertura fraccional a nivel subpixel.

Finalmente se compararon variaciones de algoritmos no paramétricos del k-vecino más cercano para la estimación de la biomasa superficial en donde se prueban diferentes matrices de ponderación de la distancia espectral para analizar su efecto en la precisión de las estimaciones.

1.2. Hipótesis

1. Los índices espectrales de vegetación basados en humedad predicen mejor la densidad forestal con un comportamiento no lineal.

2. La cobertura fraccional forestal puede ser estimada mediante métodos de clasificación difusa y datos espectrales.

3. El uso de diferentes matrices de ponderación de la distancia espectral para obtener los vecinos más cercanos puede mejorar las estimaciones de densidad forestal mediante el uso de datos espectrales y datos de inventario forestal.

1.3. Objetivos

1.3.1. Objetivo General

Estudiar el comportamiento de variables de densidad forestal en el Estado de San Luis Potosí mediante modelos estadísticos que permitan relacionar datos espectrales obtenidos de sensores remotos con información de inventario obtenida en campo.

1.3.2. Objetivos Particulares

1. Documentar la relación existente (lineal vs no lineal) que guardan los índices espectrales basados en clorofila y humedad en la predicción de la densidad forestal para el Estado de San Luis Potosí.

2. Comparar diferentes algoritmos de clasificación subpixel para la estimación de cobertura fraccional.

3. Estudiar el algoritmo no paramétrico del vecino más similar (kMSN) como una opción de optimización del kNN en la estimación de variables de densidad forestal mediante datos espectrales.

1.4. Antecedentes

1.4.1. Evaluación de los Bosques del Mundo

A nivel mundial, la Organización para la Agricultura y la Alimentación (FAO) de las Naciones Unidas (ONU) es una institución internacional con una gran historia en la evaluación de los recursos forestales enmarcada en el programa Forest Resources Assessment (FRA, por sus siglas en inglés) (FAO, 2006). El programa FRA tiene dos componentes principales: (1) evaluación global y reporte y (2) apoyo a las evaluaciones nacionales forestales. El componente de evaluación global tiene dos objetivos: (a) compilar, analizar, y reportar información forestal relacionada a todos los países, y (b) mantener mecanismos y acuerdos para la realización de reportes globales. Mientras que el componente de apoyo a las evaluaciones forestales nacionales se aplica en (c) apoyar a los países en desarrollo a administrar y usar la información forestal y (d) establecer un marco internacional para realizar reportes sobre los recursos forestales (FAO, 2006; Kangas, 2006).

Desde inicios del siglo XX, el Servicio Forestal de Estados Unidos preparó el primer reporte sobre recursos forestales globales (Zon, 1910; Zon y Sparhawk, 1923). No obstante, el primer inventario forestal mundial fue llevado a cabo por la FAO en 1947-1948 donde se usaban cuestionarios para obtener la información de países (Holmgren y Persson, 2002).

La superficie forestal a sido un tema importante en la evaluación de los recursos forestales globales, es decir, ha sido visto como un indicador del estado de los bosques del mundo y como un medio importante del monitoreo de los cambios. Los cambios en la superficie forestal pueden deberse a los cambios en el uso del suelo tales como reforestación, deforestación y expansión de los bosques naturales o también en cambios internos dentro de

una clase de cobertura como degradación o recuperación forestal (FAO¹, 2009; Simula, 2009).

La escasez de los recursos forestales ha sido siempre una preocupación, y el balance de las existencias y las necesidades requieren de información. La madera se necesita para construcción, para pulpa y papel, para combustible y energía y para secuestro de carbono entre otras cosas, de manera que el volumen maderable y la biomasa han sido de las variables más importantes en las evaluaciones de los recursos forestales globales (FAO², 2009). Los temas considerados actualmente, en la Evaluación de los Recursos Forestales del 2000 (FAO, 2001) son muchos más que aquellos mencionados para los inicios del siglo XX, e incluyen: (1) la superficie forestal y sus cambios, (2) volumen maderable y biomasa, (3) plantaciones forestales, (4) árboles fuera de los bosques, (5) biodiversidad biológica, (6) manejo forestal, (7) bosques en áreas protegidas, (8) incendios forestales, (9) abastecimiento de productos maderables y no maderables.

Es obvio que una evaluación forestal global ideal debería de incluir todos y cada uno de los aspectos mencionados anteriormente, en términos de los beneficios o servicios que proveen los bosques y una lista completa de todos los beneficiarios potenciales, desde usuarios locales hasta la población global. No obstante, cada país ha centrado sus esfuerzos de monitoreo determinándose por las limitantes de las evaluaciones y los intereses particulares en el conocimiento, haciendo de esta actividad desde el enfoque global un tanto fragmentada (Kangas, 2006; FAO², 2009).

1.4.2. Inventarios Nacionales Forestales

1.4.2.1. Finlandia

El desarrollo del inventario forestal nacional finlandés (NFI, por sus siglas en inglés) comenzó en 1989 y las primeras estimaciones operativas fueron

calculadas en 1990. En contraste con los métodos previos de clasificación de imágenes, en un inventario se requieren métodos que arrojen estimaciones de volumen, posiblemente por subcategorías como especies arbóreas, clases de edad y productos forestales. En el caso óptimo, un método debería ser útil para dar estimaciones para áreas pequeñas tan bien como el método basado en campo que da estimaciones a niveles regionales y nacionales (Figura 1.2). En este sentido se probaron análisis de regresión y análisis discriminante (Muukkonen y Heiskanen, 2007). Las variables tenían que ser predichas separadamente en estos métodos. Las experiencias principales fueron que era difícil obtener un grado suficiente de detalle de la información y la estructura de la dependencia de los datos se perdía cuando las estimaciones de los diferentes parámetros se hacían de manera independiente (Tomppo *et al.*, 2008). Para evitar este tipo de problemas, se desarrolló un método llamado el k vecino más cercano (kNN) que consiste en realizar estimaciones con base en la interpolación de la variable de interés basándose en el espacio espectral (Altman, 1992; Sironen *et al.*, 2008; Tomppo *et al.*, 2008, 2009).

1.4.2.2. Suecia

Después del ejemplo finlandés, un desarrollo similar comenzó en Suecia unos años después. Actualmente, el IFN sueco está basado en muestreo sistemático con parcelas tipo cluster rectangular dentro de las cuales existen subparcelas circulares. 12,000 parcelas son permanentes y se visitan anualmente (desde 1983), en tanto 6000 parcelas son temporales y se visitan una sola vez; en ambas se miden árboles individuales y son empleados para estimar el volumen total de los árboles, el área cubierta por un cierto tipo de vegetación entre otras cosas. Dichos clusters varían en tamaño en diferentes partes de Suecia, son distribuidos sistemáticamente en todo el país, aunque son más cercanos en el sur que en el norte (Kangas, 2006) (Figura 1.1).

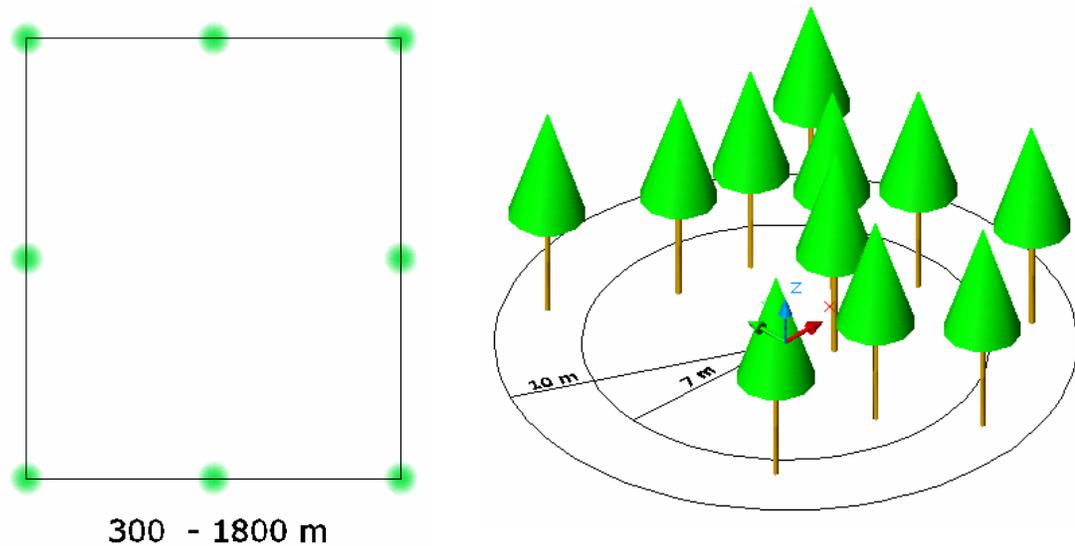


Figura 1.1. El cluster y la parcela empleada en el IFN sueco. Tomado y adaptado de Kangas (2006).

En este tipo de parcelas se miden aproximadamente 200 atributos relacionados con: (1) árboles y arbustos, (2) vegetación herbácea, (3) condiciones ecológicas y de paisaje y (4) muestras de suelo. El IFN sueco presenta sus resultados en una publicación anual llamado *Skogsdata*. Los resultados desde 1983 hasta la fecha pueden ser accedados fácilmente en <http://www-nfi.slu.se/>. Las estimaciones geoespaciales se realizan mediante el método no paramétrico del vecino más cercano (KNN) dadas sus bondades estadísticas (Kajisa *et al.*, 2008; Tomppo *et al.*, 2008) (Figura 1.2).

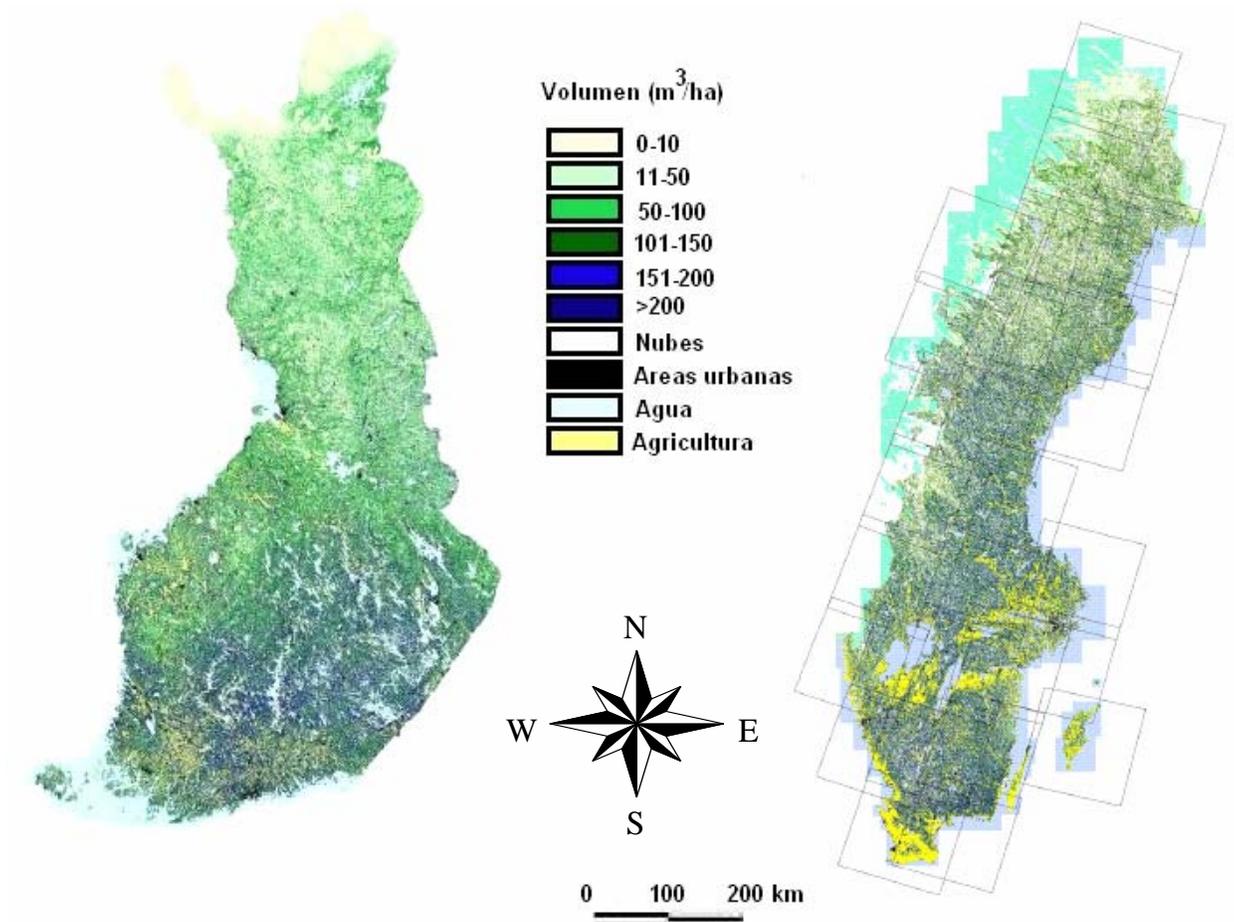


Figura 1.2. Estimación de volumen maderable resultado de los Inventarios Nacionales Forestales apoyados en datos espectrales. Izq. Finlandia. Der. Suecia. Adaptado de Tomppo *et al.* (2008).

1.4.2.3. Alemania

El uso de métodos de percepción remota en Alemania es limitado al uso de fotografía aérea u ortofotos consultadas para propósitos de actualización, el mapeo de datos forestales, la definición de rodales en el contexto de la planeación del manejo forestal operativo y como un apoyo en la orientación en inventarios al nivel de empresas forestales y plantación regional. Los inventarios en Alemania son realizados en todo el país y sirven particularmente para propósitos de política forestal; donde se determina el tamaño y la distribución de las áreas forestales y las reservas de madera (NFI), el grado del degradación

forestal (NFDI) y las condiciones prevalecientes del suelo (NSCS). El primer NFI se llevó a cabo durante el periodo 1986-1990, mientras que el trabajo de campo para el segundo inventario fue en 2001-2002 y los resultados fueron publicados por Polley (2001) y Bundeswaldinventur (2002). En la Figura 1.3 se muestra un esquema de la distribución de las parcelas de muestreo en los clusters para el IFN alemán.

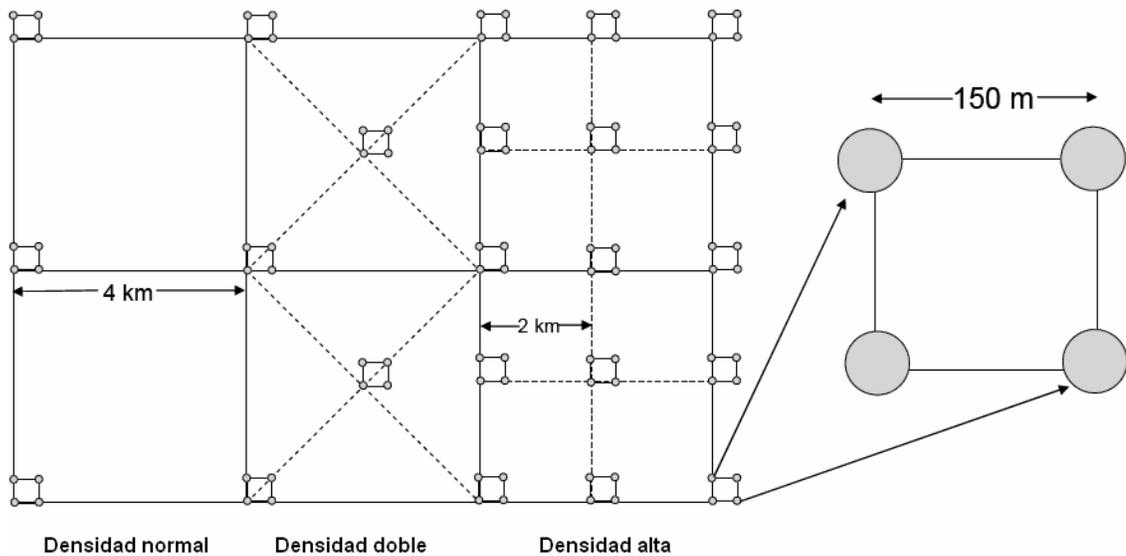


Figura 1.3. El cluster y la parcela empleada en el IFN alemán. Tomado y adaptado de Tokola (2006).

1.4.2.4. Canadá

La manera en que Canadá ha obtenido, analizado y reportado información forestal a nivel nacional ha cambiado continuamente. Antes de 1981, consistía en una compilación de cuestionarios que eran llenados por las agencias de manejo forestal (provinciales y federal). Actualmente esta tarea la realiza un consorcio de agencias provinciales independientes quien conduce la compilación del National Forest Inventory (NFI) a partir de inventarios forestales. Colectivamente, el NFI ayuda a monitorear el estado de los bosques canadienses en una manera oportuna y precisa. Las provincias y territorios individuales son quienes han administrado sus recursos naturales. Cada

jurisdicción individual tiene control sobre la definición y colección de los atributos forestales. Consecuentemente, las técnicas y los atributos colectados no son necesariamente consistentes a lo largo de todo el país (Rommel *et al.*, 2005).

No obstante, el Servicio Forestal de Canadá ha preparado un marco para compilar una serie de variables que indican el nivel mínimo de reporte para las provincias individuales y los territorios. La estructura de muestreo del NFI se encuentra distribuida en todo el territorio canadiense en una malla con un espaciamiento de 20 km (basado en la proyección de Igual Área de Albers). En los nodos de esta malla se localizan unidades de muestreo de 2 km x 2 km (aproximadamente 20,000 en todo el territorio canadiense) las cuales suman aproximadamente un 1% de la muestra nacional. Los atributos de cada unidad muestral son compilados y guardados en una base de datos vectorial (Rommel *et al.*, 2005; Gillis *et al.*, 2005).

1.4.2.5. Estados Unidos

El inventario forestal estadounidense, denominado Forest Inventory and Analysis (FIA) iniciado desde 1930, es quien reporta los indicadores cuantitativos y cualitativos del recurso forestal entre los que destacan el área forestal y sus tendencias, las especies existentes, el tamaño y salud de los árboles, el crecimiento, la mortalidad y la remoción por concepto de manejo forestal, y las tasas de uso en términos de diversos productos relacionados a la propiedad del recurso forestal.

Actualmente el programa del FIA mejorado (FIA Enhanced) incluye el levantamiento de información relacionada con las condiciones de las copas de los árboles, condiciones edáficas, plantas indicadoras de ozono, biodiversidad vegetal de herbáceas y combustibles sobre el piso forestal. Estos datos de inventario apoyan estimaciones forestales totales sobre grandes áreas geográficas. La elaboración de mapas regionales de los atributos del bosque

hace que estos datos sean más accesibles y útiles a una audiencia más grande y diversa. Algunas de las aplicaciones más importantes incluyen la migración de especies arbóreas (Woodall *et al.*, 2009), identificación de áreas aptas para la producción maderable (Blackard *et al.*, 2008), secuestro de carbono (Pan *et al.*, 2009), la localización de áreas infestadas de plantas invasoras (Huebner *et al.*, 2009), etc. Por lo que existe la necesidad de producir y distribuir datos geospaciales de atributos forestales complementando los datos de inventario del FIA.

Es reconocible el hecho de que el FIA sea un programa nacionalmente consistente y puede ser mejorado a nivel regional, estatal, o local para abordar intereses especiales. Entre las herramientas que usan se encuentran mapas, fotografías aéreas e imágenes de satélite y sistemas de posicionamiento global (GPS) para ubicar correctamente las parcelas en campo (Sivanpillai *et al.*, 2005) (Figura 1.4).

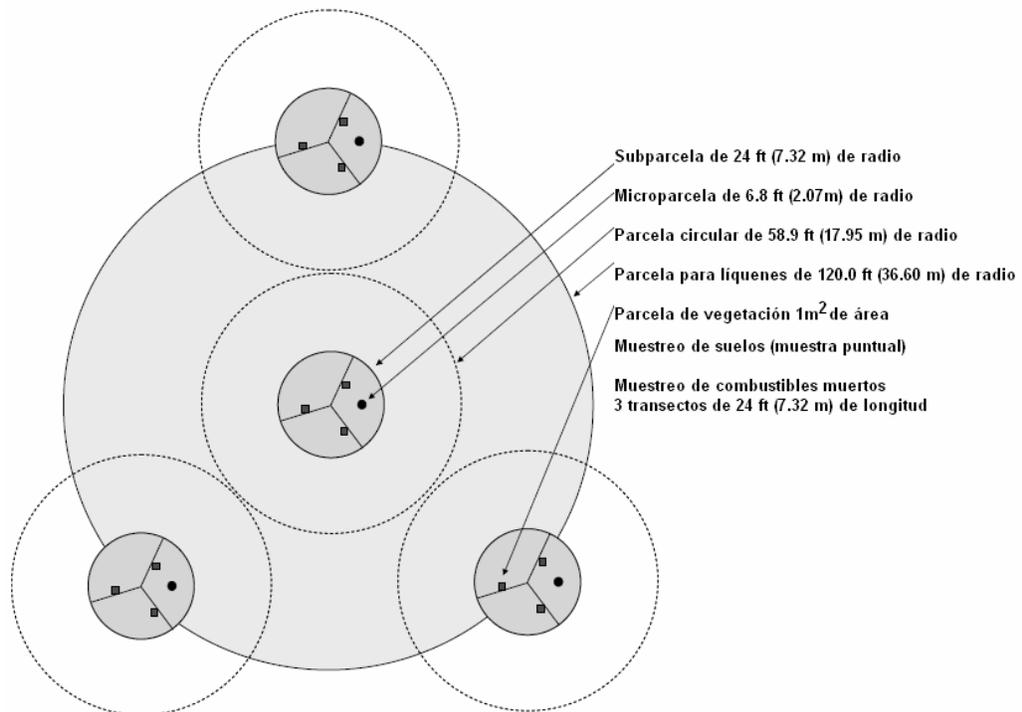


Figura 1.4. Esquema de una parcela en campo del FIA estadounidense. Tomado y adaptado de Tokola (2006).

1.4.2.6. México

México ha llevado a cabo tres proyectos que pueden ser identificados como inventarios nacionales forestales. El primer Inventario Forestal (1964-1980) se basó en el uso de imágenes aéreas apoyadas con muestreo extensivo en campo para establecer la ubicación, el tamaño, los volúmenes maderables y el valor comercial de las áreas forestales. El segundo y tercer inventario fueron básicamente actualizaciones de mapas de vegetación y uso del suelo empleando imágenes de sensores remotos. El segundo inventario fue basado en el análisis de imágenes del sensor AVHRR (Advanced Very High Resolution Radiometer) y fue publicado a escala 1:1'000,000 como parte del reporte de *Gran Visión* de la Secretaría de Agricultura y Recursos Hidráulicos (SARH). El objetivo de este inventario era producir una estimación rápida del tamaño de las tierras forestales para apoyar las necesidades de información a nivel nacional de México para los tratados internacionales (Tokola, 2006).

El tercer inventario nacional de México se completó en 1994 llamándose Inventario Nacional Forestal Periódico (1992-1994) y fue un mapa de vegetación y uso de suelo basado en interpretación visual de imágenes Landsat Thematic Mapper (TM) y mediciones de campo de aproximadamente 20,000 parcelas distribuidas en un muestreo sistemático para la determinación del tipo de vegetación. La distancia entre parcelas del INF fue de 5 km x 5 km para bosques templados y selvas, 10 km x 10 km para matorrales, y 20 km x 20 km para zonas áridas. La escala de producción cartográfica fue de 1:250,000 con el propósito de generar información sobre la ubicación, el tamaño y el volumen maderable de las tierras forestales para apoyar las necesidades operativas del país. De hecho, este inventario forestal solo produjo datos de ubicación y de tamaño para el 75% del país y no se estimaron datos de volúmenes (Tokola, 2006; Sandoval-Uribe *et al.*, 2008). La confiabilidad de este tipo de mapas ha sido evaluada por Couturier *et al.* (2008) encontrando valores ligeramente arriba del 80% (Figura 1.5).

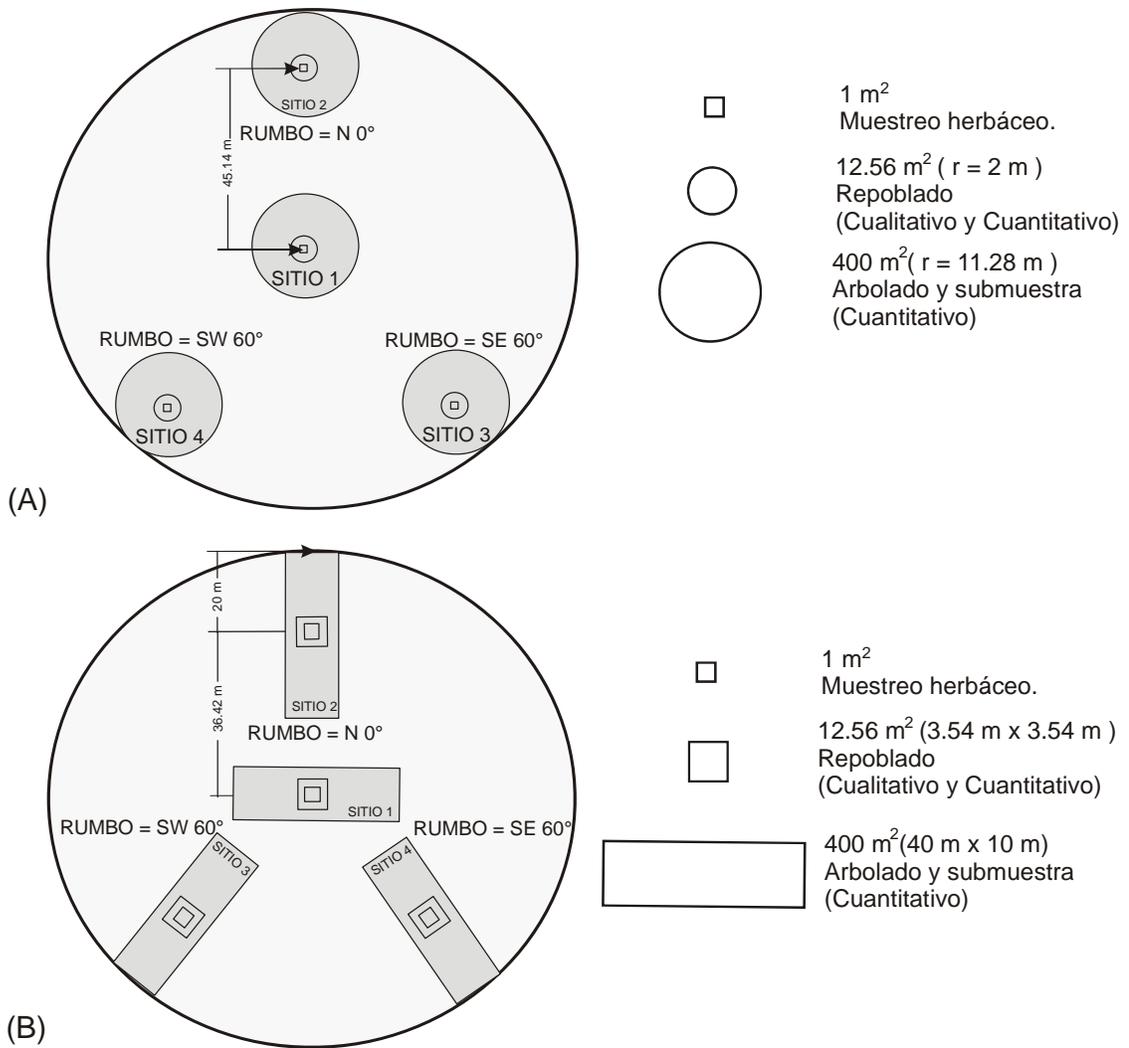


Figura 1.5. Esquema de las unidades de muestreo del Inventario Nacional Forestal de México: (A) bosque templado y (B) selva.

1.4.3. Procesamiento de Imágenes de Satélite para el Estudio del Recurso Forestal

1.4.3.1. Clasificación de los Métodos de Interpretación de Imágenes

Los métodos de interpretación de imágenes han evolucionado mostrando dos tendencias principales, la primera orientada a la clasificación de píxeles y la segunda orientada a la identificación de objetos (Figura 1.6). Los métodos basados en píxeles han sido los más desarrollados desde la puesta en órbita de los primeros satélites (ERTS, Landsat), ya que sus relativamente bajas resoluciones espaciales (30 a 1000 m de tamaño de píxel) registra el promedio de la reflectancia de todos y cada uno de los objetos que se incluyeran dentro de los píxeles. Por otro lado, los métodos basados en objetos iniciaron a la par de los métodos de fotointerpretación de fotografías aéreas (formato análogo) y se reafirmaron con la creación de los nuevos sensores satelitales de alta resolución espacial refrendando la idea del procesamiento automatizado basado en el reconocimiento de patrones (textura, forma y contexto).

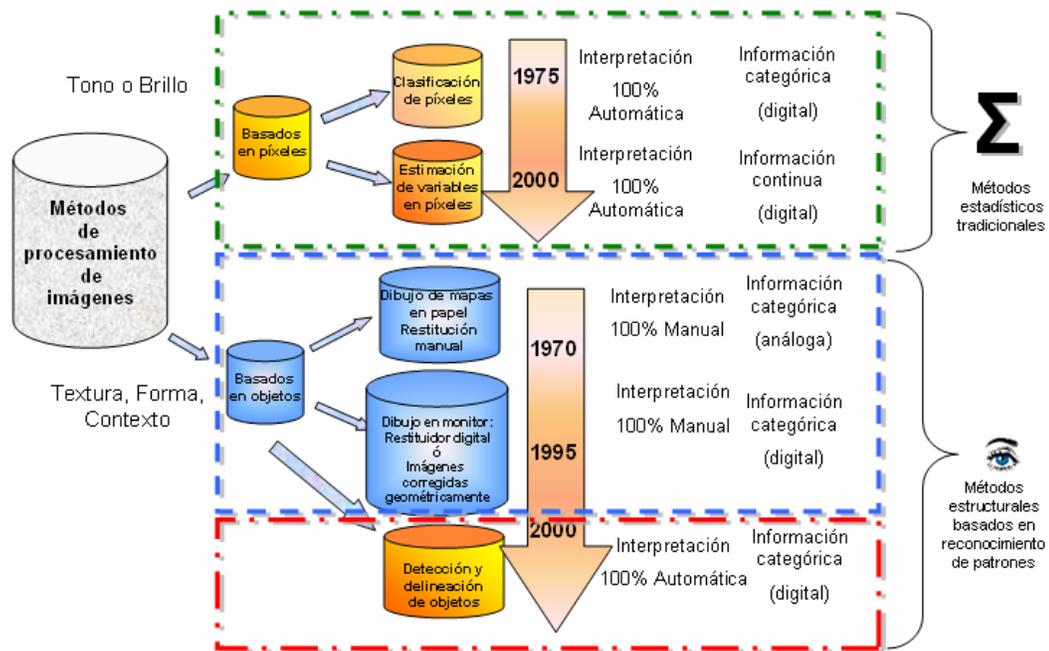


Figura 1.6. Marco conceptual de los métodos de interpretación de imágenes (elaboración propia).

1.4.3.2. Metodología General del Procesamiento de Imágenes

En términos amplios, este proceso se puede subdividir en cinco grandes fases: muestreo en campo, pre-procesamiento de la imagen de satélite, análisis de correlación de datos de campo vs datos espectrales, estimación de las variables dasométricas de interés mediante el método estadístico seleccionado y validación de los resultados obtenidos (Chirici *et al.*, 2008; Günlü *et al.*, 2008).

1.4.3.2.1. Muestreo en Campo. Según el estudio que se trate, ya sea la estimación de variables dasométricas de interés o la clasificación de coberturas y densidades de la imagen de satélite, será el tipo de información levantada en campo. Para el caso de estimación de variables dasométricas (área basal, volumen, biomasa) se levantan parcelas de muestreo donde se miden atributos de los árboles individuales, como el diámetro normal y la altura mismos que posteriormente, vía relaciones alométricas, permiten estimar su biomasa. Posteriormente, los valores individuales de cada árbol se suman para obtener estimaciones a nivel parcela y finalmente mediante extrapolaciones se calculan por unidad de superficie. Estas parcelas son georeferidas con un receptor GPS asegurando una buena exactitud posicional, para propiciar la correspondencia espacial de la información obtenida del sitio de muestro y la reflectancia de la imagen de satélite (García *et al.*, 2010).

Si el caso es la clasificación de coberturas simplemente bastará con conocer el tipo de vegetación presente en el punto de verificación, para asignar campos de entrenamiento durante el proceso de la obtención de la firma espectral característica del tipo de vegetación/densidad de que se trate.

1.4.3.2.2. Pre-procesamiento de la Imagen de Satélite. Antes de ser empleados en algún análisis, los datos espectrales tienen que ser preparados. El pre-procesamiento de la imagen normalmente incluye tres etapas: corrección geométrica, corrección radiométrica y transformaciones matemáticas.

La corrección geométrica o georeferenciación consiste en propiciar la correspondencia entre la ubicación de los píxeles de la imagen de satélite y un sistema de referencia espacial. Para ello se requiere localizar puntos de control terrestre en la imagen a los cuales se les determinan sus coordenadas, normalmente con receptores GPS. Posteriormente, mediante un análisis de regresión lineal múltiple entre las coordenadas origen (X_{origen}, Y_{origen}) y las coordenadas destino ($X_{destino}, Y_{destino}$) se calcularán los parámetros de ajuste, mismos que servirán para proyectar todos y cada uno de los píxeles al sistema de referencia espacial. Después se selecciona un método de remuestreo para otorgar a los píxeles un valor de reflectancia en la nueva ubicación. En este sentido existen tres opciones: vecino más cercano, interpolación bilinear y convolución cúbica. Para conservar la variabilidad espacial de los valores espectrales en la imagen de satélite se recomienda emplear el método del vecino más cercano, ya que los dos restantes actúan como un filtro de paso bajo que promedia los valores de reflectancia de los nueve o dieciséis píxeles más cercanos. Existen rutinas implementadas en diversos sistemas de información geográfica (SIG) que realizan el procedimiento de manera automatizada. Este procedimiento es de suma importancia ya que la precisión obtenida derivará en la correcta correspondencia espacial entre los sitios de muestreo georeferidos con GPS y los valores de reflectancia de la imagen (Toutin, 2004).

El otro procedimiento importante que debe realizarse antes de analizar la imagen particularmente cuando se pretende estimar o extraer parámetros biofísicos (biomasa, IAF, clorofila, etc), es la corrección radiométrica. Los valores de energía reflejada reportados por el sensor remoto son números o niveles digitales (ND) que son afectados por las condiciones de toma de la escena. La topografía, la hora de toma y la condición atmosférica condicionarán la iluminación de la imagen, influyendo así en la calidad de los datos captados. Para minimizar estos efectos, se emplean diversos procedimientos denominados de manera genérica, calibración o estandarización radiométrica,

los cuales consisten inicialmente en convertir los niveles digitales (ND) de la imagen (valores de 0 a 255, para imágenes en formato de 8 bits) a valores adimensionales (reflectancia) que sean numéricamente comparables entre escenas, entre fechas e inclusive entre sensores (Schaepman-Strub *et al.*, 2006).

Una vez lograda la correspondencia espacial (corrección geométrica) entre la muestra y los valores espectrales, y minimizados los efectos de la condición de toma de la escena en los valores de la imagen (corrección radiométrica), es posible intensificar la detección de las variables biofísicas de interés (clorofila, humedad, biomasa, carbono e IAP) mediante transformaciones matemáticas que incorporan los datos originales registrados en las bandas del sensor: entre éstas destacan los índices de vegetación (Glenn *et al.*, 2008), propiciando al final una mejor correlación entre la información óptica de la imagen y los valores de la o las variables obtenidos en campo.

Las transformaciones matemáticas se pueden entender analizando la firma espectral típica de la vegetación sana (Figura 1.7), donde de manera natural existen cambios de la reflectancia en las diferentes porciones del espectro electromagnético *v.gr.* rojo e infrarrojo. Dichos cambios hacen posible que el cociente normalizado entre ellas indique aspectos relacionados a la cantidad de superficie fotosintéticamente activa y por ende, la descripción de la densidad forestal (Chen *et al.*, 2005). En la Figura 4 se indica que la clorofila de las plantas absorbe fuertemente energía para el proceso fotosintético. El aspecto verde que observamos en la vegetación se debe a que en la parte visible del espectro electromagnético (azul-verde-rojo), las hojas absorben la parte azul y roja, complementariamente reflejando el verde. En el infrarrojo cercano ocurre una interacción similar a la del verde, esta energía no es utilizada para la fotosíntesis por lo que es fuertemente reflejada, no obstante no se percibe por el ojo humano. El efecto diferencial entre la absorción en el rojo y

el reflejo en el infrarrojo cercano ha sido el principio físico fundamental de diversos índices cuantitativos para la caracterización de la vegetación, denominados genéricamente: índices de vegetación (Glenn *et al.*, 2008).

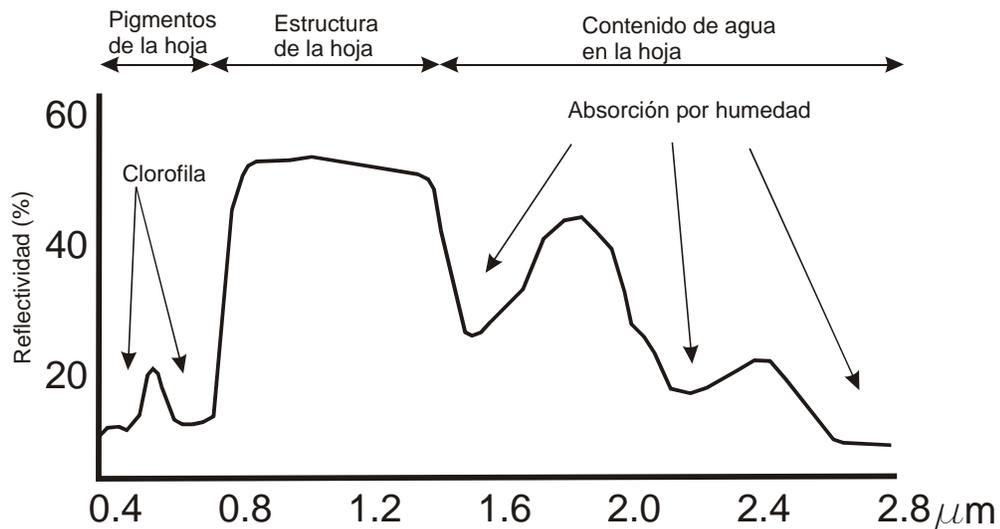


Figura 1.7. Comportamiento teórico de la reflectividad de la vegetación vigorosa (CCRS, 1998).

1.4.3.2.3. Relación Datos de Campo vs Datos Espectrales. Para el caso de variables continuas, una manera de investigar el comportamiento de las variables forestales *versus* la reflectancia o transformaciones matemáticas obtenidas de las imágenes de satélite, es mediante un análisis de correlación y gráficas de dispersión de los datos. La idea es conocer el grado de asociación y signo de la correlación existente entre las variables de densidad forestal y los datos espectrales, y de esta manera poder realizar una estimación confiable (McRoberts *et al.*, 2008).

Mientras que en el caso de variables categóricas *v.gr.* tipos de vegetación/grados de densidad, dicha relación se podrá averiguar mediante la construcción de firmas espectrales, la estimación de la separabilidad de clases, y análisis de matrices de confusión construidas con los datos de entrenamiento.

Aquí también se podría incluir el uso de modelos estadísticos experimentales (paramétricos y no-paramétricos) para estimar las variables de interés con los datos espectrales en los lugares donde se realizaron los sitios de muestreo; conociendo el valor estimado con el valor observado en campo se podrá obtener un error *a priori* o error estándar de la estimación.

1.4.3.2.4. Estimación de las Variables de Interés o Clasificación de la Imagen. Esta fase también es conocida como el mapeo de la variable seleccionada (categórica o continua) a partir del modelo estadístico seleccionado y construido con los datos de la muestra (entrenamiento).

- Análisis de regresión (paramétrico). Mediante el análisis de regresión se puede encontrar el tipo de relación existente entre los datos espectrales provenientes del sensor remoto y la variable de respuesta (discreta o continua) de interés. Las variables independientes son las bandas espectrales de las imágenes, cada una cuenta con diferente variabilidad y detectan aspectos complementarios de las coberturas: clorofila (rojo e infrarrojo cercano), humedad (infrarrojo de onda corta e infrarrojo lejano) y temperatura (infrarrojo térmico). Mientras que las variables dependientes incluyen a las variables de la densidad forestal (volumen en m^3/ha o, biomasa en Mg/ha). La estimación simplemente bastará con evaluar la ecuación obtenida con los datos de la imagen; es decir, multiplicando las ponderaciones o coeficientes obtenidos para cada banda espectral para obtener la estimación espacialmente explícita de la variable de interés. La literatura reporta casos donde se emplea la regresión lineal entre las reflectancias por banda o sus transformaciones matemáticas (índices de vegetación u otras) y la variable de respuesta (Anaya *et al.*, 2009).

- Vecino más cercano (kNN) (no paramétrico). Este método ha sido muy utilizado en la clasificación del uso de suelo y estimación de parámetros forestales con el uso de imágenes de satélite principalmente en países de la

península escandinava. Las estimaciones continuas se hacen mediante el cálculo de una media ponderada de los k vecinos espectrales más cercanos. Mientras que para variables categóricas, se realiza mediante la moda de los k vecinos espectrales más cercanos. Dicha ponderación puede implementarse mediante el inverso de la distancia euclidiana (McRoberts *et al.*, 2008; Zhou *et al.*, 2008).

- Máxima Verosimilitud. Este método sirve para generar información categórica (tipos de vegetación) a partir de los datos de reflectancia. Es un clasificador que utiliza información de campos de entrenamiento, como la media, varianza y covarianza para estimar la probabilidad posterior de que un píxel pertenezca a determinada clase. Existe una gran similitud entre este procedimiento y el de mínima distancia a la media con la opción de distancias estandarizadas. La diferencia está en que máxima verosimilitud toma en cuenta la correlación entre bandas y la varianza interna. Existen una buena cantidad de trabajos que utilizan este tipo de clasificación encontrando resultados satisfactorios (Miranda-Aragón *et al.*, 2009).

- Análisis discriminante de Fisher. Este método también sirve para generar información categórica (tipos de vegetación) a partir de los datos de reflectancia, mediante el análisis de los datos de los campos de entrenamiento para formar un conjunto de funciones lineales que expresan el grado de soporte de cada clase. La clase asignada para cada píxel es entonces aquella clase que recibe el mayor soporte después de evaluar a todas las funciones. Estas funciones tienen la forma similar de una ecuación de regresión lineal múltiple, donde las variables independientes son las bandas y la variable dependiente es la clase deseada. De hecho, las ecuaciones son calculadas de manera que maximicen la varianza entre las clases y minimicen la varianza dentro de las clases. El número de ecuaciones será igual al número de bandas, cada una describiendo un hiperplano de soporte. Las intersecciones de estos planos

forman los límites entre clases en términos del espacio espectral (Mallinis *et al.*, 2004; Salovaara *et al.*, 2005).

- Redes neuronales. Es una técnica de estimación y clasificación perteneciente a la inteligencia artificial. Tiene como principio el proceso de aprendizaje que intenta simular la conducta cognitiva del cerebro humano. Una red neuronal se caracteriza por: (1) el patrón de conexiones entre neuronas (arquitectura o topología), (2) el método para determinar los pesos (ponderaciones) de las conexiones (entrenamiento o aprendizaje) y (3) su función de activación. La neurona X_i , es el elemento básico de una red neuronal; cada neurona tiene una salida, la cual está generalmente relacionada al estado de la neurona (activación) y puede dirigir salidas a varias neuronas. Cada neurona recibe sinapsis o ligas de conexión. El vector de entradas se conforma por el estado de las neuronas de entrada multiplicada por los pesos de las sinapsis (W_i). La activación de la neurona es procesada por la aplicación de una función umbral (q) a esta neurona (Razi y Athappilly, 2005).

- Árboles de decisión. Los modelos de árboles de regresión (Blackard *et al.*, 2008) y clasificación (Hansen *et al.*, 2008) son un conjunto de condiciones (if-then) que permiten la predicción de variables de interés o clasificación de casos. Un modelo de árbol que predice el valor de variables continuas a partir de un conjunto de variables predictoras (categóricas o continuas) se le llama árbol de regresión. Por su parte, un modelo de árbol que predice variables categóricas a partir de un conjunto de variables predictoras categoricas o continuas toma el nombre de árbol de clasificación (Razi y Athappilly, 2005). Este método promete buenas expectativas ya que el proceso genera una mayor confiabilidad *per se*, debido a que la estimación se realiza con un algoritmo en multifase, permitiendo una mayor flexibilidad: para el caso de datos categóricos será con las clases asignadas o mientras que en los datos continuos será con los valores estimados para cada píxel (Reich *et al.*, 2008).

- Clasificación orientada a objetos. La gran mayoría de los algoritmos están enfocados a estudiar imágenes donde el valor del píxel representa la mezcla y/o el promedio de la reflectancia de los objetos que se encuentran dentro del mismo. No obstante, los sensores con resolución espacial fina (Ikonos o Quickbird, entre otros) detectan la variabilidad natural ocasionada por la rugosidad de las ramas y el follaje, la sombra de la copa de los árboles; en este sentido, las técnicas tradicionales no funcionan como se espera (Lewinski y Zaremski, 2004).

Más que medir o caracterizar la reflectancia de píxeles individuales, resulta necesaria la lógica de la fotointerpretación automatizada donde se involucren factores como la forma y la textura de los objetos que los hagan claramente distinguibles de los demás. La relación con las variables forestales estará basada en la relación del número y tamaño de las copas detectadas con el atributo individual medido en campo, ya sea biomasa o volumen de madera por árbol.

1.4.3.2.5. Validación del Mapa Obtenido. La fase final consistirá en averiguar la fiabilidad de los resultados obtenidos mediante la comparación de los valores del mapa (estimados) *versus* valores obtenidos nuevamente en campo. Cuando se trata de variables cuantitativas, *v.gr.* volumen o biomasa, esto se puede realizar mediante el cálculo del error estándar; mientras que si se trata de variables cuantitativas, *v.gr.* clases de densidad por tipos de vegetación, dicha evaluación se tendrá que realizar mediante matrices de confusión, contingencia, o error (Treitz y Rogan, 2004). A pesar de lo costoso que resulta realizar una validación espacialmente explícita de los modelos desarrollados para estimar variables asociadas al recurso forestal, es necesario esforzarse por llevarla a cabo, de lo contrario los resultados (mapas) obtenidos son de utilidad limitada para la toma de decisiones.

Cuando se considera el uso de la percepción remota en los inventarios forestales es necesario tomar en cuenta sus costos en relación con los beneficios; es decir, las reducciones de los costos del trabajo de campo, contra los costos relativos a la adquisición, preprocesamiento e interpretación de las imágenes de sensores remotos. Si se considera que las imágenes de sensores remotos traerán un beneficio, el siguiente paso será considerar el tipo de sensor a usar y la metodología de procesamiento a emplear. Es decir, si las imágenes se adquirirán mediante vuelos fotogramétricos o por satélite, y cuál sería la resolución (espacial, espectral, radiométrica y temporal) de la imagen más apta. La respuesta a estas preguntas dependerá del tamaño del área de inventario, el propósito del inventario, el nivel deseado de exactitud y los costos de las imágenes. Es decir, si el propósito del inventario será, cuantificar la superficie de áreas boscosas para propósitos de producción maderable, tal vez no se requerirá información altamente detallada. Por otro lado, si el propósito es evaluar las características de árboles individuales, entonces el uso de imágenes aéreas ópticas o el uso de LIDAR será una mejor opción (Holopainen y Kalliovirta, 2006; Anderson *et al.*, 2009).

1.5. Literatura Citada

- Altman, N. 1992. Introduction to kernel and nearest neighbour nonparametric regression. *The American Statistician*. 46(3):175-184.
- Anaya, J.A., Chuvieco, E. and Palacios-Orueta, A. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management*. 257:1237–1246.
- Anderson, K., Evans, T.P. and Richards, K.R. 2009. National forest carbon inventories: policy needs and assessment capacity. *Climatic Change*. 93:69-101.
- Antonio-Némiga, X., Treviño-Garza, E.J. y Jurado-Ybarra, E. 2008. Fragmentación forestal en la subcuenca del río Pilón: diagnóstico y prioridades. *Madera y Bosques*. 14(3):5-23.
- Blackard., J.A., Finco, M.V., Helmer, E.H., Holden, E.H., Hoppus, M.L., Jacobs, D.M., Lister, A.J., Moisen, G.G., Nelson, M.D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D.L., Winterberger, K.C., Brandeis, T.J., Czaplewski, R.L., McRoberts, R.E., Patterson, P.L. and Tymcio, R.P. 2008. Mapping U.S. forest biomass using nation wide forest inventory data and moderate resolution information. *Remote Sensing of Environment*. 112:1658-1677.
- Bundeswaldinventur. 2002. Disponible en: <http://www.bundeswaldinventur.de/enid/3c5cb22ae9bb0871ca98dcc648736adf,0/a9.html>. Consultada: 18 de mayo de 2010.
- CCRS. 1998. Fundamentals of remote sensing. A Canada Centre for Remote Sensing Remote Sensing Tutorial. Natural Resources Canada. 258 p.
- Chen, X., Vierling, L., Deering, D. 2005. A simple and effective radiometric correction method to improve landscape change detection across sensors and across time. *Remote Sensing of Environment*. 98:63-79.
- Chirici, G., Barbati, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F. and Bertini, R. 2008. Non-parametric and parametric methods using satellite

- images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sensing of Environment*. 112: 2686-2700.
- Couturier, S., Vega, A., Mas, J.F., Tapia, V. y López-Granados, E. 2008. Evaluación de confiabilidad del mapa del Inventario Nacional Forestal 2000: diseños de muestreo y caracterización difusa de paisajes. *Investigaciones Geográficas. Boletín del Instituto de Geografía UNAM*. 67:20-38.
- DeFries, R. 2008. Terrestrial vegetation in the coupled human-earth system: contributions of remote sensing. *Annual Review on Environmental Resources*. 33:369-390.
- FAO. 2001. *Global Forest Resources Assessment 2000*. FAO Forest Paper, No. 140. FAO, Rome.
- FAO. 2006. *Global Forest Resources Assessment 2005, Main Report. Progress Towards Sustainable Forest Management*. FAO Forestry Paper 147, Rome.
- FAO. 2009. *State of World's forests 2009*. Food and Agriculture Organization of the United Nations. Rome, Italy. 168 p.
- FAO¹. 2009. *FRA 2010 remote sensing survey - An outline of objectives, data, methods and approach*. Forest Resources Assessment Programme. FAO Working Paper 155. Rome, Italy. 16 p.
- Fisher, P. 2010. Remote sensing of land cover classes as type 2 fuzzy sets. *Remote Sensing of Environment*. 114:309–321.
- Fraser, R.H. and Pouliot, I.O.D. 2009. Monitoring land cover change and ecological integrity in Canada's national parks. *Remote Sensing of Environment*. 113:1397–1409.
- Gallaun, H., Zanchi, G., Nabuurs, G.J., Hengeveld, G., Schardt, M. and Verkerk, P. 2009. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*. 260(3):252-261.

- García, M., Riaño, D., Chuvieco, E., Danson, F.M. 2010. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment*. 114:816–830.
- Gillis, M.D., Omule, A.Y. and Brierley, T. 2005. Monitoring Canada's forests: The National Forest Inventory. *The Forestry Chronicle*. 81(2):214-221.
- Glenn, E.P., Huete, A.R. Nagler, P.L. and Nelson, S.G. 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. *Sensors* 8: 2136-2160.
- Günlü, A., Sivrikaya, F., Zeki, B.E. Keles, S., Çakir, G. and İhsan, K.A. 2008. Estimation of stand type parameters and land cover using Landsat-7 ETM Image: A case study from Turkey. *Sensors*. 8:2509-2525.
- Hansen, M.C. Roy, D.P., Linquist, E., Adusei, B., Justice, C., Altstatt, A, 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*. 112:2495-2513.
- Holmgren, P. and Persson, R. 2002. Evolution and prospects of global forest assessments. *Unasylva*. 53(210):1-10.
- Holopainen, M. and Kalliovirta, J. 2006. Modern data acquisition for forest inventories. In: Kangas, A. and Maltamo, M. 2006. *Forest Inventory: Methodology and Applications*. Springer. Dordrecht, The Netherlands. pp. 279-291.
- Huebner, C.D., Morin, R.S., Zubriggen, A., White, R.L., Moore, A. and Twardus, D. 2009. Patterns of exotic plant invasions in Pennsylvania's Allegheny National Forest using intensive Forest Inventory and Analysis plots. *Forest Ecology and Management*. 257:258–270.
- Kajisa, T., Murakami, T., Mizoue, N., Kitahara, F. and Yoshida, S. 2008. Estimation of stand volumes using the k-nearest neighbors method in Kyushu, J. *Journal of Forest Research*. 13:249-254.

- Kangas, A. 2006. Assessing the world's forests. In: Kangas, A. and Maltamo, M. 2006. *Forest Inventory: Methodology and Applications*. Springer. Dordrecht, The Netherlands. pp. 279-291.
- Köhl, M., Magnussen, S.S., Marchetti, M. 2006. *Sampling Methods, Remote Sensing and GIS Multisource Forest Inventory*. Springer. p. 190.
- Lewinski, S., and Zaremski, K. 2004. Examples of object-oriented classification performed on high resolution satellite images. *Miscellanea Geographica*. 11: 349-358.
- Mallinis, G., Koutsias, N., Makras, A., Karteris, M. 2004. Forest Parameters Estimation in a European Mediterranean Landscape Using Remotely Sensed Data. *Forest Science*. 50(4):450–460.
- McRoberts, R.E. 2008. Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. *Remote Sensing of Environment*. 112: 2212-2221.
- Merem, E.C. and Twumasi, Y.A. 2008. Using geospatial information technology in natural resources management: the case of urban land management in West Africa. *Sensors* 8: 607-619.
- Miranda-Aragón, L., Ibañez-Castillo, L.A., Valdez-Lazalde, J.R. y Hernández de la Rosa, P. 2009. Modelación hidrológica empírica del gasto de 100 años de periodo de retorno del Río Grande, Tlalchapa Guerrero en dos escenarios de uso del suelo. *Agrociencia*. 43:333-344.
- Muukkonen, P. and Heiskanen, J. 2007. Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories. *Remote Sensing of Environment*. 107:617–624.
- Nelson, R., Ranson, K.J., Sun, G., Kimes, D.S., Kharuk, V. and Montesano, P. 2009. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of Environment*. 113:691–701.
- Pan, Y., Birdsey, R., Hom, J. and McCullough, K. 2009. Separating effects of changes in atmospheric composition, climate and land-use on carbon

- sequestration of U.S. Mid-Atlantic temperate forests. *Forest Ecology and Management*. 259:151–164.
- Polley, H, ed. 2001. Survey instructions for Federal Forest Inventory II, (2001-2002) 2nd corrected and revised reprint, May 2001. Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft (BMVEL), Bonn, Germany. 109 pp.
- Razi, M.A. and Athappilly, K. 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*. 29:65-79.
- Reich, R.M. Aguirre-Bravo, C. and Bravo, V.A. 2008. New approach for modeling climatic data with applications in modeling tree species distributions in the states of Jalisco and Colima. *Journal of Arid Environments*. 72(7):1343-1357.
- Rommel, T.K., Csillag, F., Mitchell, S. and Wulder, M.A. 2005. Integration of forest inventory and satellite imagery: a Canadian status assessment and research issues. *Forest Ecology and Management*. 207:405-428.
- Salovaara, K.J., Thessler, S., Malik, R.N. and Tuomisto, H. 2005. Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sensing of Environment*. 97(1):39-51.
- Sandoval-Uribe, A., Healey, S.P., Moisen, G.G., Palafox-Rivas, R., González-Aguilar, E., Meneses-Tovar, C.L., Díaz-Ponce Dávalos, E.S. and Silva-Mascorro, S. 2008. Mexican forest inventory expands continental carbon monitoring. *EOS, Transactions American Geophysical Union*. 89(47):485.
- Schaepman-Strub, G., Schaepman, M.E. Painter, T.H. Dangel, S. and Martonchik, J.V. 2006. Reflectance quantities in optical remote sensing - definitions and case studies. *Remote Sensing of Environment*. 103: 27-42.
- Simula, M. 2009. Towards defining forest degradation: comparative analysis of existing definitions. *Forest Resources Assessment Programme. Working Paper 154*. Rome, Italy. 62 p.

- Sironen, S., Kangas, A., Maltamo, M. and Kalliovirta, J. 2008. Localization of growth estimates using non-parametric imputation methods. *Forest Ecology and Management*. 256:674–684.
- Sivanpillai, R., Smith, C.T., Srinivasan, R., Messina, M.G and Wu, X.B. 2005. Estimating regional forest cover in East Texas using Enhanced Thematic Mapper (ETM+) data. *Forest Ecology and Management*. 218: 342-352.
- Tokola, T. 2006. Europe. *In: Kangas, A. and Maltamo, M. 2006. Forest Inventory: Methodology and Applications*. Springer. Dordrecht, The Netherlands. pp. 295-308.
- Tomppo, E., Olsson, H., Stahl, G., Nilsson, M., Hagner, O. and Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*. 112:1982–1999.
- Toutin, T. 2004. Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing*. 25(10):1893-1924.
- Treitz, P. and Rogan, J. 2004. Remote sensing for mapping and monitoring land cover and land use change: an introduction. *Progress in Planning*. 61:269-279.
- Woodall, C.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D and Finley, A.O. 2009. An indicator of tree migration in forests of the eastern United States. *Forest Ecology and Management*. 257:1434-1444.
- Zhou, P., Luukknen, O., Tokola, T. and Nieminen, J. 2008. Effect of vegetation cover on soil erosion in a mountainous watershed. *Catena*. 75(3):319-325.
- Zon, R. 1910. The forest resources of the world. United States Department of Agriculture. Forest Service. Bulletin No. 83. Washington DC, USA, Government printing Office. 91 p.
- Zon, R. and Sparhawk, W.N. 1923. *Forest Resources of the World*. McGraw-Hill. New York, NY.

CAPÍTULO 2

CONSTRUCCIÓN DE MODELOS DE BIOMASA SUPERFICIAL CON TECNOLOGÍA DE PERCEPCIÓN REMOTA EN LA ZONA INTERTROPICAL EN MÉXICO ¹

Resumen

La estimación geoespacial de la biomasa arbórea aérea (B) tiene un papel fundamental en la generación de políticas de acción relacionadas con la mitigación del cambio climático, ya que el carbono retenido en la biomasa es vital para regular la temperatura del planeta. Este trabajo estudia la biomasa arbórea aérea mediante el uso de índices de vegetación normalizados basados en clorofila (infrarrojo cercano) y humedad (infrarrojo medio) construidos con datos MCD43A4 obtenidos del sensor MODerate-resolution Imaging Spectroradiometer (MODIS) y adicionalmente los campos continuos de vegetación MOD44B (VCF) en una región que cubre una parte del límite septentrional de la zona intertropical: San Luis Potosí, México. Se realizaron estimaciones de biomasa arbórea aérea (Mg/ha) con datos de árboles individuales provenientes del Inventario Nacional Forestal (INF) de México y el uso de ecuaciones alométricas reportadas en la literatura científica. Se probaron modelos lineales y no lineales (exponencial) para averiguar el potencial predictivo. Se encontró una correlación altamente significativa ($p = 0.01$) entre la biomasa arbórea aérea y todas las variables explicatorias sobresaliendo el NDVI62 sensible a la clorofila y al estrés hídrico de la vegetación. El modelo no lineal (NDVI62, VCF) presentó el índice de ajuste más alto (Pseudo - $r^2 = 0.77$) y el error más bajo (raíz del error cuadrático medio = 26.00 Mg/ha). Los coeficientes de correlación en la validación fueron similares en ambos modelos: lineal ($r = 0.87^{**}$) y no lineal ($r = 0.86^{**}$).

Palabras clave: MODIS, MCD43A4, MOD44B, inventario forestal, regresión.

¹ Publicado: Aguirre-Salado, C.A., Treviño-Garza, E.J., Aguirre-Calderón, O.A., Jiménez-Pérez, J., González-Tagle, M.A., Valdez-Lazalde, J.R., Miranda-Aragón, L. and Aguirre-Salado, A. 2012. Construction of aboveground biomass models with remote sensing technology in the intertropical zone in Mexico. *J. Geogr. Sci.* 22(4):669-680. DOI: 10.1007/s11442-012-0955-9.

Abstract

Geospatial estimation of aboveground biomass (AGB) has an important role for generating action policies related with mitigation of climate change, since carbon (C) retained in biomass is vital for regulating planet temperature. This work studies the AGB using both chlorophyll (red, near infrared) and moisture (middle infrared) based normalized vegetation indexes constructed with MCD43A4 MODerate-resolution Imaging Spectroradiometer (MODIS) data; additionally MOD44B vegetation continuous fields (VCF) were used. This study was located in a region that comprises a part of the upper limit of intertropical zone: San Luis Potosí, México. AGB estimations were made using both individual tree data from the National Forest Inventory of Mexico and allometric equations reported in scientific literature. Linear and nonlinear (exponential) models were probed to find out the predictive potential with satellite data. A highly significant correlation ($p = 0.01$) was found between all the explaining variables, excelling NDVI62 which is sensible to chlorophyll and moisture stress. The obtained nonlinear model (NDVI62, VCF) showed the highest index of fit (Pseudo – $r^2 = 0.77$) and the smallest error (root of squared mean error = 26.00 Mg/ha). In validation, correlation coefficients were similar in both models: linear ($r = 0.87^{**}$) and nonlinear ($r = 0.86^{**}$).

Key words: MODIS, MCD43A4, MOD44B, forest inventory, regression.

2.1. Introducción

El cambio climático es una realidad y cada día es más urgente la generación de alternativas que lo mitiguen (GOFC-GOLD, 2010). Todos los servicios ambientales que proveen los bosques son vitales para el funcionamiento adecuado del planeta. El crecimiento de los árboles acumula biomasa/carbono (Rodríguez-Laguna *et al.*, 2006; 2009) en sus tejidos vegetales extrayéndolo de la atmósfera y contribuyendo a mitigar el efecto invernadero. El conocer detalladamente la ubicación, la cantidad y la dinámica de los recursos forestales resulta por demás importante para la generación de políticas ambientales que coadyuven a frenar el cambio climático via prevención de la deforestación y fomento de la captura de carbono (DeFries, 2008; Andersson *et al.*, 2009).

Actualmente, una manera de extender y/o espacializar dichas estimaciones ha sido mediante el uso de técnicas de procesamiento e interpretación de imágenes captadas por sensores remotos (Xie *et al.*, 2008). Estas imágenes censan radiométricamente el área de interés en una malla (píxel) a diferentes longitudes de onda del espectro electromagnético y son sensibles a tres rasgos fundamentales de la vegetación: intensidad de verdor-clorofila (infrarrojo cercano) (Glenn *et al.*, 2008), contenido de humedad (infrarrojo de onda corta) (Gong *et al.*, 2003) y temperatura (infrarrojo térmico) (Amiri *et al.*, 2009). Esto ha permitido a los científicos determinar espaciotemporalmente (Marquez-Linares *et al.*, 2005; Antonio-Némiga *et al.*, 2008) los diferentes tipos de cobertura para la detección del cambio y también correlacionar determinadas variables biofísicas con su comportamiento espectral y espacio-temporal (Lu *et al.*, 2004; Powell *et al.*, 2010) y hacer de la teledetección una herramienta importante para el estudio de los recursos forestales (Xie *et al.*, 2008).

Existe una diversidad de trabajos basados en el uso de datos espectrales de baja/media resolución espacial con cubrimientos regionales/nacionales que sirven de base para la elaboración de diagnósticos relacionados con el monitoreo de la vegetación. Algunos han sido enfocados solamente en el mapeo cualitativo e incluso cuantitativo de la vegetación (Blackard *et al.*, 2008; Anaya *et al.*, 2009; Gallaun *et al.*, 2009; Coops *et al.*, 2009), mientras que otros han estudiado los cambios presentados a través del tiempo (Marquez-Linares *et al.*, 2005; Antonio-Némiga *et al.*, 2008; Baldi *et al.*, 2008). Más aún, el enfoque observado en la literatura ha sido el hecho de emplear diferentes fuentes de información complementaria a los datos espectrales para mejorar la precisión de las estimaciones *v.gr.* información climática (Reich *et al.*, 2008; Gallaun *et al.*, 2009). Por ejemplo, Anaya *et al.* (2009) publicaron un método para incrementar el nivel de detalle de las estimaciones de biomasa aérea a una escala regional para Colombia. Ellos usaron datos obtenidos del sensor MODIS, particularmente los campos continuos de vegetación MOD44B (VCF – vegetation continuous field, por sus siglas en inglés) como variable explicatoria de la densidad en bosques primarios y secundarios encontrando una relación exponencial; mientras que el índice de vegetación mejorado (EVI – enhanced vegetation index, por sus siglas en inglés) fue usado para describir el comportamiento de la densidad en pastizales encontrando una relación lineal. Blackard *et al.* (2008) mapearon la biomasa forestal de los Estados Unidos mediante la modelación de los datos del inventario forestal contra datos espectrales del satélite MODIS y variables auxiliares de tipo topográfico y climático entre otras.

El IFN de México organizado por la Comisión Nacional Forestal (CONAFOR) representa una vasta fuente de información levantada en campo sobre la situación de los recursos forestales de México. El reporte de las actuales estimaciones espaciales de los almacenes de biomasa/carbono a nivel nacional (GOFC-GOLD, 2010) se realizó con datos de densidad forestal promedio asignados a polígonos de los mapas de vegetación y uso del suelo

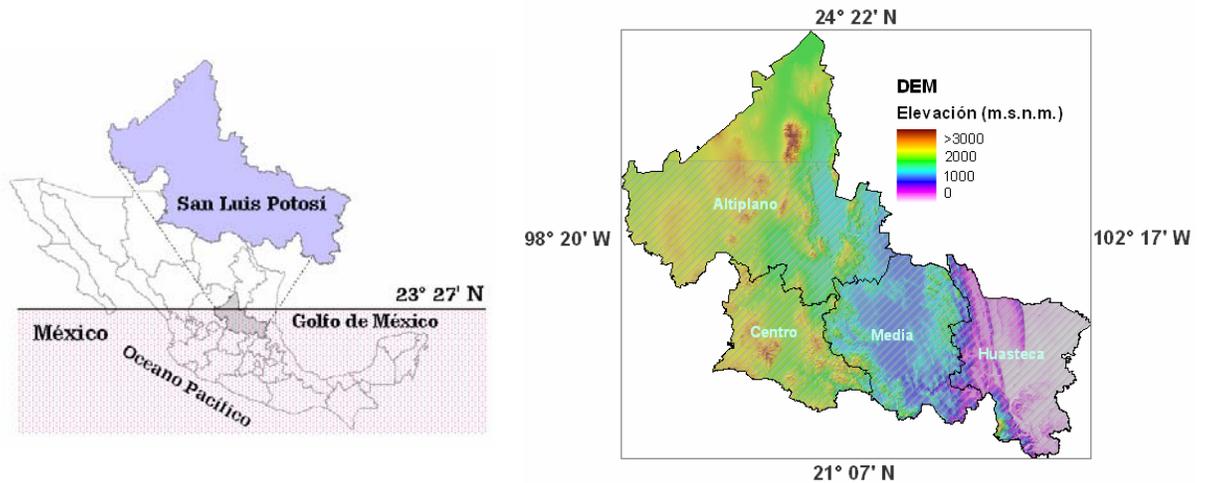
serie IV (INEGI, 2007), los cuales fueron generados mediante técnicas de fotointerpretación visual de imágenes satelitales Landsat y SPOT. Aunque dicha metodología es precisa, no aprovecha la riqueza espectral de los datos captados por el sensor remoto para la generación de estimaciones continuas, píxel a píxel, como lo requieren variables numéricas que caracterizan la densidad forestal. En este sentido, los objetivos de este trabajo fueron: (1) averiguar la relación existente entre la biomasa arbórea aérea (Mg/ha) y diferentes índices espectrales normalizados de vegetación sensibles a clorofila y humedad, (2) construir modelos (lineales y no lineales) para la estimación de la biomasa arbórea aérea, (3) generar mapas de biomasa arborea aérea con los modelos seleccionados, (4) validar las estimaciones con datos externos al modelo y (5) comparar las estimaciones de los dos modelos en un transecto este-oeste en el área de estudio.

2.2. Materiales y Métodos

2.2.1. Área de Estudio

San Luis Potosí se localiza en la porción centro norte de México y tiene una superficie de 6'116,360.91 ha. Se ubica entre los paralelos 21° 07' y 24° 22' latitud norte y, 98° 20' y 102° 17' longitud oeste (Figura 2.1). El estado se encuentra en el límite norte de la zona intertropical, cuenta con una riqueza de paisajes y condiciones de vegetación que puede ser comprendida analizando las zonas geográficas que lo componen: (1) Altiplano, (2) Centro, (3) Media y (4) Huasteca (INEGI, 2002). Esta distribución de ecosistemas obedece a la conformación orográfica, al patrón de distribución de precipitación que proviene del Golfo de México y a la relación inversa existente entre la temperatura y la altitud propia en estas latitudes. Los vientos cargados de humedad que precipitan en la Sierra Madre Oriental (Zona Media y Huasteca) y aunado al gradiente altitudinal, propician formaciones vegetales de clima tropical (selva) y templado (bosques); mientras que los vientos avanzan continente adentro, la

vegetación va cambiando progresivamente hasta observar vegetación propia de climas áridos *v.gr.* matorral (Figura 2.2).



A)

B)

Figura 2.1. Localización del área de estudio. A) Contexto nacional y zona intertropical (hachurado) (Límite 23° 27' N) y B) Zonas geográficas del Estado.

2.2.2. Datos de Campo

Después de realizar un análisis de control de calidad en los datos, se seleccionaron 235 conglomerados de muestreo de la colección del IFN levantados en el Estado de San Luis Potosí durante el periodo 2005-2007 (Figura 2.2) donde la información dasométrica incluyó básicamente diámetro normal, altura, especie. Cada uno de los conglomerados se compone de 4 sitios de muestreo de 400 m² sumando 1600 m² de área (CONAFOR, 2010). Las coordenadas de los sitios de muestreo se encontraban bajo la proyección Plate-Carrée y fueron proyectados al sistema de coordenadas UTM-14n datum WGS84. Para el cálculo de la biomasa aérea de cada una de las 364 especies encontradas en la base de datos del IFN para el Estado de San Luis Potosí, se emplearon 25 ecuaciones diferentes reportadas en Návar (2009). Las formas matemáticas de las ecuaciones varían para cada autor y especie *v.gr.* lineal, exponencial, potencial; en las cuales se ingresa el diámetro normal (cm) y la

altura del individuo (m) para obtener la biomasa total individual (kg). Los datos dasométricos individuales (20731 árboles) fueron sumados por parcela y estimados a la hectárea para ser relacionados con las variables espectrales obtenidas de la imagen de satélite. El procesamiento de los datos se llevó a cabo en el software SAS (SAS Institute Inc., 2004).

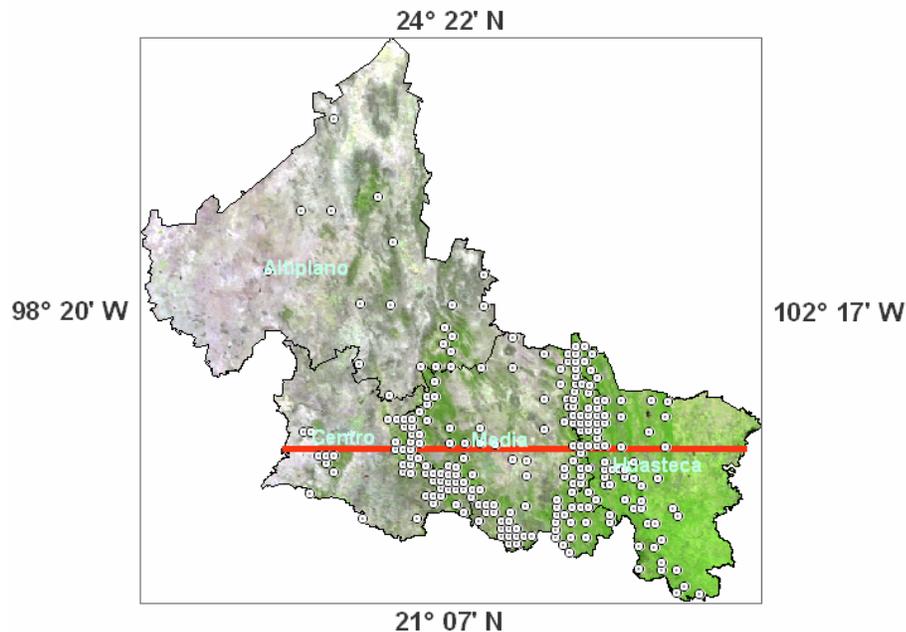


Figura 2.2. Distribución espacial de los sitios de muestreo del Inventario Nacional Forestal para el Estado de San Luis Potosí y transecto para comparación de estimaciones.

2.2.3. Datos Espectrales

El tamaño del área de estudio y la escala deseada de análisis resultan clave al momento de seleccionar un sensor remoto como fuente de datos para estudiar los recursos forestales (Wu y Zhao-Liang, 2009). En la zona intertropical existen limitaciones visuales ocasionadas por la nubosidad lo que dificulta conseguir imágenes de alta/media resolución espacial como Landsat que cubran grandes áreas y con la misma fecha de toma en todas las escenas. Para que las estimaciones geospaciales resulten espacialmente coherentes,

los datos espectrales deben ser de fecha similar de lo contrario se tiene que seleccionar un sensor de menor resolución espacial, por ejemplo MODIS. No todos los productos MODIS se encuentran a 250 m de resolución espacial. Dado que los objetivos del trabajo fueron examinar el comportamiento de índices normalizados con bandas espectrales específicas y radiométricamente corregidas, se seleccionó el producto MCD43A4. Este producto se obtiene del sensor MODIS instalado en los satélites Aqua y Terra; se le denomina reflectancia (ρ) ajustada al nadir (MCD43A4) usando una función de distribución de reflectancia bidireccional; integra datos multifecha tomados durante periodos de 16 días y se encuentra corregido atmosféricamente y libre de nubes. Tiene una resolución espacial de 500 m y una resolución espectral de siete bandas distribuidas en las siguientes longitudes de onda: 1. Rojo (620 - 670 nm), 2. Infrarrojo cercano (841 - 876 nm), 3. Azul (459 - 479 nm), 4. Verde (545 - 565 nm), 5. Infrarrojo medio₁ (1230 - 1250 nm), 6. Infrarrojo medio₂ (1628 - 1652 nm), 7. Infrarrojo medio₃ (2105 - 2155 nm) (Bhaduri *et al.*, 2010).

Para el análisis se usaron datos del periodo 11 al 26 de febrero de 2006 (día juliano 42 al 57) los cuales fueron obtenidos de la interface denominada Modis Reprojection Tool (MRTWeb) que combina la posibilidad de búsqueda de información con capacidades de preprocesamiento de datos espectrales (<http://lpdaac.usgs.gov/datapool/datapool.asp>) (Daucsavage *et al.*, 2010). Estos datos fueron obtenidos para las coordenadas extremas del área de interés y fueron re proyectados al sistema UTM-14N. Se construyeron los índices normalizados de vegetación: 1) NDVI₂₁ tradicional, calculado como NDVI₂₁ tradicional = $(\rho_{irc} - \rho_{rojo}) / (\rho_{irc} + \rho_{rojo})$; 2) NDVI₂₄, calculado como NDVI₂₄ = $(\rho_{irc} - \rho_{verde}) / (\rho_{irc} + \rho_{verde})$; 3) NDVI₆₁, calculado como NDVI₆₁ = $(\rho_{irm2} - \rho_{rojo}) / (\rho_{irm2} + \rho_{rojo})$; 4) NDVI₆₂, calculado como NDVI₆₂ = $(\rho_{irm2} - \rho_{irc}) / (\rho_{irm2} + \rho_{irc})$; 5) NDVI₆₄, calculado como NDVI₆₄ = $(\rho_{irm2} - \rho_{verde}) / (\rho_{irm2} + \rho_{verde})$. La banda 3 sensible en la porción azul no fue usada debido a que existen fallas en su calibración pudiendo generar errores en las estimaciones (Gallaun *et al.*, 2010).

También se empleó el producto MODIS denominado campos continuos de vegetación versión 3, colección 3 (MOD44B) (VCF, por sus siglas en inglés) (500 m de resolución espacial) que indica la cobertura proporcional (%) para vegetación arbórea, vegetación herbácea y suelo desnudo y se encuentra en tres capas de información, respectivamente (Hansen *et al.*, 2003). En este trabajo se emplearon los datos VCF correspondientes a la cobertura arbórea.

2.2.4. Procesamiento de la Información

Inicialmente se efectuó un análisis de correlación de Pearson (r) entre la biomasa y los índices de vegetación para conocer el grado y el sentido de asociación de los datos. Posteriormente se ajustaron modelos de regresión lineal y no lineal para determinar el que mejor se ajustara a la distribución espectral de los datos. Los modelos probados fueron 1) lineal $B = \beta_0 + \beta_1(X_1) + \varepsilon$; y 2) exponencial $B = \beta_0 e^{\beta_1(X_1)} + \varepsilon$, donde B = biomasa arbórea aérea (Mg/ha); X_i = índice de vegetación y/o VCF. Como indicadores de ajuste se usó el coeficiente de determinación (r^2) calculado con la fórmula

$$r^2 = \frac{SCR}{SCT} = \frac{SCR}{SCR + SCE} \text{ y la pseudo } r^2 \text{ para modelos no lineales (SAS Inc.,}$$

2008) con la fórmula $Pseudo\ r^2 = 1 - \left(\frac{SCE}{SCT_{corregida}} \right)$ donde SCR = suma de

cuadrados de regresión; SCT = suma de cuadrados totales; SCE = suma de cuadrados del error.

2.2.5. Validación

Se seleccionó una muestra aleatoria del 10% de las parcelas de muestreo que no se incluyeron en la construcción del modelo para validar la metodología. Se evaluó el error absoluto (RECM) y relativo (RECM %) con las

fórmulas: $RECM = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$ y $RECM \% = \frac{RECM}{\bar{y}} * 100$, donde \hat{y}_i = valor

estimado; y_i = valor observado; \bar{y} = promedio de las estimaciones; n = número de observaciones. Además, se usó el coeficiente de correlación de Pearson (r) como una medida de ajuste lineal entre las predicciones y observaciones para ambos métodos.

2.3. Resultados y Discusión

2.3.1. Relación entre las Variables de Densidad Forestal y los Índices de Vegetación

La biomasa arbórea aérea (B) mostró una correlación positiva ($\alpha=0.01$) en casi todos los índices de vegetación excepto para el NDVI62 (Tabla 2.1, Figura 2.3). El NDVI21 (NDVI tradicional) y los VCF fueron las variables que mostraron la correlación positiva más alta, ya que se encuentran íntimamente relacionados a la cantidad de clorofila de la vegetación sana y a mayores valores denotan un incremento de la densidad forestal (Glenn et al., 2008). Mientras que el NDVI62, fue la variable que mostró la correlación más alta (-0.77) confirmando los resultados de Gong *et al.* (2003) y Aguirre-Salado *et al.* (2009) donde se indica que la densidad de bosques se explica mejor por índices sensibles a la humedad-clorofila, que aquellos que solamente se basan en la radiometría de la clorofila.

El signo negativo encontrado en la correlación de la biomasa arbórea aérea vs NDVI62 se puede explicar analizando el orden de las bandas espectrales (en el numerador) que lo calculan, primero la banda 6 (Infrarrojo medio₂, 1628 – 1652 nm) y después la banda 2 (Infrarrojo cercano (841 - 876 nm). En la firma espectral de la vegetación sana (Gao, 2009), el infrarrojo medio registra una baja reflectancia (asociada a la humedad) mientras que el infrarrojo

cercano se caracteriza por registrar una alta reflectancia (asociada a la clorofila) donde el cociente normalizado genera siempre un valor negativo que describe al índice de estrés hídrico (Rock *et al.*, 1986).

Tabla 2.1. Coeficientes de correlación de Pearson (r) obtenidos para biomasa y las variables satelitales.

	NDVI21	NDVI24	NDVI61	NDVI62	NDVI64	VCF
B	0.74	0.73	0.65	-0.77	0.39	0.75

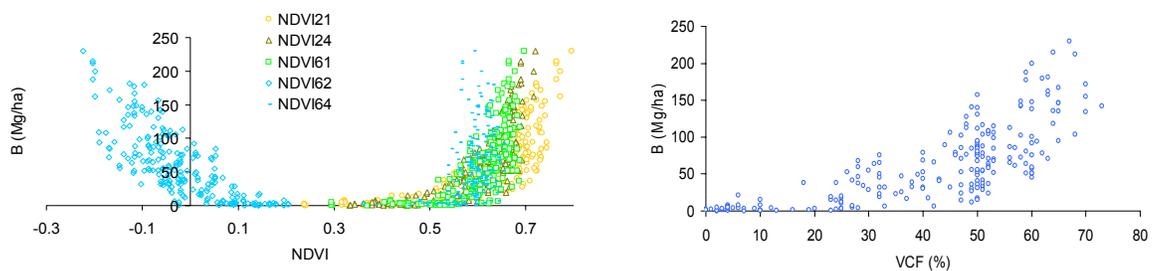


Figura 2.3. Distribución de la biomasa arbórea aérea respecto a los índices de vegetación construidos y los VCF.

2.3.2. Modelos de Regresión

En la Tabla 2.2. se presentan los modelos probados que relacionan la biomasa con las variables explicatorias (índices de vegetación y VCF). En la Ecuación 1 se probaron todas las variables explicatorias y las únicas que demostraron ser altamente significativas fueron el NDVI62 y los VCF ($p = 0.01$). No obstante, también se generaron otras combinaciones de variables explicatorias en los modelos probados y se evaluaron los índices de ajuste. La Ecuación 2 fue la que linealmente describe mejor la densidad forestal ya que recoge la información espectral relacionada con la humedad (infrarrojo medio), clorofila (infrarrojo cercano) y toma en cuenta el aporte al modelo de los VCF, variable sintética validada a nivel mundial para modelar la densidad forestal (Hansen *et al.*, 2003). En la Tabla 2.2 se muestra la evidente mejora en los indicadores de

ajuste de los modelos no lineales (Tipo exponencial) respecto a los modelos lineales, alcanzando el máximo potencial con la variable NDVI62 y los VCF; mientras que en la Figura 2.4, se puede apreciar la distribución de los datos en el espacio tridimensional. No obstante, el NDVI21 individual mostró una interesante capacidad predictiva.

Tabla 2.2. Modelos probados

Modelo	Ecuación	r ² o Pseudo - r ²	Raíz de SCE
B = -53.87 -985.27(NDVI21) -903.44(NDVI24) + 1166.45(NDVI61) -1451.24(NDVI62**) + 819.02(NDVI64) + 0.95(VCF**)	1	0.7240	28.29
B = 15.86 -266.28(NDVI62**) + 1.02(VCF**)	2	0.6392	32.35
B = -78.49 +155.15(NDVI21**) + 1.16(VCF**)	3	0.5778	35.00
B = -125.64 + 312.15 (NDVI21**)	4	0.5521	36.04
B = -191.83 + 430.18 (NDVI24**)	5	0.5366	36.66
B = -214.26 + 461.23 (NDVI61**)	6	0.4146	41.21
B = 54.72 – 425.50 (NDVI62**)	7	0.5944	34.30
B = -208.05 + 463.23 (NDVI64**)	8	0.1495	49.68
B = -24.17 + 2.12 (VCF**)	9	0.5582	35.84
B = 0.1610e ^{NDVI21*(9.1299)}	10	0.7328	27.97
B = 0.0148e ^{NDVI24*(13.2918)}	11	0.7168	28.73
B = 0.0136e ^{NDVI61*(13.5128)}	12	0.5453	36.40
B = 40.6081e ^{NDVI62*(-6.8794)}	13	0.5366	36.95
B = 0.5135e ^{NDVI64*(8.1424)}	14	0.1435	49.97
B = 7.3800e ^{VCF*(0.0453)}	15	0.6339	32.67
B = 0.2905e ^{NDVI21*(6.8344) + VCF*(0.0185)}	16	0.7381	27.62
B = 13.8926e ^{NDVI62*(-4.0086) + VCF*(0.0275)}	17	0.7691	26.00

B: Biomasa total (Mg/ha); NDVI: Índice de Vegetación de Diferencias Normalizadas construido con la bandas de la imagen; P value: valor de probabilidad, * = Significancia al 0.05%; ** = Significancia al 0.01%.

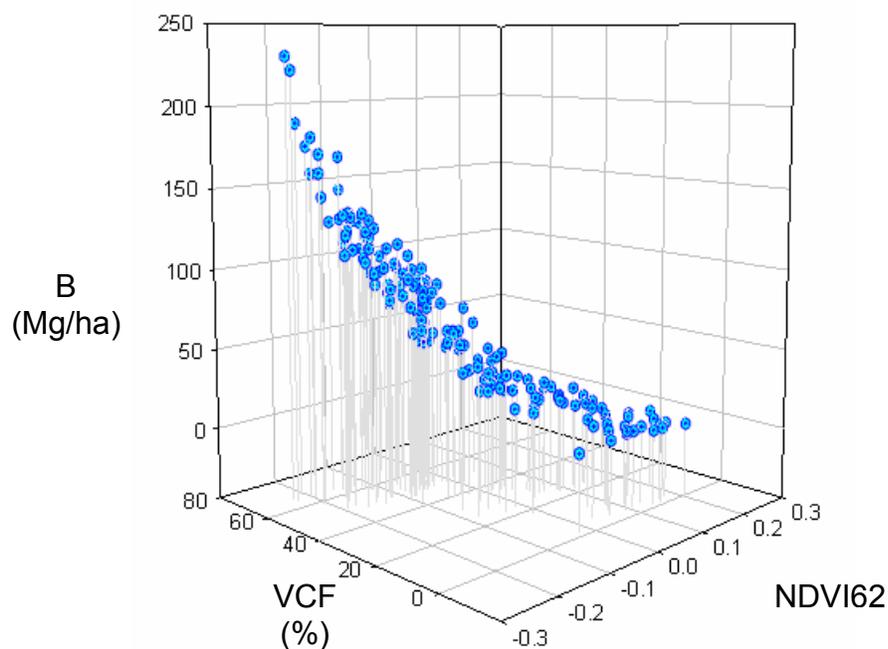


Figura 2.4. Biomasa en función del NDVI62 y los VCF.

2.3.3. Mapas de Densidad

La estimación píxel a píxel de la biomasa arbórea aérea se realizó con las Ecuaciones 2 (lineal) y 17 (exponencial). Este método aprovecha la sensibilidad radiométrica y fineza espectral de los datos captados por los sensores remotos y resulta valioso para describir la variabilidad espacial de la biomasa arbórea aérea para el área de estudio (Figura 2.5). Los valores más pequeños (<5 Mg/ha) corresponden a la región árida de San Luis Potosí, que incluye la mayor parte del Altiplano, y unas porciones de la zona Centro y Media. Por otro lado, los valores más altos (>150 Mg/ha) corresponden a la porción templada y tropical, zona Media y Huasteca. Las estimaciones de biomasa coinciden por las reportadas por Anaya *et al.*, (2009) que van de los <5 Mg/ha en las zonas áridas y pastizales hasta >150 Mg/ha en las zonas tropicales. Si se desean transformar las estimaciones a carbono simplemente bastará con multiplicar la biomasa (Mg/ha) por el factor de 0.5 (Barrio-Anta *et al.*, 2006).

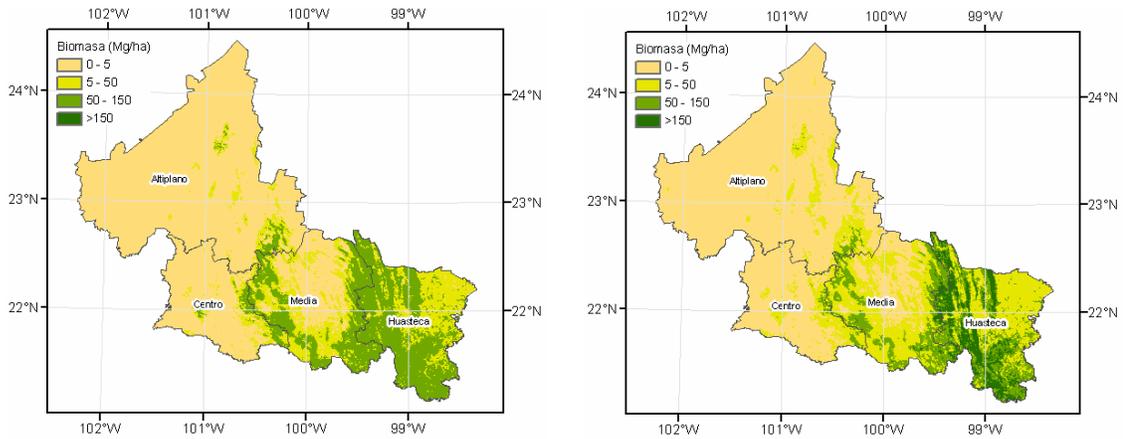


Figura 2.5. Biomasa arbórea aérea estimada mediante los modelos seleccionados y las variables NDVI62 y VCF. A) Lineal, B) Exponencial.

En la Figura 2.6. se aprecian las diferencias en las estimaciones realizadas por los modelos lineal vs no lineal a lo largo del perfil este-oeste trazado en área de estudio. La tendencia coincide con las diferentes zonas boscosas de la zona Centro, Media y Huasteca. Se aprecian los valores negativos para el modelo de regresión en las zonas áridas mientras que para las zonas con mayor densidad el modelo exponencial realiza estimaciones más grandes.

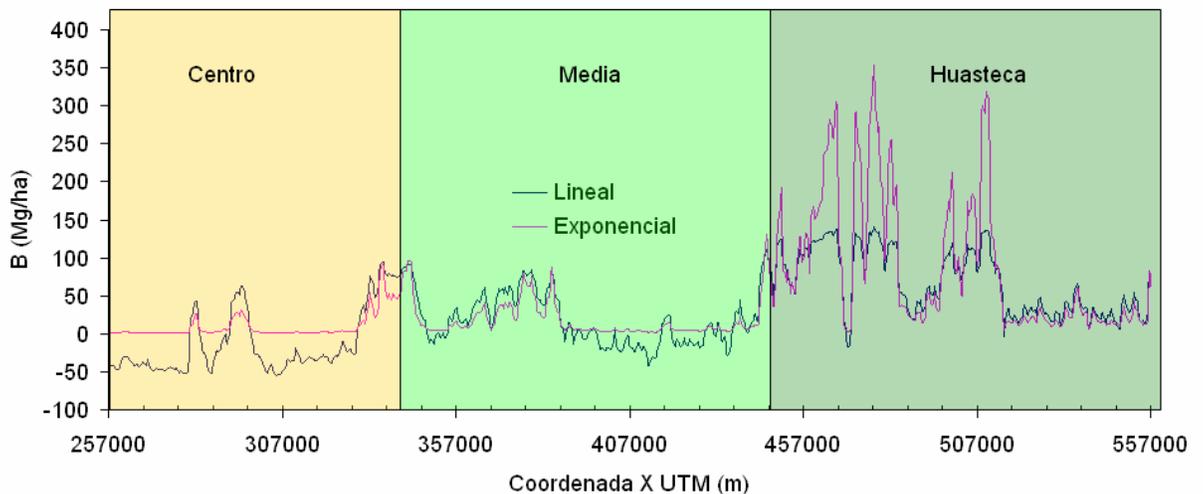


Figura 2.6. Comparación de estimaciones (píxel) en un transecto dirección oeste-este.

Rodríguez-Laguna *et al.*, (2006) estimaron en 56.7 Mg/ha el contenido de carbono arbóreo en un bosque de niebla (Biosfera El Cielo) similar al que existe bajo las condiciones climáticas y ecológicas en los límites de la zona Media y Huasteca de San Luis Potosí. En este estudio, para el bosque de niebla (bosque mesófilo de montaña: Serie IV de Vegetación y Uso de Suelo de INEGI) se obtuvo un promedio estimado con el modelo lineal de 108.00 B Mg/ha (54.00 C Mg/ha), y con el modelo no lineal de 133.69 B Mg/ha (66.84 C Mg/ha). Esto confirma el potencial predictivo del presente método, más aún con la ventaja de recoger la variabilidad espacial de la biomasa arbórea aérea, píxel a píxel (Figura 2.7).

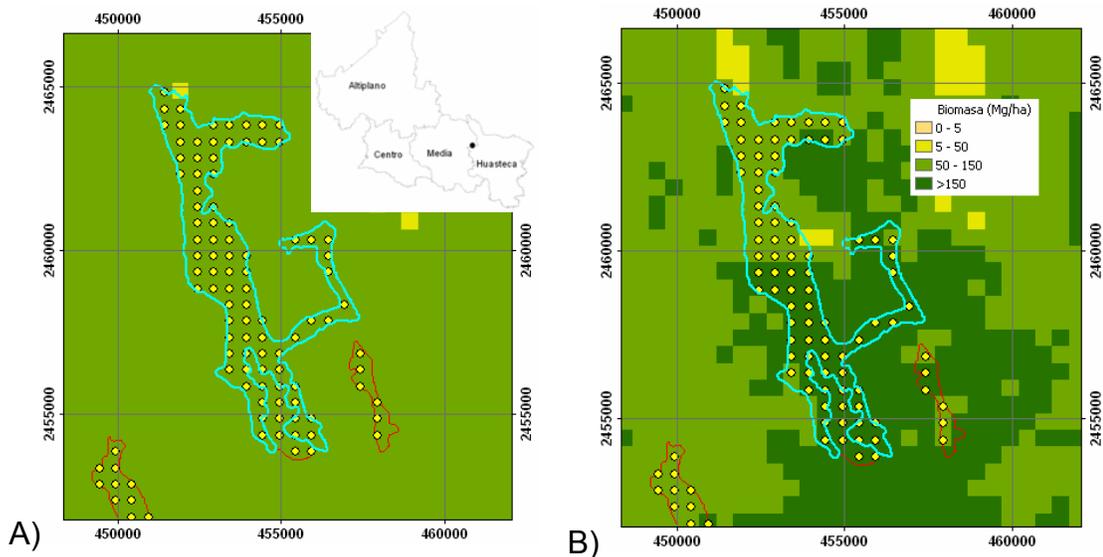


Figura 2.7. Biomasa arbórea aérea píxel a píxel para bosque mesófilo de montaña. A) Modelo Lineal. B) Modelo no lineal.

2.3.4. Validación

Los errores del modelo lineal (RECM = 21.91 Mg/ha, RECM% = 36.81 %) obtenidos con la muestra independiente de validación fueron menores respecto a los del modelo exponencial (RECM = 22.24 Mg/ha, RECM% = 39.56 %). Esto probablemente sucedió así porque el modelo exponencial arroja estimaciones más contrastantes respecto al modelo lineal. El modelo lineal tiende a realizar

estimaciones negativas en píxeles con valores bajos de biomasa, mientras que el modelo exponencial está matemáticamente definido para que la estimación se vuelva asintótica al eje X y nunca sea negativa. Por otro lado, ambos modelos lineal ($r = 0.87^{**}$) y no lineal ($r = 0.86^{**}$) fueron validados satisfactoriamente al comparar las estimaciones con datos externos a la muestra para entrenar los modelos (Figura 8). Estos resultados son comparables con los encontrados en la literatura; Zheng *et al.* (2004) con datos Landsat, como variable la biomasa forestal, y como método estadístico el de regresión obtuvieron una correlación entre valores observados y estimados de 0.67; Blackard *et al.* (2008) con datos MODIS, mediciones de biomasa, y como método el de árboles de regresión, obtuvieron correlaciones entre valores observados y estimados (10-fold cross validation) desde 0.31^* hasta 0.92^{**} .

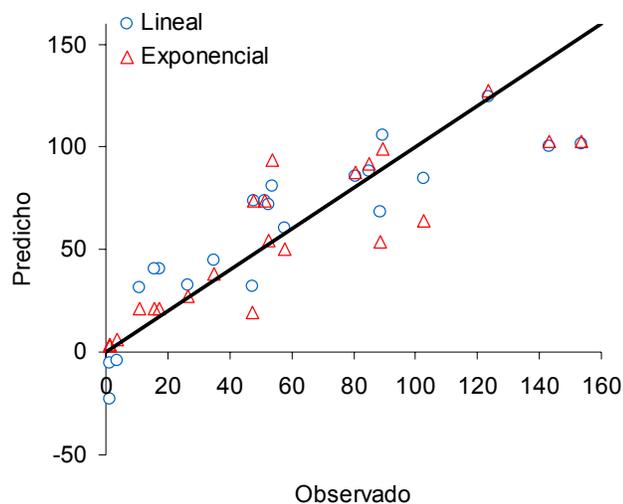


Figura 2.8. Valores de biomasa arbórea aérea observados vs predichos.

2.4. Conclusiones

Todos los índices normalizados de vegetación construidos con datos de reflectancia ajustada al nadir (MCD43A4) y los campos continuos de vegetación (MOD44B) obtenidos del sensor MODIS tuvieron un potencial predictivo altamente significativo ($p = 0.01$). No obstante la máxima capacidad la presentó el NDVI62, que involucra las bandas sensibles en el infrarrojo medio (Banda 6:

1628 – 1652 nm) e infrarrojo cercano (Banda 2: 841 - 876 nm) y los VCF. El modelo no lineal fue el que presentó el índice de ajuste más alto con los datos de la muestra ($Pseudo-r^2 = 0.7691$; Raíz de SCE = 26.00 Mg/ha), no obstante puede generar sobreestimaciones. Con los datos de validación, ambos métodos estadísticos (lineal y no lineal) tuvieron el mismo potencial predictivo, aunque en el análisis del perfil generado con el transecto experimental, las diferencias entre las predicciones por el modelo lineal y el no lineal fueron evidentes. El modelo lineal presentó estimaciones más conservadoras de la biomasa arbórea aérea respecto al modelo no lineal. Con los datos del IFN, el modelo lineal y el modelo no lineal se estima que la biomasa arbórea aérea total en las zonas arboladas del Estado de San Luis Potosí (aprox. 3.9 millones de ha) es de 93.8 millones de Mg (93,808,669.39 Mg), mientras que con el modelo no lineal es de 116 millones de Mg (116,255,822.93 Mg).

2.5. Literatura Citada

- Antonio-Némiga, X., Treviño-Garza, E.J. y Jurado-Ybarra, E. 2008. Fragmentación forestal en la subcuenca del río Pilon: diagnóstico y prioridades. *Madera y Bosques*. 14(3):5-23.
- Aguirre-Salado, C.A., Valdez-Lazalde, J.R., Ángeles-Pérez, G., De los Santos-Posadas, H.M., Happanen, R., Aguirre-Salado, A.I. 2009. Mapeo de carbono arbóreo aéreo en bosques manejados de *Pinus patula* en Hidalgo, México. *Agrociencia* 43:209-220.
- Amiri, R., Weng, Q., Alimohamadi, A. and Alavipanah, S.K. 2009. Spatial-temporal dynamics of land surface temperature in relation to fractional vegetation cover and land use/cover in the Tabriz urban area, Iran. *Remote Sensing of Environment*. 113:2606–2617.
- Anaya, J.A., Chuvieco, E. and Palacios-Orueta, A. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management*. 257:1237–1246.

- Andersson, K., Evans, T.P. and Richards, K.R. 2009. National forest carbon inventories: policy needs and assessment capacity. *Climatic Change*. 93:69-101.
- Barrio-Anta, M., Balboa-Murias, M.A., Castedo-Dorado, F., Diéguez-Aranda, U., Álvarez-González, J.G. 2006. An ecoregional model for estimating volume, biomass and carbon pools in maritime pine stands in Galicia (northwestern Spain). *Forest Ecology and Management*. 223:24-34.
- Bhaduri, K., Das, K. and Votava, P. 2010. Distributed anomaly detection using satellite data from multiple modalities. *Proceedings of The 2010 Conference on Intelligent Data Understanding (CIDU-NASA)*. San Francisco Bay Area. Oct 5-7. pp.109-123.
- Blackard, J.A., Finco, M.V., Helmer, E.H., Holden, E.H., Hoppus, M.L., Jacobs, D.M., Lister, A.J., Moisen, G.G., Nelson, M.D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D.L., Winterberger, K.C., Brandeis, T.J., Czaplewski, R.L., McRoberts, R.E., Patterson, P.L. and Tymcio, R.P. 2008. Mapping U.S. forest biomass using nation wide forest inventory data and moderate resolution information. *Remote Sensing of Environment*. 112:1658-1677.
- CONAFOR. 2010. Manual de Procedimientos para el muestreo de campo. Remuestreo-2010. Inventario Nacional Forestal y de Suelos. Comisión Nacional Forestal. Secretaría del Medio Ambiente y Recursos Naturales. Zapopan, Jalisco, México. 140 p.
- Coops, N.C., Ferster, C.J., Waring, R.H. and Nightingale, J. 2009. Comparison of three models for predicting gross primary production across and within forested ecoregions in the contiguous United States. *Remote Sensing of Environment*. 113:680–690.
- Daucsavage, J. Kaminski, M., Ramachandran, B., Jenkerson, C., Sprenger, K., Faust, R. and Rockvam, T. 2010. ASTER and MODIS Land Data Management at the Land Processes, and National Snow and Ice Data Centers. In: Ramachandran, B., Justice, C.O., and Abrams, M.J. *Land Remote Sensing and Global Environmental Change*. NASA Earth

- Observing System and the Science of ASTER and MODIS. Springer. pp. 167-182.
- DeFries, R. 2008. Terrestrial vegetation in the coupled human-earth system: contributions of remote sensing. *Annual Review on Environmental Resources*. 33:369-390.
- Gao, J. 2009. *Digital Analysis of Remotely Sensed Imagery*. McGraw Hill. pp. 18. ISBN:978-0-07-160466-6.
- Gallaun, H., Zanchi, G., Nabuurs, G.J., Hengeveld, G., Schardt, M. and Verkerk, P. 2009. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*. 260(3):252-261.
- Glenn, E.P., Huete, A.R. Nagler, P.L. and Nelson, S.G. 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. *Sensors* 8: 2136-2160.
- GOFC-GOLD. 2010. A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals caused by deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. GOFC-GOLD Report version COP16-1. GOFC-GOLD Project Office, Natural Resources Canada, Alberta, Canada. 203 p.
- Gong, P., Pu, R., Biging, G.S. Larrieu, M.R. 2003. Estimation of forest leaf area index using vegetation indices derived from Hyperion Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing* 41(6):1355-1362.
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C., Sohlberg, R.A., 2003. Global percent tree cover at a spatial resolution of 500meters: first results of the MODIS Vegetation Continuous Fields algorithm. *Earth Interactions* 7:1–15.
- INEGI. 2002. *Síntesis Geográfica del Estado de San Luis Potosí*. Instituto Nacional de Estadística y Geografía e Informática. México.

- INEGI. 2007. Vegetación y Uso de Suelo Serie IV. Instituto Nacional de Estadística y Geografía e Informática. México.
- Lu, D., Mausel, P., Brondizio, E. and Moran, E. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*. 198:149-167.
- Márquez-Linares, M.A., Treviño-Garza, E.J. and Jurado-Ybarra, E. 2005. Reemplazo de áreas arboladas por chaparrales y comunidades herbáceas en el periodo 1970-2000 en una microcuenca de Durango, México. *Investigaciones Geográficas*. 58:54-65.
- Návar, J. 2009. Biomass component equations for latin american species and groups of species. *Annals of Forest Science*. 66(2):208.
- Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B. and Ohmann, J.L. 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*. 114:1053–1068.
- Reich, R.M. Aguirre-Bravo, C., Bravo, V.A. 2008. New approach for modeling climatic data with applications in modeling tree species distributions in the states of Jalisco and Colima. *Journal of Arid Environments* 72(7):1343-1357.
- Rock, B.N., Vogelmann, J.E., Williams, D.L., Vogelmann, A.F. and Hoshizaki, T. 1986. Remote detection of forest damage. *Bioscience*. 36:439-445.
- Rodríguez-Laguna, R., Jiménez-Pérez, J., Aguirre-Calderón, O.A. y Treviño-Garza, E.J. 2006. Estimación del carbono almacenado en un bosque de niebla en Tamaulipas, México. *Ciencia UANL*. 9(2):179-188.
- Rodríguez-Laguna, R., Jiménez-Pérez, J., Aguirre-Calderón, O.A., Treviño-Garza, E.J. y Razo-Zárte, R. 2009. Estimación de carbono almacenado en el bosque de pino-encino en la Reserva de la Biosfera "El cielo", Tamaulipas, México. *Ra Ximhai*. 5(3):317-327.
- SAS Institute Inc. 2004. SAS/STAT 9.1 User's Guide. SAS Institute Inc., Cary, NC, USA. 5121 p.

- Wu, H. and Zhao-Liang, L. 2009. Scale issues in Remote Sensing: a review on analysis processing and modeling. *Sensors* 9:1768-1793.
- Xie, Y., Sha, Z. and Yu, M. 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*. 1(1): 9-23.
- Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J, and Ryu, S.R. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*. 93:402-411.

CAPÍTULO 3
MAPEO DE LA COBERTURA FORESTAL EN UNA PORCIÓN NORTE
CENTRO DE MÉXICO: UNA COMPARACIÓN DE MÉTODOS DE
PROCESAMIENTO DIGITAL DE IMÁGENES¹

Resumen

La vegetación natural es un atributo ambiental que varía de manera continua en el paisaje y se encuentra mezclado a diferentes niveles. A pesar de que la mayoría de los mapas existentes de vegetación describen tal realidad de manera discreta, existen métodos avanzados que pueden caracterizar la cobertura fraccional de los tipos de vegetación dentro de cada píxel. Este capítulo se enfoca en la comparación de métodos para la estimación de coberturas forestales a nivel subpixel probados en una porción del límite septentrional de la zona intertropical en México usando datos espectrales de resolución espacial moderada (MCD43A4) combinados con información obtenida del Inventario Nacional Forestal (INF) Mexicano y del mapa de vegetación y uso de suelo serie IV (2007) generado por el Instituto Nacional de Geografía y Estadística. Se probaron tres métodos para determinar la cobertura fraccional dentro de los píxeles: 1) probabilidad posterior bayesiana, 2) el k vecino más cercano (FkNN, por sus siglas en inglés) en su versión difusa y 3) análisis de mezcla espectral (LMA, por sus siglas en inglés). Tales métodos fueron entrenados con datos espectrales de clases puras para bosque y matorral. Las estimaciones fueron validadas con información obtenida del procesamiento de imágenes de una resolución más detallada. La validación muestra que el LMA ($r = 0.81$, raíz del error cuadrático medio RMSE = 0.14) es más apropiado para describir la variación dentro del píxel para las clases empleadas.

Palabras clave: MODIS, análisis de mezclas, inventario forestal, biomasa.

¹ Publicado: Aguirre-Salado, C.A., Treviño-Garza, E.J., Aguirre-Calderón, O.A., Jiménez-Pérez, J., González-Tagle, M.A., Miranda Aragón, L., Valdez-Lazalde, J.R., Aguirre-Salado, A.I., Sánchez-Díaz, G. 2012. Forest cover mapping in a north central portion of Mexico: a comparison of digital image processing methods. *GISci. Remote Sens.* 49(6), <http://dx.doi.org/10.2747/1548-1603.49.6.1>.

Abstract

Natural vegetation is an environmental feature that varies continuously through the landscape and is represented in different levels of purity. Despite of the most of existent vegetation maps describe such reality in a discrete way, there are advanced approaches that characterize fractional cover of selected thematic classes within a pixel. Therefore, this chapter focus in the comparison of methods for estimating forest cover at subpixel level in a portion of the septentrional border of the intertropical zone in Mexico using medium spatial resolution data (MCD43A4) combined with information obtained from National Forest Inventory and vegetation and land use map (serie IV, 2007) generated by the Instituto Nacional de Geografía y Estadística. Three methods were probed for determining fractional cover within pixels: 1) Bayesian posterior probability, 2) the Fuzzy K nearest neighbor and 3) the Linear Mixture Analysis. Such methods were applied with spectral signatures of pure classes (endmember) for forest and shrubland. The estimations were validated with information obtained via processing of a more spatially detailed image. The validation stage shows that LMA is more suitable to describe the intrapixel variation of the selected thematic classes.

Key words: MODIS, mixture analysis, forest inventory, biomass.

3.1. Introducción

El estudio de la vegetación forestal mediante técnicas de procesamiento de imágenes de satélite es una tarea común en el monitoreo de los recursos forestales. Normalmente los métodos tradicionales de procesamiento se enfocan en determinar la clase de vegetación de un píxel. Aunque esto teóricamente funciona bien para coberturas que son espacialmente uniformes sobre escalas amplias, los resultados no son precisos para tipos de vegetación o usos del suelo que no cubren completamente el píxel. Este problema se aborda mejor en el enfoque difuso con el que es posible determinar la proporción de las coberturas que ocupan cada píxel (Gao, 2009).

Existen varios métodos que generan estimaciones no definitivas sobre la identidad de un píxel. Ellos determinan probabilidades del grado de pertenencia del píxel a cada clase y pueden ser interpretados como proporciones de mezclas a nivel subpíxel. El análisis de mezcla espectral es un modelo físicamente determinístico que puede ser empleado para abordar el problema del píxel mezclado (Myint y Okin, 2009); asume que la señal registrada en un determinado píxel es una combinación lineal del espectro, que puede ser separada en componentes llamados endmembers (clases puras), y que las clases puras son coberturas reconocibles que contienen propiedades espectrales homogéneas en la imagen. Un aspecto clave en el éxito del método es la selección de un número óptimo (mínimo) de clases puras lo cual está determinado por la dimensionalidad de los datos espectrales, es decir, por los rasgos biofísicos que el sensor es capaz de detectar más que por el número de bandas (Gao, 2009). Dada su efectividad para manejar el problema de mezclas espectrales, este método ha sido empleado para generar estimaciones continuas de variables biofísicas: 1) estimación de cubierta forestal (Song *et al.*, 2005; Vohland *et al.*, 2007), 2) degradación forestal (Matricardi *et al.*, 2010), 3)

determinación de superficie pavimentada (Myint y Okin *et al.*, 2009) y 4) determinación de temperatura (Amiri *et al.*, 2009).

Otro método de clasificación muy empleado es el de máxima verosimilitud que considera la varianza y la covarianza de la respuesta espectral de cada categoría y crea superficies en forma de campana llamadas funciones de densidad de probabilidades para cada categoría; las regiones de igual probabilidad son delineadas para expresar la sensibilidad del clasificador a la covarianza, mientras se aleje del centro de un cluster determinado (Narumalani *et al.*, 2002; Gao, 2009). Su versión difusa implica que la función de densidad de probabilidades posteriores de cada clase pueda ser interpretada como una proporción de coberturas (Gallaun *et al.*, 2009). A este método se le conoce como clasificador de probabilidad posterior bayesiana.

Por el lado no paramétrico, existe un algoritmo versátil aplicado en estudios forestales con el uso de datos espectrales y es el k vecino más cercano. Si la variable de interés es categórica el algoritmo funciona como clasificador, ya sea que se escoja al más cercano o como una moda entre los más cercanos (Zhou *et al.*, 2008; Tomppo *et al.*, 2009); si la variable a determinar es continua, entonces se aplica una variante donde el valor a estimar se calcula como un promedio que se pondera inversamente a la distancia espectral entre los k vecinos más cercanos (Gjertsen *et al.*, 2007; Fuchs *et al.*, 2009). Si el resultado deseado es difuso, se puede calcular la proporción de las clases encontradas entre los k vecinos más cercanos (Pham, 2005), aunque una condición es que las muestras de las clases tengan un tamaño similar para evitar sesgos en las estimaciones.

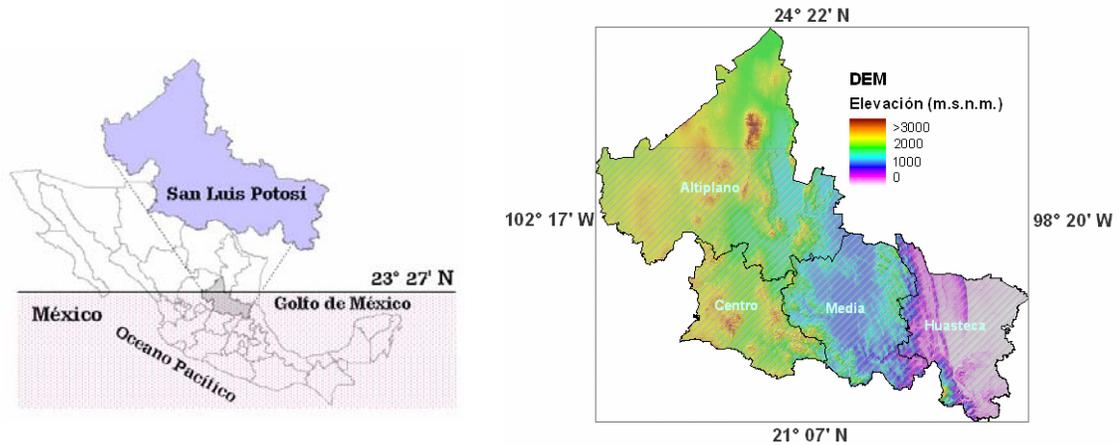
Aunque a la fecha se reporta una diversidad de métodos que determinan la composición proporcional de un píxel, no existen aplicaciones metodológicas para la estimación de la cobertura fraccional por grupos de vegetación en México; mas aún, la existencia de datos espectrales combinados MODIS MCD43 de buena calidad (Opazo y Chuvieco, 2007) para áreas extensas y

calibrados radiométricamente (NBAR, reflectancia ajustada bidireccionalmente al nadir) motiva sobre el potencial que tienen los métodos subpixel para la determinación de cobertura fraccional de formaciones vegetales. En este sentido, los objetivos de este trabajo fueron (1) analizar las firmas espectrales de clases puras de bosque y matorral, (2) generar mapas subpixel de cobertura fraccional a partir de datos combinados del producto MODIS MCD43A4 (500 m de resolución espacial) y los métodos de probabilidad posterior bayesiana, el estimador no paramétrico del k-vecino más cercano en su versión difusa y el análisis de mezcla espectral. Y finalmente, (3) validar las estimaciones con datos obtenidos del procesamiento de una imagen Landsat 5 TM (30 m resolución espacial).

3.2. Materiales y Métodos

3.2.1. Área de Estudio

San Luis Potosí es un estado que se localiza en la porción centro norte de México (entre 21° 07' - 24° 22' N y 98° 20' - 102° 17' W) y cuenta con una superficie de 6'116,360 ha (Figura 3.1). El estado se encuentra en el límite norte de la zona intertropical, cuenta con una riqueza de paisajes y condiciones de vegetación que puede ser comprendida analizando las zonas geográficas que lo componen: (1) Altiplano, (2) Centro, (3) Media y (4) Huasteca (INEGI, 2002). Esta distribución de ecosistemas obedece a la conformación orográfica, al patrón de distribución de precipitación que proviene del Golfo de México y a la relación inversa existente entre la temperatura y la altitud propia de estas latitudes. Los vientos cargados de humedad que precipitan en la Sierra Madre Oriental (Zona Media y Huasteca) y aunado al gradiente altitudinal, propician formaciones vegetales de clima tropical (selva) y templado (bosques); mientras que los vientos avanzan continente adentro, la vegetación va cambiando progresivamente hasta observar vegetación propia de climas áridos *v.gr.* matorral (Figura 3.2).



A)

B)

Figura 3.1. Localización del área de estudio. A) Contexto nacional y zona intertropical (hachurado) (Límite 23° 27' N) y B) Mapa topográfico.

3.2.2. Materiales

Después de realizar un análisis de control de calidad en los datos, se seleccionaron 235 conglomerados de muestreo de la colección del INF para San Luis Potosí que fueron medidos durante el periodo 2005-2007 (Figura 3.2). Cada uno de los conglomerados se compone de 4 sitios circulares de muestreo de 400 m² sumando 1600 m² de área (CONAFOR, 2010). Los datos se estratificaron en bosque (arbolado, incluyendo bosque y selva) y matorral. Las coordenadas de los sitios de muestreo se encontraban bajo la proyección Plate-Carrée y fueron proyectados al sistema de coordenadas UTM-14n datum WGS84. El procesamiento de los datos se llevó a cabo mediante programación en lenguaje SAS (SAS Institute Inc., 2004).

También se emplearon los datos de vegetación y uso de suelo serie IV (2007) escala 1:250,000 generados por el Instituto Nacional de Geografía y Estadística (INEGI) mediante técnicas de fotointerpretación de imágenes de satélite Landsat y SPOT con un cubrimiento nacional. Las claves de las cartas seleccionadas fueron como sigue: Fresnillo (F13-3), Zacatecas (F13-6), Matehuala (F14-1), San Luis Potosí (F14-4), Ciudad Mante (F14-5), Guanajuato

(F14-7), Ciudad Valle (F14-8) y Concepción del Oro (G14-10). Esta información fue usada para la selección de las clases puras así como la generación de una máscara espacial para extraer las zonas con vegetación natural dentro del área de estudio.

3.2.3. Datos Espectrales

Se empleó el producto combinado MCD43A4 que se obtiene del sensor MODIS instalado en los satélites Aqua y Terra; se le denomina reflectancia (ρ) ajustada al nadir (MCD43A4) con una función de distribución de reflectancia bidireccional; integra datos multifecha tomados durante periodos de 16 días y se encuentra corregido atmosféricamente y libre de nubes. Tiene una resolución espacial de 500 m y una resolución espectral de siete bandas distribuidas en las siguientes longitudes de onda: 1. Rojo (620 - 670 nm), 2. Infrarrojo cercano (841 - 876 nm), 3. Azul (459 - 479 nm), 4. Verde (545 - 565 nm), 5. Infrarrojo medio₁ (1230 - 1250 nm), 6. Infrarrojo medio₂ (1628 - 1652 nm), 7. Infrarrojo medio₃ (2105 - 2155 nm) (Bhaduri *et al.*, 2010). Para el análisis se usaron datos espectrales del periodo 11 al 26 de febrero de 2006 (día juliano 42 al 57), que es aproximada a la fecha de levantamiento de datos del INF de México; fueron obtenidos mediante la interface denominada Modis Reprojection Tool (MRTWeb) (<http://lpdaac.usgs.gov/datapool/datapool.asp>) (Daucsavage *et al.*, 2010).

Se realizó una imagen compuesta en falso color con los primeros tres componentes principales para enfatizar visualmente la variabilidad espectral y espacial de las diferentes coberturas y facilitar la selección de los píxeles de las clases puras de interés (Figura 3.2).

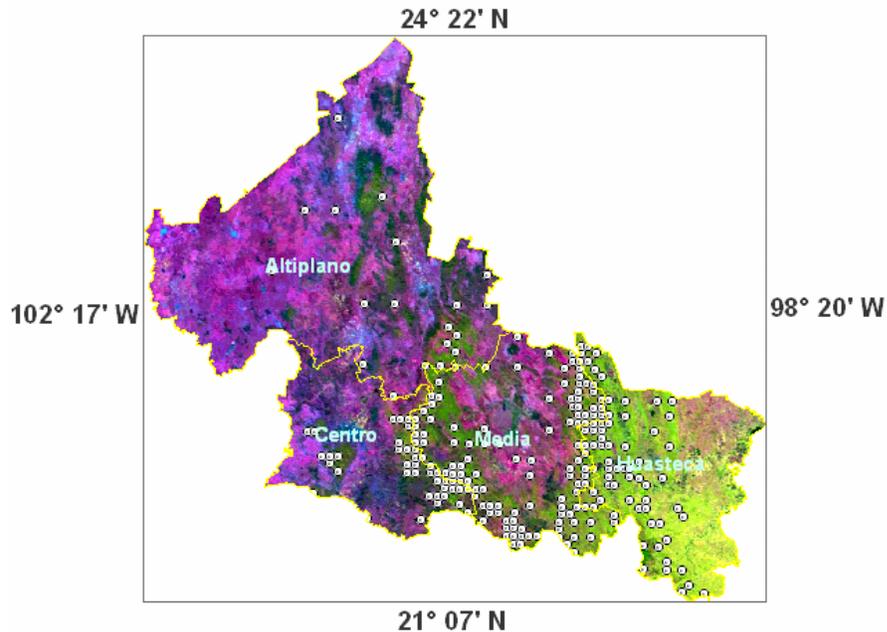


Figura 3.2. Distribución espacial de los sitios de muestreo seleccionados del Inventario Forestal Nacional en el área de estudio sobre imagen en falso color compuesta con los tres primeros componentes principales.

3.2.4. Procesamiento de la Información

Existen dos tipos de clasificadores de imágenes, rígidos y blandos. Los rígidos producen una decisión definitiva sobre la identidad de cada píxel, mientras que los blandos expresan el grado en el que un píxel pertenece a cada una de las clases consideradas. Uno de los motivos para emplear un clasificador blando es determinar la mezcla de clases de cobertura presentes en un píxel, lo que comúnmente se conoce como clasificación subpíxel (Eastman y Laney, 2002). Como datos de entrenamiento se seleccionaron píxeles puros para cada una de las categorías seleccionadas: (1) bosque y (2) matorral. Como variables en la clasificación subpíxel se emplearon las bandas espectrales 1-7. Se aplicó el algoritmo no paramétrico cluster para la purificación de firmas espectrales y así maximizar la separabilidad de las clases, calculada con el Índice de Divergencia Transformado (IDT) (Matheron, 2004; Eastman, 2006).

Para la determinación de la cobertura fraccional se emplearon tres métodos: 1) probabilidad posterior bayesiana, 2) el k vecino más cercano en su versión difusa y 3) el análisis de mezcla espectral. Para el caso del clasificador bayesiano, se determinó la probabilidad posterior de la membresía de clase

segun Strahler (1980):
$$\Phi_k(X_i) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(X - \mu_k)^T \Sigma_k^{-1}(X - \mu_k)\right\},$$

donde p = número de variables (bandas espectrales) que caracterizan a cada observación; X_i = vector de mediciones de las p variables asociadas con la i-ésima observación; $\Phi_k(X_i)$ = valor de densidad de probabilidades asociada con el vector de observación X para la clase k; Σ_k = matriz de varianzas-covarianzas asociada con la k-ésima clase; μ_k = vector de medias asociado con la k-ésima clase.

El clasificador del k-vecino más cercano en su versión difusa (Kuo *et al.*, 2009) se empleó para determinar los mapas de cobertura fraccional. El grado de membresía de un píxel a una categoría se calcula como la proporción de cada categoría entre los k-vecinos espectralmente más cercanos. Para el cálculo de la distancia espectral se empleó la fórmula $D(p, q, W) = [(p - q)^T W (p - q)]^{1/2}$ donde D = distancia Euclidiana espectral; W = matriz identidad, p = conjunto de observaciones de referencia (muestra), q = conjunto de observaciones objetivo. Se empleó k = 30 para que se tuvieran suficientes observaciones para estimar la proporción de clases. Una consideración importante de este algoritmo es que las clases empleadas tengan un tamaño similar para disminuir el sesgo ocasionado por la influencia del tamaño de clase, situación que sí fue prevista en este trabajo.

Otro método empleado fue el análisis de mezcla espectral para determinar la cobertura fraccional de las clases puras dentro de cada uno de los píxeles combinados; el valor de la reflectancia de la banda i , R_i , puede ser expresada matemáticamente como: $R_i = \sum_{j=1}^n f_j \cdot RE_{ij} + \varepsilon_i$, donde R_i = reflectancia espectral mezclada de un píxel en la i -ésima banda, f_j = abundancia de la j -ésima clase pura (endmember), RE_{ij} = reflectancia del endmember j en la banda i , ε_i = residual en la i -ésima banda espectral y n = número de endmembers. Si el ruido se encuentra no correlacionado, la ecuación lineal puede ser invertida para calcular f_j basada en el principio de mínimos cuadrados, o $f = (E' E)^{-1} E' r$, donde r es el vector de reflectancia observada $[r_1, r_2, \dots, r_n]$ y E es una matriz $n \times j$ (número de espectro de endmember). Esta ecuación tiene dos supuestos: $\sum f = 1$, la suma de las ponderaciones debe ser igual a 1 y $f \geq 0$, el principio de no negatividad (Gao, 2009).

3.2.5. Validación

Para evaluar la fiabilidad de métodos empleados, los mapas de cobertura fraccional fueron validados con información obtenida de la clasificación supervisada (máxima verosimilitud) de una imagen de satélite Landsat 5 TM: escena, path 27 / row 47; fecha de toma, 10 de enero de 2006; resolución espacial, 30 m; bandas espectrales empleadas, azul (banda 1) al infrarrojo medio (banda 7). Los números digitales (ND) de cada banda espectral fueron corregidos radiométricamente: 1) inicialmente se convirtieron a radianza espectral (L_λ) (Amiri *et al.*, 2009) y posteriormente a reflectancia exoatmosférica adimensional (ρ_λ) (Chander *et al.*, 2010). Dado que la zona de validación presenta relieve montañoso, la reflectancia fue corregida topográficamente con el método de la corrección-C (Teferi *et al.*, 2010). Finalmente, se usó el

coeficiente de correlación de Pearson (r) y el RMSE para comparar la estimación fraccional de la imagen MODIS vs Landsat.

3.3. Resultados y Discusión

3.3.1. Selección de Endmembers

Se realizó un análisis de separabilidad de firmas espectrales con el IDT (Matheron, 2004) que se calcula de manera pareada para cada clase temática donde se obtuvo el valor máximo ($IDT = 100$) que representa una mínima confusión de las firmas espectrales, explicado por la purificación de la firma espectral mediante el algoritmo no paramétrico cluster. En la Figura 3.3 se observa que el bosque presenta una mayor absorción de energía electromagnética en el rojo, mientras que en el infrarrojo ostenta una mayor reflectancia (Gao, 2009). La pequeña diferencia positiva de la reflectancia registrada en el verde comparado con el azul y el rojo (espectro visible) explica la tonalidad verdosa de la zona boscosa, mientras que para el caso de los matorrales existe un incremento en el rojo comparado contra los dos colores primarios restantes. En la porción de los infrarrojo medio, las bandas 6 (1628 – 1652 nm) y 7 (2105 – 2155 nm) que son sensibles al contenido de humedad en la vegetación, discriminaron espectralmente mejor ambas clases.

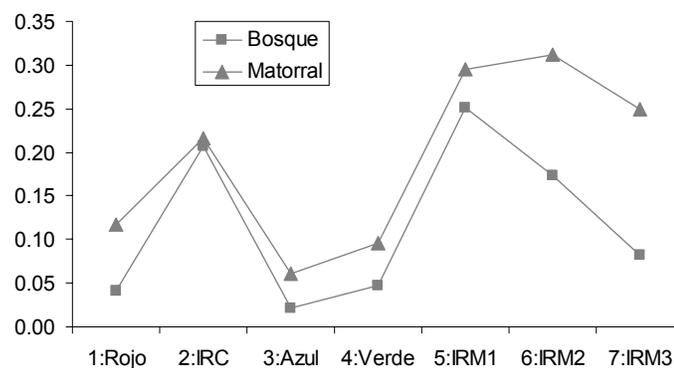


Figura 3.3. Firmas espectrales puras basadas en datos MODIS MCD43A4. IRC: Infrarrojo Cercano, IRM: Infrarrojo Medio.

3.3.2. Selección del Método

Para seleccionar el mejor método de mapeo subpixel se realizó una validación de las estimaciones de cobertura fraccional de cada uno de los métodos (bayesiano de probabilidad posterior, kNN difuso y análisis de mezcla espectral). La comparación se realizó con la información obtenida mediante clasificación supervisada con el algoritmo de máxima verosimilitud de la porción de una escena Landsat 5 TM (30 m de resolución espacial) tomada en los límites de la zona centro y media donde se presentan las clases de interés (bosque y matorral). En la Figura 3.5 se observa resolución espacial ofrecida por MODIS (500 m) contra el alto nivel de detalle alcanzado por el sensor TM de Landsat (Figura 3.4).

La imagen MODIS clasificada con el algoritmo subpixel fue validada mediante la comparación de las proporciones por tipos de vegetación para cada píxel. Es interesante el comportamiento de algoritmo bayesiano ($r = -0.64$ y -0.54 , para bosque y matorral respectivamente) ya que no fue como se esperaba mostrando resultados incoherentes, siendo superado considerablemente por FkNN ($r = 0.78$, en ambas categorías, con un RMSE = 0.23%) y LMA ($r = 0.81$, en ambas categorías, con un RMSE = 0.14%) quienes sí presentaron correlaciones significativas y generaron una estimación más cercana a la realidad (Figura 3.6). Eastman y Laney (2002) concluyeron que bajo las condiciones del mundo real, el éxito de la aplicación del método de la probabilidad posterior bayesiana para la clasificación subpixel resulta altamente inconsistente y difícilmente controlado.

No obstante para el caso de los dos métodos restantes (FkNN y LMA), el error alcanzado en esta investigación es comparable con algunos trabajos similares. Song *et al.* (2005) estimaron la cobertura fraccional mediante el análisis espectral de mezclas dentro de píxeles de imágenes Landsat (30 x 30 m de resolución espacial) y validaron con imágenes Ikonos (4 x 4 m de resolución espacial); el RMSE varió de 0.10 a 0.18%. Xiao y Moody (2005)

estimaron cobertura fraccional con datos Landsat (30 m de resolución espacial) y validaron con una ortoimagen en color verdadero (3 m de resolución espacial); el RMSE varió de 0.10 a 0.15%.

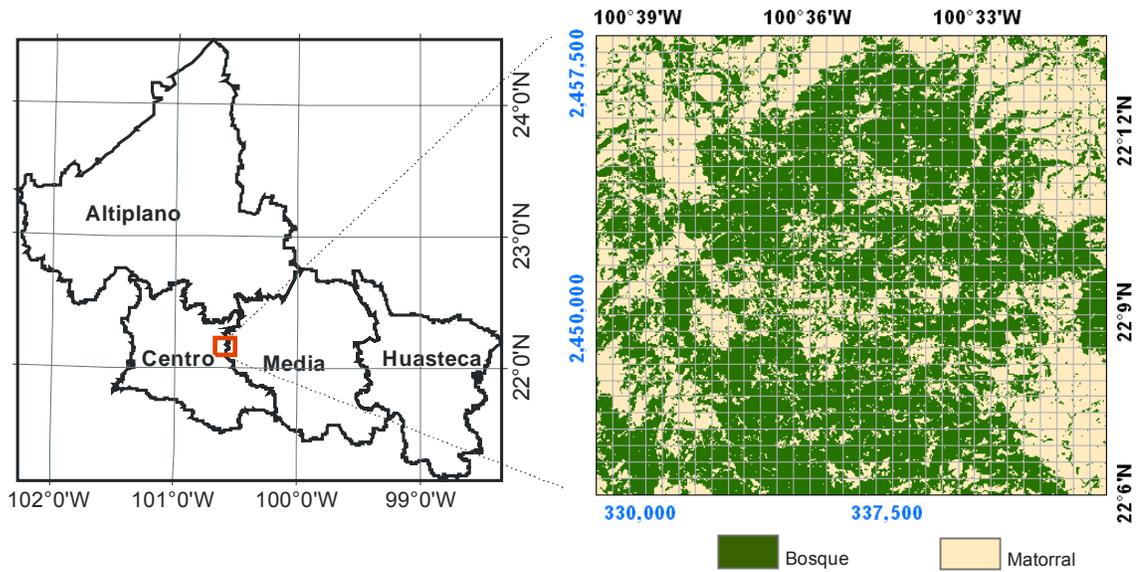


Figura 3.4. Izq. Localización de la zona de validación y Der. Clasificación con el algoritmo de máxima verosimilitud. Coordenadas UTM-14N (m) en Azul, Coordenadas Lat-Long en Negro.

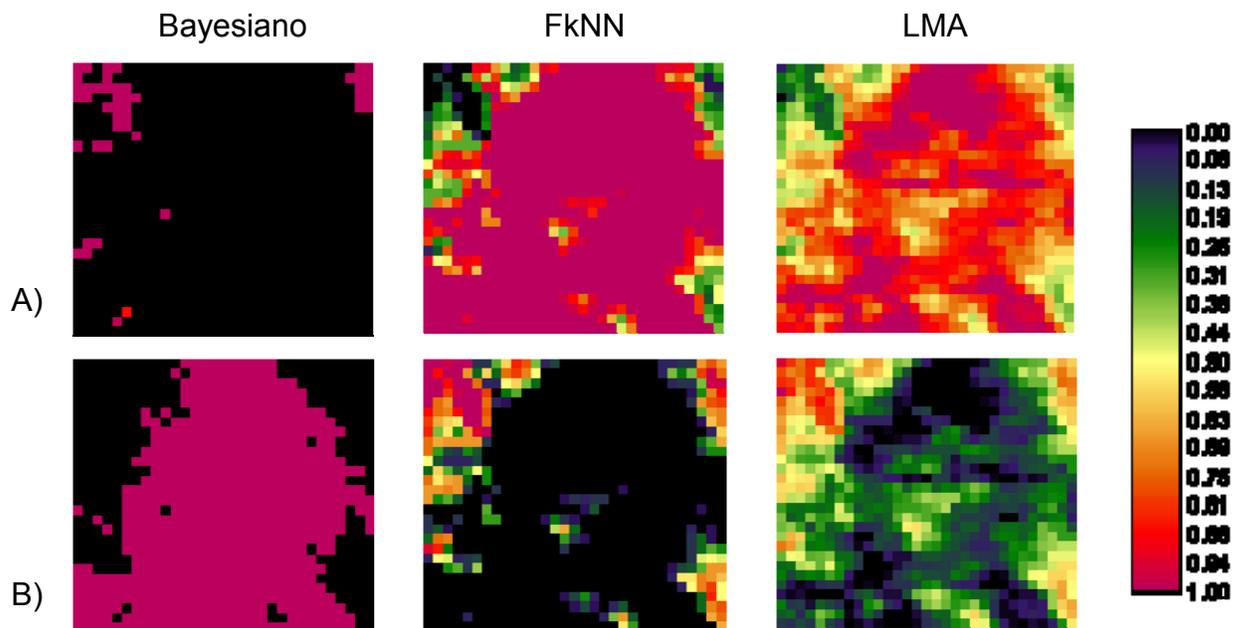


Figura 3.5. Cobertura fraccional en el área de validación. A) Bosque y B) Matorral.

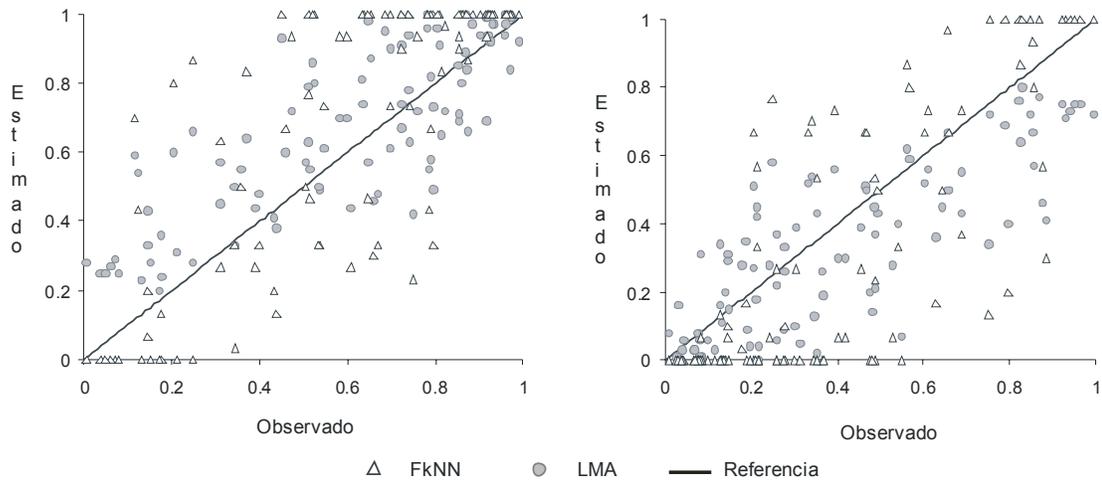


Figura. 3.6. Cobertura fraccional estimada vs datos de referencia. Izq. Bosque. Der. Matorral.

3.3.3. Mapas de Cobertura Fraccional

En la Figura 3.7 A y B se muestran los mapas obtenidos con el análisis de mezcla espectral para la estimación de cobertura fraccional para bosque y matorral. La mayor probabilidad de encontrar zonas arboladas se concentra en las zonas elevadas coincidiendo con las áreas que captan la precipitación que proviene del Golfo de México; mientras que la presencia de matorrales obedece a la distribución del clima árido característico de la zona altiplano con precipitaciones que varían de los 0 a los 600 mm. Los terrenos no forestales fueron extraídos mediante una máscara generada con el mapa de vegetación y uso del suelo serie IV (2007).

Aunque el principal objetivo de este trabajo no es la estimación de biomasa (B), resulta pertinente tratar brevemente este punto para contrastar mejor las estimaciones de cobertura fraccional en unidades de referencia absoluta (Mg/ha). Por lo que los mapas de cobertura fraccional fueron escalados con estimaciones de densidad forestal realizadas con el modelo no lineal $B = 13.8926e^{NDVI62*(-4.0086) + VCF*(0.0275)}$, con Pseudo $r^2 = 0.77$, donde B = Biomasa aérea (Mg/ha); $NDVI62 = (\rho_{irm2} - \rho_{irc}) / (\rho_{irm2} + \rho_{irc})$, ρ_{irm2} = reflectancia en

el infrarrojo medio (1628 – 1652 nm), ρ_{irc} = reflectancia en el infrarrojo cercano (841 – 876 nm); VCF = Campos continuos de vegetación. Este modelo se generó con datos del Inventario Nacional Forestal para la zona de estudio.

Las estimaciones de biomasa para bosques templados reportadas oscilan entre 80-120 Mg/ha (Rodríguez Laguna *et al.*, 2006, 2009; Acosta-Mireles *et al.*, 2009) y la mayor parte se encuentra en la zona Media del área de estudio. Aunque también existen zonas de mayor densidad forestal que pueden ser superiores a 150 Mg/ha B y normalmente se encuentran en la zona montañosa de la Huasteca, donde existen mejores condiciones ecológicas para el crecimiento de árboles de gran porte (Návar-Chaidez *et al.*, 2011). Para matorral xerófilo también se obtuvieron estimaciones consistentes con la literatura. Espinoza-Bretado y Návar (2005) determinaron la biomasa aérea en un matorral espinoso Tamaulipeco (MET) y encontraron una diversidad de densidades que varían conforme la posición de la parcela de medición en un gradiente topográfico; las zonas más conservadas presentan 80 Mg/ha B llegando hasta un mínimo de 10 Mg/ha. Dado que la firma espectral de este tipo de vegetación es característica, ya que la composición de especies presenta una mayor producción foliar y por ende un mayor contenido clorofílico, el MET difiere del matorral xerófilo tradicional; para el caso particular de este trabajo, MET puede ser interpretado e incluido dentro del mapa de cobertura fraccional arbolada. No obstante, para el resto de tipos de vegetación correspondientes al matorral xerófilo (*v.gr.* crassicaule, micrófilo, etc.) se pueden comparar con el mapa generado para matorral *per se*. La mayoría de trabajos reportan estimaciones de biomasa para zonas áridas entre 2 y 10 Mg/ha. Por ejemplo, Robles-Esparza *et al.*, 2008, reportan 3.6 Mg/ha B para matorral de *Dasyllirion spp* en Zacatecas, México. Navarro-Cerrillo y Blanco-Oyonarte (2006) reportan para ecosistemas del sur de España, desde 4 a 11 Mg/ha B para matorrales con especies representativas de porte bajo y con excepciones de hasta 19 Mg/ha para matorrales con especies de tamaño medio.

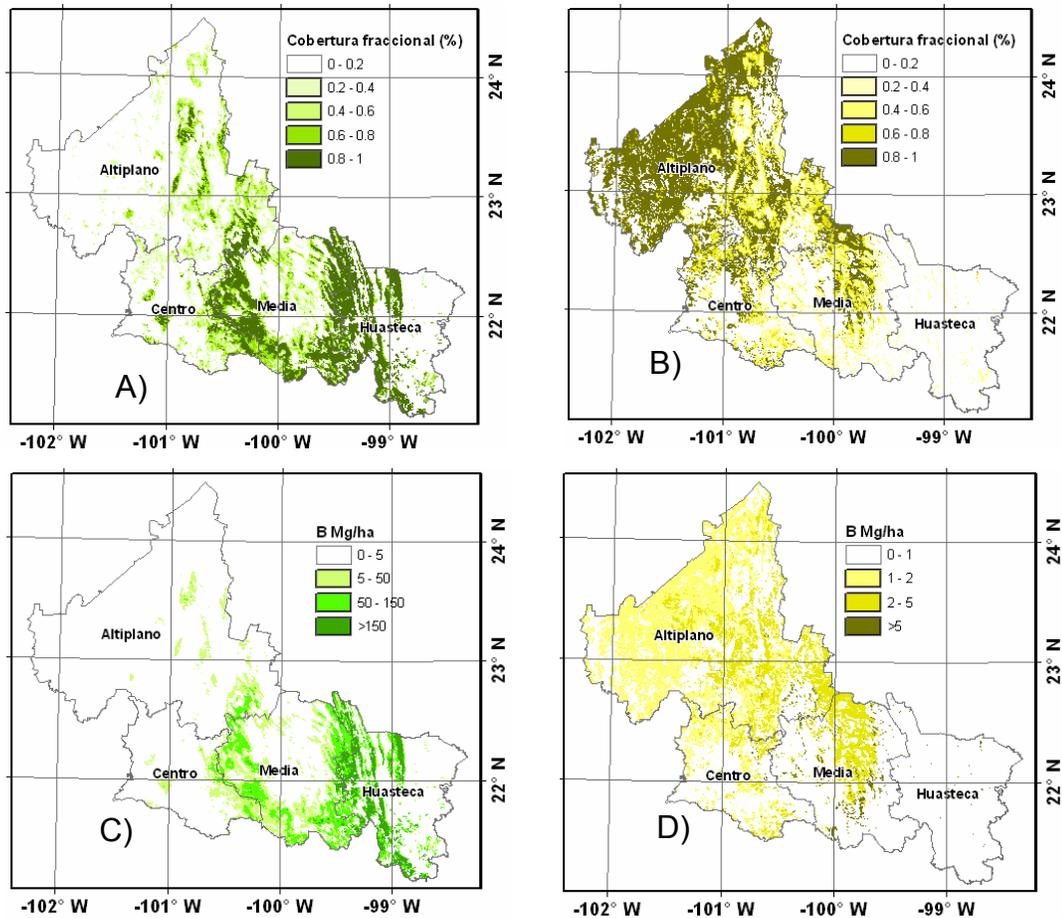


Figura 3.7. Mapas de cobertura fraccional con el método de análisis de mezcla espectral y escalamiento de la cobertura fraccional con el mapa de biomasa arbórea aérea. A) y C) Bosque, B) y D) Matorral.

3.3.4. Distribución espacial del error

De los tres métodos empleados, el error solamente puede ser estimado para el método bayesiano y el LMA ya que tienen funciones de probabilidad asociadas (*v.gr.* distribución normal). Dado que el método bayesiano resultó por demás incoherente en la determinación de la cobertura fraccional, no se presenta su error. No obstante, para el caso de LMA que fue el método que sí estimó correctamente la cobertura fraccional, se presenta la distribución espacial del error obtenido para las estimaciones que varía de 0 a 0.1, con media de 0.05 y se encuentra dentro rango del ejercicio de validación (Figura 3.8). Aunque los

errores obtenidos son bajos, es interesante comentar que los mayores sesgos en el ajuste se encuentran espacialmente en los píxeles espectralmente mas alejados de los valores medios de la firma espectral (García, 2003).

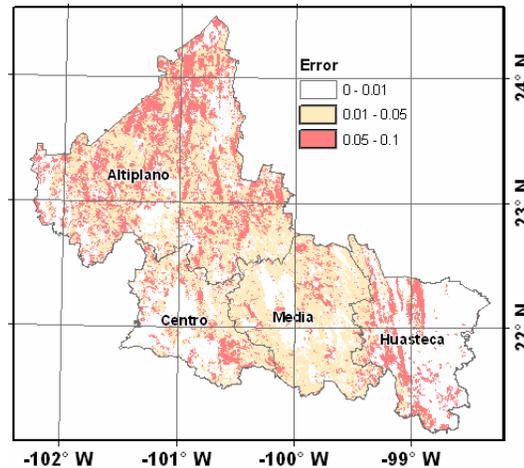


Figura 3.8. Distribución espacial del error para LMA.

3.4. Conclusiones

En este capítulo se desarrolló una metodología para la estimación de cobertura fraccional a nivel subpixel mediante datos espectrales de resolución espacial moderada (MCD43A4) en una porción del límite septentrional de la zona intertropical en México, particularmente el estado de San Luis Potosí. De los métodos empleados (además del bayesiano y FkNN), el LMA ($r = 0.81$, $RMSE = 0.14$) fue el que presentó un mejor ajuste con los datos de referencia obtenidos con la estimación fraccional por la clasificación supervisada de una imagen de satélite Landsat para documentar a nivel subpixel la cobertura fraccional de bosque vs matorral. Finalmente, la cobertura fraccional fue escalada a biomasa arbórea aérea con un modelo no lineal que relaciona datos de densidad forestal y variables satelitales (NDVI62 y VCF). Esto indica el potencial predictivo de la clasificación subpixel para la determinación de la proporción de las principales coberturas en los inventarios forestales usando firmas espectrales puras.

3.5. Literatura Citada

- Acosta-Mireles, M., Carrillo-Anzures, F. y Díaz-Lavariega, M. 2009. Determinación del carbono total en bosques mixtos de *Pinus patula* Schl. et Cham. Terra Latinoamericana. 27(2):105-114.
- Amiri, R., Weng, Q., Alimohamadi, A. and Alavipanah, S.K. 2009. Spatial-temporal dynamics of land surface temperature in relation to fractional vegetation cover and land use/cover in the Tabriz urban area, Iran. Remote Sensing of Environment. 113:2606–2617.
- Bhaduri, K., Das, K. and Votava, P. 2010. Distributed anomaly detection using satellite data from multiple modalities. Proceedings of The 2010 Conference on Intelligent Data Understanding (CIDU-NASA). San Francisco Bay Area. Oct 5-7. pp.109-123.
- Chander, G., Xiong, X., Choi, T. and Angal, A. 2010. Monitoring on-orbit calibration stability of the Terra MODIS and Landsat 7 ETM+ sensors using pseudo-invariant test sites. Remote Sensing of Environment. 114:925–939.
- CONAFOR. 2010. Manual de Procedimientos para el muestreo de campo. Remuestreo-2010. Inventario Nacional Forestal y de Suelos. Comisión Nacional Forestal. Secretaría del Medio Ambiente y Recursos Naturales. Zapopan, Jalisco, México. 140 p.
- Daucsavage, J. Kaminski, M., Ramachandran, B., Jenkerson, C., Sprenger, K., Faust, R. and Rockvam, T. 2010. ASTER and MODIS Land Data Management at the Land Processes, and National Snow and Ice Data Centers. In: Ramachandran, B., Justice, C.O., and Abrams, M.J. Land Remote Sensing and Global Environmental Change. NASA Earth Observing System and the Science of ASTER and MODIS. Springer. pp. 167-182.
- Eastman, J.R. and Laney, R.M. 2002. Bayesian soft classification for subpixel analysis: a critical evaluation. Photogrammetric Engineering & Remote Sensing. 68(11):1149-1154.

- Eastman, R.J. 2006. IDRISI Andes Guide to GIS and Image Processing. Clark University. 327 p.
- Espinoza-Bretado, R. y Návar, J. 2005. Producción de biomasa, diversidad y ecología de especies en un gradiente de productividad en el matorral espinoso tamaulipeco del nordeste de México. *Revista Chapingo. Serie Ciencias Forestales y del Ambiente*. 11(1):25-31.
- Fuchs, H., Magdon, P., Klein, C. and Flessa, H. 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: combining satellite imagery and field inventory. *Remote Sensing of Environment*. 113:518–531.
- Gallaun, H., Zanchi, G., Nabuurs, G.J., Hengeveld, G., Schardt, M. and Verkerk, P. 2009. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*. 260(3):252-261.
- Gao, J. 2009. *Digital Analisis of Remotely Sensed Imagery*. McGraw Hill. 645 p. ISBN:978-0-07-160466-6.
- García, M. 2003. Respuesta de la vegetación a variaciones climáticas en praderas y sistemas adehesados Mediterraneos: Metodología de análisis utilizando datos hiperespectrales y multiespectrales. PhD Thesis in English. Universidad Politecnica de Madrid. pp. 4. Disponible en: <http://oa.upm.es/257/1/02200322.pdf>
- Gjertsen, A.K. 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment*. 110:420–430.
- Kuo, B.C., Yang, J.M. and Li, C.H. 2009. Remote sensing data classification with kernel nonparametric feature extractions. In: Camps-Valls, G. and Bruzzone, L. *Kernel Methods for Remote Sensing Data Analysis*. p. 414.
- Matheron, P.M. 2004 *Computer processing of remotely sensed images*. 3rd Ed. John Wiley & Sons Ltd. ISBN: 0-470-84919-3. p. 243.
- Matricardi, E.A.T., Skole, D.L., Pedlowski, M.A., Chomentowski, W. and Fernandez, L.C. 2010. Assessment of tropical forest degradation by

- selective logging and fire using Landsat imagery. *Remote Sensing of Environment*. 114:1117–1129.
- Myint, S.W. and Okin, G.S. 2009. Modelling land-cover types using multiple endmember spectral mixture analysis in a desert city. *International Journal of Remote Sensing*. 30(9):2237-2257.
- Narumalani, S., Hlady, J.T. and Jensen, J.R. 2002. Information extraction from remotely sensed data. In: Bossler, J.D., Jensen, J.R., McMaster, R.B. and Rizos, C. *Manual of Geospatial Science and Technology*. Taylor & Francis. 1 Ed. ISBN: 0-7484-0924-6. p. 312.
- Návar-Chaidez, J. 2011. The spatial distribution of aboveground biomass in tropical forests of Mexico. *Tropical and Subtropical Agroecosystems*. 13:149-158.
- Navarro-Cerrillo, R.M. and Blanco-Oyonarte, P. 2006. Estimation of above-ground biomass in shrubland ecosystems of southern Spain. *Investigación Agraria. Sistemas y Recursos Forestales*. 15(2):197-207.
- Opazo, S. y Chuvieco, E. 2007. Utilización de productos MODIS para la cartografía de áreas quemadas. *Revista de Teledetección*. 27:27-43.
- Pham, T.D. 2005. An optimally weighed Fuzzy k-NN algorithm. *In*: Singh, S., Singh, M., Apte, C. and Perner, P. *Pattern recognition and data mining. Third International Conference on Advances in Pattern Recognition. ICAPR 2005*. Bath, UK, August 2005. Springer Berlin Heidelberg New York. ISSN: 3-540-28757-4. pp.239-247.
- Robles-Esparza, A., España-Montoya, J.L., Robles-Berumen, H. 2008. Biomasa y forraje, distribución espacial y abundancia de la planta de sotol (*Dasyllirion spp*) en el ejido el Jazmín, Mazapil, Zacatecas, México. *Revista Investigación Científica*. 4(2):1-9.
- Rodríguez-Laguna, R., Jiménez-Pérez, J., Aguirre-Calderón, O.A. y Treviño-Garza, E.J. 2006. Estimación del carbono almacenado en un bosque de niebla en Tamaulipas, México. *Ciencia UANL*. 9(2):179-188.
- Rodríguez-Laguna, R., Jiménez-Pérez, J., Aguirre-Calderón, O.A., Treviño-Garza, E.J. y Razo-Zárte, R. 2009. Estimación de carbono almacenado

- en el bosque de pino-encino en la Reserva de la Biosfera "El cielo", Tamaulipas, México. *Ra Ximhai*. 5(3):317-327.
- SAS Institute Inc. 2004. *SAS/STAT 9.1 User's Guide*. SAS Institute Inc., Cary, NC, USA. 5121 p.
- Song, C. 2005. Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?. *Remote Sensing of Environment*. 95:248-263.
- Strahler, A.H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*. 10:135–163.
- Teferi, E., Uhlenbrook, S., Bewket, W., Wenningerm J. and Simane, B. 2010. The use of remote sensing to quantify wetland loss in the Choke Mountain range, Upper Blue Nile Basin, Ethiopia. *Hydrology and Earth System Sciences*. 14:2415-2428.
- Tomppo, E.O., Gagliano, C., De Natale, F., Katila, M. and McRoberts, R. 2009. Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery. *Remote Sensing of Environment*. 113:500–517.
- Vohland, M., Stoffels, J., Hau, C. and Schüler, G. 2007. Remote sensing techniques for forest parameter assessment: multispectral classification and linear spectral mixture analysis. *Silva Fennica*. 41(3):441-456.
- Xiao, J. and Moody, A. 2005. A comparison of methods for estimating fractional green vegetation cover within a desert to upland transition zone in central New Mexico, USA. *Remote Sensing of Environment*. 98:237-250.
- Zhou, P., Luukknen, O., Tokola, T. and Nieminen, J. 2008. Effect of vegetation cover on soil erosion in a mountainous watershed. *Catena* 75(3):319-325.

CAPÍTULO 4

COMPARACIÓN DE ALGORITMOS NO PARAMÉTRICOS DEL K VECINO MÁS CERCANO PARA LA ESTIMACIÓN DE LA BIOMASA SUPERFICIAL EN EL NORTE CENTRO DE MÉXICO¹

Resumen

Este capítulo se enfoca en la comparación de cuatro variantes del algoritmo no paramétrico del vecino más cercano (kNN) para la modelación geoespacial de la biomasa forestal en una porción del límite septentrional de la zona intertropical en México mediante la combinación de datos espectrales (MCD43A4), topográficos y climáticos a una resolución moderada con datos obtenidos del Inventario Nacional Forestal. Se evaluó el desempeño de diferentes matrices de ponderación de la distancia espectral aplicada al método kNN mediante un algoritmo de validación cruzada dejando uno fuera. Los resultados obtenidos indican que el método del vecino más similar (MSN), maximizó correctamente la correlación entre las variables explicatorias y la variable de respuesta ($r = 0.9$). Tanto los errores obtenidos como las estimaciones presentadas se validaron con estudios similares reportados en la literatura. Esto confirma el potencial predictivo del MSN para la realización de inventarios forestales a nivel de píxel bajo el contexto de la iniciativa de REDD+.

Palabras clave: KNN, Mahalanobis, vecino más similar, MCD43A4, inventario forestal, REDD+.

¹ Sometido a arbitraje: Aguirre-Salado, C.A., Treviño-Garza, E.J., Aguirre-Calderón, O.A., Jiménez-Pérez, J., González-Tagle, M.A., Valdez-Lazalde, J.R., Sánchez-Díaz, G., Haapanen, R., Aguirre-Salado, A.I. and Miranda Aragón, L. 2012. Mapping aboveground biomass integrating geospatial and forest inventory data through a k nearest neighbor strategy in North Central Mexico.

Abstract

This chapter is focused on a comparison of four variations of the nonparametric k-nearest neighbor for geospatial modeling of aboveground biomass in a portion of the septentrional border of the intertropical zone combining spectral, climatic and topographic predictor variables with Mexican National Forest Inventory data. The performance of including several weighting matrices applied into kNN algorithm is evaluated using a cross validation leave one out technique. The results obtained indicate that the MSN approach, correctly maximized the correlation between predictor and response variables ($r = 0.9$). Both estimations and errors obtained were successfully validated with similar studies reported in literature. This confirms the predictive potential of MSN approach for making forest inventories at pixel level under the REDD+ initiative.

Key words: kNN, Mahalanobis, most similar neighbor, MCD43A4, forest inventory, REDD+.

4.1. Introducción

La Convención Marco de las Naciones Unidas sobre el Cambio Climático (UNFCCC, por sus siglas en inglés) propone la posibilidad de introducir un esquema de compensación para Reducir Emisiones por Deforestación y Degradación (REDD+) posterior al protocolo de Kioto. Dado que la deforestación representa aproximadamente el 18% de las emisiones de gases de efecto invernadero, los bosques se centran como uno de los objetivos principales para el manejo del cambio climático (Streck *et al.*, 2008; Köhl *et al.*, 2009).

Los países que se encuentren interesados en beneficiarse con tales incentivos, deben desarrollar un sistema propio para monitorear (M), reportar (R) y verificar (V) los almacenes de carbono forestal de una manera consistente, comparable y completa; esto servirá para llevar un control transparente de la contabilidad del carbono y fomentará la confianza de quienes deseen invertir fondos económicos para compensar o neutralizar sus emisiones de gases de efecto invernadero bajo el esquema REDD+ (FAO, 2009; Havemann *et al.*, 2009).

Para la correcta integración de un sistema de monitoreo para REDD+ se requiere del uso complementario de diferentes métodos. La fase inicial involucra la generación de modelos alométricos por género o especie para estimar biomasa individual o incluso carbono usando diámetro normal y altura del árbol. Afortunadamente a la fecha, ya existe un considerable banco de ecuaciones alométricas en literatura científica para transformar el volumen individual (m^3) en biomasa (Mg) o en carbono (Mg) (Návar, 2009). Tales datos individuales son sumados y extrapolados a la hectárea para obtener una estimación por sitio o conglomerado de muestreo. Para espacializar las estimaciones de carbono de las parcelas de inventario, existen dos métodos generales definidos por la información espacial fuente: 1) mapas vectoriales o 2) imágenes de satélite. En

el primer método, se requieren mapas de vegetación y uso de suelo que cuentan inherentemente con estimaciones de superficie por tipo de clase temática. El carbono forestal extrapolado a la hectárea es asumido para cada uno de los polígonos de interés y después de multiplicar los valores promedio de carbono (Mg/ha) por la superficie total de cada polígono se obtiene la estimación generalizada de los almacenes de carbono. De Jong *et al.* (2010) emplearon este método para estimar las emisiones de gases de efecto invernadero por cambio de uso del suelo en la República Mexicana (1993-2002), mismas que fueron empleadas en la 4ta. Comunicación Nacional ante la Convención Marco de las Naciones Unidas sobre el Cambio Climático (SEMARNAT-INE, 2009). Este método se menciona como parte de los procedimientos válidos para el monitoreo y reporte de las emisiones de gases de efecto invernadero antropogénicos revisados en la COP16 (GOFC-GOLD, 2010). Otros países como India (Kaul *et al.*, 2009) y Uganda (Nakakaawa *et al.*, 2011) también han empleado exitosamente esta metodología para estimar los flujos de carbono en sus ecosistemas.

Por otro lado, el segundo método (y es el propuesto por este trabajo) para realizar estimaciones totales de almacenes de carbono es utilizado por países que tienen tradición en el manejo de información proveniente de sensores remotos a través el procesamiento digital de imágenes de satélite. La información cuantitativa y/o cualitativa generada en los inventarios forestales se relaciona estadísticamente con los datos provenientes de la imagen de satélite. Este método aprovecha el poder de emplear el vector espectral multivariado que se refleja en la variabilidad espacial de las estimaciones en cada píxel. Existen experiencias, reportadas en literatura reciente, como la de Blackard *et al.* (2008) en Estados Unidos, como la de Anaya *et al.* (2009) en Colombia, o Gallaun *et al.* (2009) en Europa que hacen de esta metodología un método bastante recomendable para la realización de estimaciones de carbono forestal de países completos, pero con el detalle que ofrece la información espectral, píxel a píxel. Esto evidencia la importancia del valor agregado que ofrecen las

imágenes de satélite para la construcción del sistema MRV de cualquier país (Holmgren *et al.*, 2008; De Leeuw *et al.*, 2010) y es una de las principales motivaciones para el presente estudio.

Un algoritmo denominado el k vecino más cercano (kNN, por sus siglas en inglés), ha sido usado ampliamente para estimar la densidad forestal píxel a píxel usando imágenes de satélite. Este algoritmo presenta ventajas sobre los modelos paramétricos porque representa una manera dinámica (cuasi-irrepetible para cada píxel) de estimar la densidad forestal, convirtiéndolo en un método muy poderoso (Chirici *et al.*, 2008; McRoberts, 2008). Algunas aplicaciones recientes para este algoritmo incluyen la estimación de área basal (Finley y McRoberts, 2008), índice de área foliar y cobertura arbórea (Aguirre-Salado *et al.*, 2011), volumen de madera (Finley y McRoberts, 2008; Baffeta *et al.*, 2009; Barth *et al.*, 2009; McRoberts, 2009), biomasa (Breidenbach *et al.*, 2011), carbono (Aguirre-Salado *et al.*, 2009; Fuchs *et al.*, 2009; Stümer *et al.*, 2010), composición de especies forestales (Gjertsen *et al.*, 2007), tipos de vegetación (Zhou *et al.*, 2008) entre otros.

Actualmente existen algunas variantes en la aplicación del algoritmo kNN que radican en la determinación de la distancia espectral. En este sentido, los objetivos de este trabajo fueron 1) determinar las variables predictoras (espectrales, climáticas y topográficas) significativas en la estimación de la densidad forestal, 2) comparar el desempeño de algoritmos no paramétricos del vecino más cercano que incluyen variaciones en la matriz de ponderación de la distancia espectral, 3) generar mapas de biomasa aérea con el mejor estimador no paramétrico del k-vecino más cercano y finalmente, 4) validar las estimaciones con estudios reportados.

4.2. Materiales y Métodos

4.2.1. Área de Estudio

El estado de San Luis Potosí se localiza en la porción centro norte de México y abarca una superficie de 6'116,360.91 ha. Se ubica entre los paralelos 21° 07' N - 24° 22' N y, 98° 20' O y 102° 17' O (Figura 4.1); El estado se encuentra en el límite norte de la zona intertropical, cuenta con una riqueza de paisajes y condiciones de vegetación que puede ser comprendida analizando las zonas geográficas que lo componen: (1) Altiplano, (2) Centro, (3) Media y (4) Huasteca (INAFED, 2005). Esta distribución de ecosistemas obedece a la conformación orográfica, al patrón de distribución de precipitación que proviene del Golfo de México y a la relación inversa existente entre la temperatura y la altitud propia de estas latitudes. Los vientos cargados de humedad que precipitan en la Sierra Madre Oriental (Zona Media y Huasteca) y aunado al gradiente altitudinal, propician formaciones vegetales de clima tropical (selva) y templado (bosques); mientras que los vientos avanzan continente adentro, la vegetación va cambiando progresivamente hasta observar vegetación propia de climas áridos *v.gr.* matorral (Figura 4.2).

4.2.2. Estimación de la Biomasa con Datos de Campo

Después de realizar un análisis de control de calidad en los datos, se seleccionaron 235 conglomerados de muestreo de la colección del IFN levantados en el Estado de San Luis Potosí (Figura 4.2) donde la información dasométrica incluyó básicamente diámetro normal, altura, especie. Cada uno de los conglomerados se compone de 4 sitios de muestreo de 400 m² sumando 1600 m² de área (CONAFOR, 2010). Las coordenadas de los sitios de muestreo se encontraban bajo la proyección Plate-Carrée y fueron proyectados al sistema de coordenadas UTM-14n datum WGS84. Para el cálculo de la biomasa aérea de cada una de las 364 especies encontradas en la base de datos del IFN para el Estado de San Luis Potosí, se emplearon 25 ecuaciones

diferentes reportadas en Nívar (2009). Las formas matemáticas de las ecuaciones varían para cada autor y especie *v.gr.* lineal, exponencial, potencial; en las cuales se ingresa el diámetro normal (cm) y la altura del individuo (m) para obtener la biomasa total individual (kg). Los datos de biomasa arbórea aérea individual (20731 árboles) fueron sumados por parcela y estimados a la hectárea para ser relacionados con las variables predictoras. El procesamiento de los datos se llevó a cabo en el software SAS (SAS Institute Inc., 2004).

4.2.3. Variables Predictoras

Se empleó el producto MCD43A4 que se obtiene del sensor MODIS instalado en los satélites Aqua y Terra; se le denomina reflectancia (ρ) ajustada al nadir (MCD43A4) con una función de distribución de reflectancia bidireccional; integra datos multifecha tomados durante periodos de 16 días y se encuentra corregido atmosféricamente y libre de nubes. Tiene una resolución espacial de 500 m y una resolución espectral de siete bandas distribuidas en las siguientes longitudes de onda: 1. Rojo (620 - 670 nm), 2. Infrarrojo cercano (841 - 876 nm), 3. Azul (459 - 479 nm), 4. Verde (545 - 565 nm), 5. Infrarrojo medio₁ (1230 - 1250 nm), 6. Infrarrojo medio₂ (1628 - 1652 nm), 7. Infrarrojo medio₃ (2105 - 2155 nm) (Bhaduri *et al.*, 2010). Para el análisis se usaron datos del periodo 11 al 26 de febrero de 2006 (día juliano 42 al 57) los cuales fueron obtenidos de la interface denominada Modis Reprojection Tool (MRTWeb) que combina la posibilidad de búsqueda de información con capacidades de preprocesamiento de datos espectrales (<http://lpdaac.usgs.gov/datapool/datapool.asp>) (Daucsavage *et al.*, 2010). Estos datos fueron obtenidos para las coordenadas extremas del área de interés y fueron re proyectados al sistema UTM-14N. Se construyó un índice normalizado de vegetación denominado NDVI₆₂, calculado como $NDVI_{62} = (\rho_{irm2} - \rho_{irc}) / (\rho_{irm2} + \rho_{irc})$. Este índice es sensible al estrés hídrico de la vegetación y guarda una correlación negativa con la biomasa forestal (Aguirre-Salado *et al.*, 2009). También se empleó el producto MODIS denominado campos continuos

de vegetación versión 3, colección 3 (MOD44B) (VCF, por sus siglas en inglés) (500 m de resolución espacial) que indica la cobertura proporcional (%) para vegetación arbórea, vegetación herbácea y suelo desnudo y se encuentra en tres capas de información, respectivamente (Hansen *et al.*, 2003). En este trabajo se emplearon los datos VCF correspondientes a la cobertura arbórea.

La Misión Topográfica de Radar en Transbordador (SRTM) que fue llevada a cabo por la Agencia Espacial Aeronáutica Nacional (NASA) cuenta con datos de elevación para el 80% del globo terrestre, los cuales pueden obtenerse en el sitio del Consorcio para la Información Espacial del Grupo Consultivo para la Investigación Agrícola Internacional (CGIAR-CSI) disponible en <http://srtm.csi.cgiar.org/>. Los datos SRTM se encuentran a una resolución espacial de 1" de arco (aproximadamente 90 m) con un error vertical 16 m, y una sensibilidad de 1 m entre píxel (Huggel *et al.*, 2008; Getirana *et al.*, 2010). Se descargaron datos altimétricos correspondientes a la zona de estudio, mediante la especificación de las coordenadas del área de interés (lat-long). El archivo obtenido se proyectó al sistema de coordenadas UTM, zona 14, datum WGS84. Las variables topográficas que se obtuvieron con el conjunto de datos SRTM fueron como sigue: 1) elevación, 2) grado de la pendiente, 3) orientación de la pendiente (azimuth 0 – 360°), 4) desviación estandar de la orientación de la pendiente y la distancia horizontal al escurrimiento más cercano (Blackard *et al.*, 2008) que se obtuvo a partir de la cartografía vectorial escala 1:250,000 (INEGI).

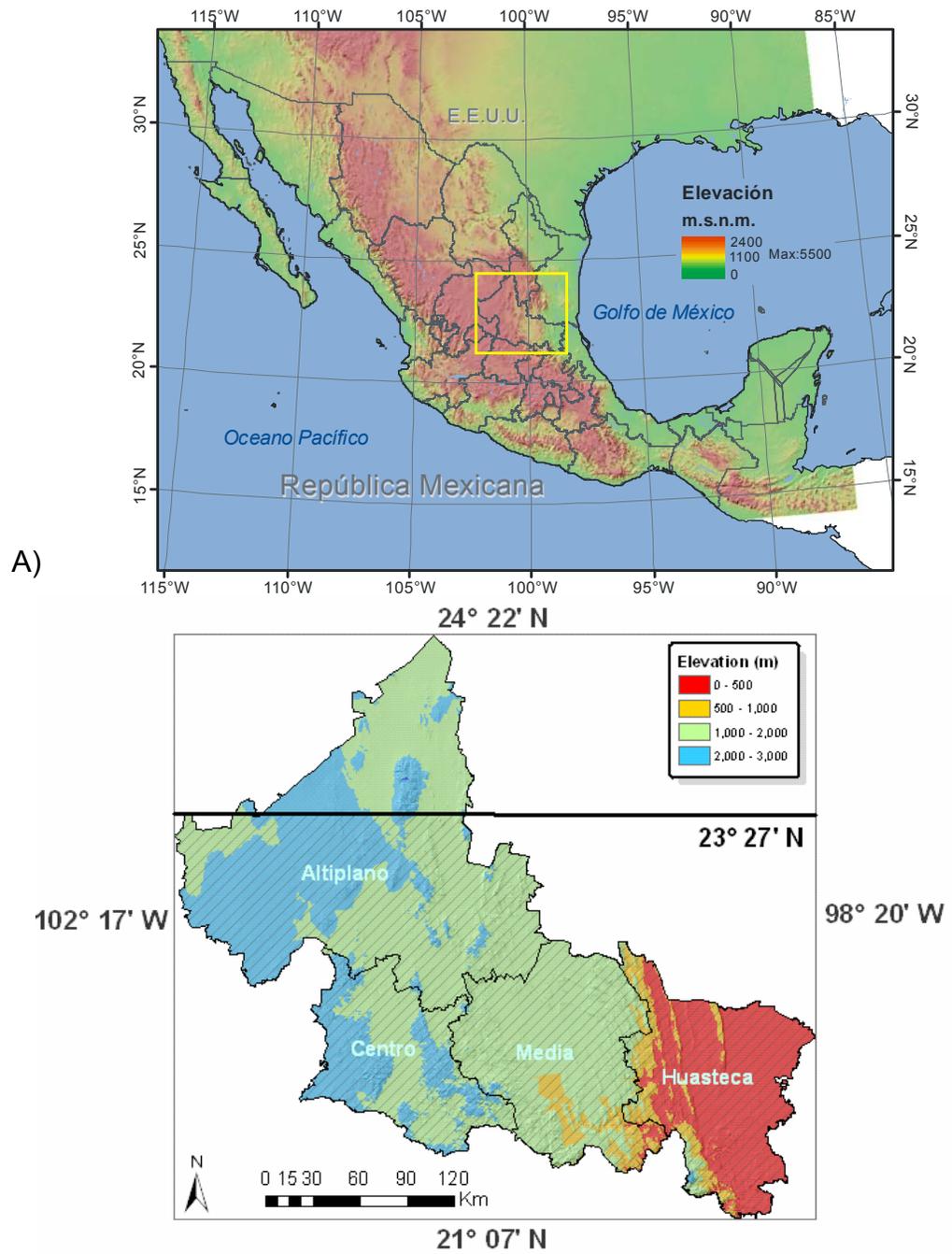


Figura 4.1. Localización del área de estudio: A) Contexto nacional y B) Mapa topográfico y zona intertropical (hachurado).

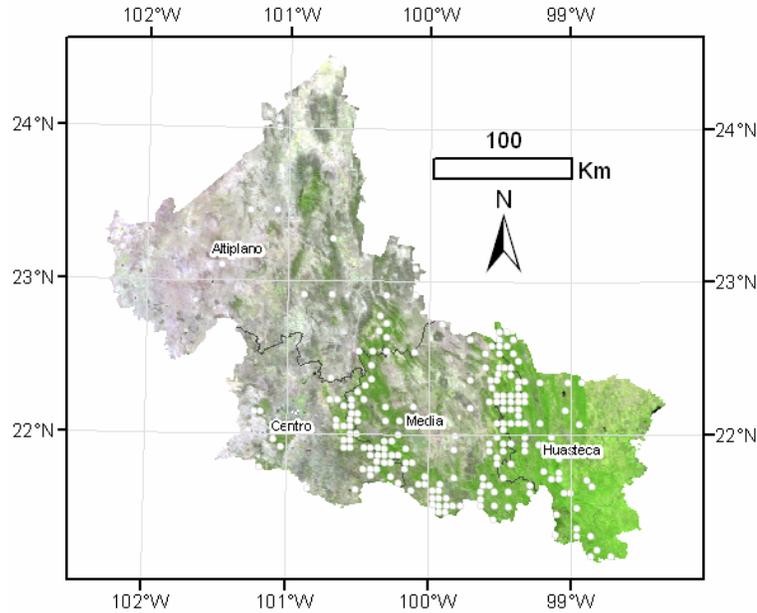


Figura 4.2. Distribución espacial de los sitios de muestreo seleccionados del Inventario Forestal Nacional en el área de estudio.

Como variables predictoras de tipo climático se emplearon datos de precipitación media anual (PMA) y temperatura mínima del mes más frío (TMF) de 308 estaciones climatológicas ubicadas dentro del área de estudio. Estas variables tienen un significado biológico y ecológico (Chen *et al.*, 2007). Los datos fueron obtenidos del Extractor Rápido de Información Climática (ERIC) (IMTA, 2006); para fines de espacialización, estos datos se interpolaron mediante el método Kriging ordinario (modelo esférico). El error medio cuadrático obtenido para la PMA fue de 182.7 mm ($r^{**}=0.91$ observaciones vs predicciones), mientras que para la TMMF fue de 1.61° ($r^{**}=0.79$), respectivamente. Todas las variables predictoras (satelitales, topográficas y climáticas) fueron estandarizadas a la resolución espacial del producto MCD43A4 (500 m). También se empleó el mapa de vegetación y uso de suelo serie IV (2007) generado por el Instituto Nacional de Estadística y Geografía de México (INEGI, 2009) para obtener los promedios de la biomasa arbórea aérea (estimada para cada píxel) por cada tipo de vegetación; esto con el propósito de realizar algunas comparaciones con estimaciones reportadas en literatura reciente.

4.2.4. Procesamiento de la Información

Se empleó el algoritmo kNN que realiza una interpolación entre los k vecinos espectralmente más cercanos. Para ello calcula la distancia espectral entre el píxel que se desea estimar a todos los píxeles de la muestra, posteriormente se ordenan de menor a mayor distancia para seleccionar los k vecinos espectralmente más cercanos, y finalmente se realiza un promedio que se pondera de manera inversa a tal distancia espectral entre los k vecinos más cercanos (Franco-López et al., 2001; Gjertsen et al., 2007; Barth et al., 2009; Baffeta et al., 2009, 2011; Breidenbach et al., 2011). Una de las bondades de este método radica en que no se requieren suposiciones estadísticas sobre la naturaleza de los datos *v.gr.* normalidad, haciéndolo versátil y robusto. Para el cálculo de la distancia espectral se empleó la fórmula $D(p, q, W) = \sqrt{[(p - q)^T W (p - q)]}$ donde W es la matriz de ponderación, con cuatro variantes que se muestran en la Tabla 4.1. Una motivación para la realización de esta prueba fue la hipótesis que indica que el emplear una distancia espectral estandarizada, minimiza el efecto de las unidades de medida de cada variable explicatoria (X's) y problemas de multicolinealidad sobre las estimaciones.

Tabla 4.1. Estimación de la matriz de ponderación para la estandarización de las distancias espectrales (Crookston y Finley, 2008).

Método	Valor de W
Raw	Matriz identidad, I.
Euclideana	Inversa de la suma directa de la matriz de covarianza de las X.
Mahalanobis	Inversa de la matriz de covarianza de las X.
MSN	$W = \Gamma \Lambda^2 \Gamma^T$, donde Γ es la matriz de vectores canónicos correspondientes a las X's y es obtenida mediante el análisis de correlación canónica entre las X y las Y, y Λ es la matriz de correlación canónica.

Debido a que la selección de un conjunto apropiado de variables explicatorias resulta un tanto complicado, se decidió evaluar de manera heurística tanto el desempeño predictivo de cada grupo de variables, como la determinación del k óptimo; es decir, a partir del cálculo de la raíz de la diferencia cuadrática media (RMSD) generado en la validación cruzada dejando uno fuera, para una secuencia desde $k = 1$ hasta 30 (Chirici *et al.*, 2008; McRoberts, 2008). El RMSD es comúnmente conocido como raíz del cuadrado medio del error (RMSE, por sus siglas en inglés); no obstante, Crookston y Finley (2008) sugieren el uso de cambiar el término de error por diferencia, ya que el error del modelo no resulta de minimizar la raíz cuadrada de las diferencias entre dos conjuntos de datos. Para validar este método de selección de variables se realizó un análisis de regresión por pasos (Stepwise). Algunos autores han empleado este procedimiento para realizar la selección de variables que serán empleadas para la determinación de la distancia espectral en el algoritmo kNN usando un nivel de significancia de $\alpha=0.15$ (Fuchs *et al.*, 2009); para el caso de este trabajo se empleó una significancia de $\alpha=0.05$ y $\alpha=0.01$ como umbral para admitir y remover variables, respectivamente (Gjertsen *et al.*, 2007).

Para el caso del k óptimo varios autores han recomendado la aplicación de un algoritmo de validación cruzada, ya sea dejando uno fuera (Chirici *et al.*, 2008; Aguirre-Salado *et al.*, 2009; Fuchs *et al.*, 2009; Canisius *et al.*, 2010) o dividiendo el conjunto de datos en submuestras de manera iterativa (k-fold cross validation) (García-Pedrajas y Ortiz-Boyer, 2009). Todos ellos registran una estabilización del error cuando se aumenta el número de vecinos más cercanos empleados para ponderar las estimaciones de manera inversa a la distancia (Fuchs *et al.*, 2009; Tian *et al.*, 2012). En este trabajo se asume que cada conjunto de datos (X's y Y's) es diferente, por lo tanto debe existir un cálculo del k óptimo exclusivo para cada conjunto de datos. El procesamiento de la información se realizó en el software R usando la librería yalmpute (Crookston y Finley, 2011).

4.3. Resultados y Discusión

4.3.1. Selección de Variables

El indicador empleado para seleccionar las variables explicatorias fue el RMSD estimado en la validación cruzada dejando uno fuera, como una forma para averiguar el potencial predictivo de las variables explicatorias de cada grupo de variables estudiadas. Las variables explicatorias que presentaron la menor diferencia entre los valores observados y estimados de biomasa arbórea aérea fueron NDVI62 y VCF (de 30 a 36 Mg/ha) (Figura 4.3 A). Posteriormente le siguió el grupo de las variables climáticas (Figura 4.3 B), donde la temperatura del mes mas frío presentó el RMSD más bajo de su grupo (≈ 48 Mg/ha); finalmente le siguió el grupo de variables topográficas (Figura 4.3 C) destacando la pendiente (≈ 49 Mg/ha) y la variabilidad de la orientación (≈ 51 Mg/ha) que mostraron la menor diferencia de dicho grupo. También se compararon tales desempeños predictivos por grupos de variables siendo evidente la disminución del RMSD para el caso de las variables satelitales (NDVI62 y VCF) (Figura 4.3 D).

La selección de variables realizada mediante la comparación del RMSD obtenido de la validación cruzada dejando uno fuera, se validó exitosamente mediante la aplicación del procedimiento Stepwise (regresión por pasos), con base en los niveles de significancia indicados anteriormente ($\alpha=0.05$ y $\alpha=0.01$ para admitir y remover variables, respectivamente) (Fuchs *et al.*, 2009). Por lo que, las únicas variables explicatorias que cumplieron estos criterios fueron el NDVI62 y VCF ($p < 0.0001$); mientras que las variables restantes (*i.e.* climáticas y topográficas) no fueron estadísticamente significativas.

La investigación reportada sobre el potencial de los índices espectrales de vegetación en la estimación de la densidad forestal, había sido acaparada por índices construidos basados solamente en el rango del rojo-infrarrojo. No

obstante, en literatura actual (2007 a la fecha) es más común encontrar trabajos que revisen el potencial de índice espectrales basados en el infrarrojo de onda corta (SWIR) que es sensible a los cambios de humedad en la vegetación (Gjertsen *et al.*, 2007; Fuchs *et al.*, 2009 Wolter *et al.*, 2009). Por ejemplo, un estudio interesante es el de Gjertsen *et al.* (2007) quienes estudiaron algunas variables como volumen de madera (m^3/ha) usando datos espectrales Landsat TM. Las variables espectrales que mejor predijeron la densidad fueron la TM_4 (NIR) y TM_7 (SWIR). Estos resultados soportan los obtenidos en este trabajo, ya que precisamente dichas bandas espectrales se pueden comparar con las bandas del producto MODIS MCD43A4 usadas para construir el índice espectral de vegetación ($NDVI_{62}$) utilizado en esta investigación. Otro estudio relacionado es el de Aguirre-Salado *et al.* (2009) quienes emplearon el $NDVI_{42}$ (índice de estrés hídrico), construido con datos del sensor SPOT5 HRG (10 m de resolución espacial), ya que fue el más apropiado para describir la densidad forestal en bosques de coníferas. Tal índice es completamente equivalente con el $NDVI_{62}$ presentado en esta investigación, por lo tanto comparable. Otro estudio relacionado es el de Fuchs *et al.* (2009) quienes emplearon datos del sensor ASTER (15 m de resolución espacial) para estudiar el carbono arbóreo en la tundra siberiana. Las mejores correlaciones entre los datos espectrales fueron obtenidas empleando las bandas del infrarrojo de onda corta (SWIR) con correlaciones que varían entre ± 0.52 a ± 0.57 . En otro trabajo se acepta la importancia del poder predictivo del infrarrojo de onda corta para mejorar las correlaciones en las estimaciones (Canisius *et al.*, 2010).

Blackard *et al.* (2009) estimaron la biomasa en Estados Unidos empleando datos espectrales y auxiliares (topográficos, climáticos, tipos de suelo) y árboles de regresión. Los datos espectrales obtenidos a partir de datos MODIS siempre mostraron desempeños adecuados para explicar el comportamiento de la biomasa forestal. Cruz-Leyva *et al.* (2010) modelaron algunas variables de densidad forestal como área basal y volumen en un bosque de *Pinus patula* y *P. teocote* en Hidalgo, México usando variables

espectrales del satélite SPOT 5 HRG e información auxiliar (datos climáticos y topográficos). Ellos concluyen que los mejores modelos fueron los desarrollados usando datos espectrales. Una de las motivaciones para el empleo de variables auxiliares es el supuesto incremento en la precisión de las estimaciones. No obstante, con este trabajo y algunos estudios previos se confirma que para el caso de la estimación de la densidad forestal, la mejora en la precisión cuando se emplean variables auxiliares no resulta necesariamente cierto. Sin embargo, para el caso de investigaciones enfocadas diferente, *v.gr.*, para el caso de la determinación de hábitats potenciales para especies vegetales y animales, particularmente con el análisis de regresión logística que determina probabilidades, el uso de variables auxiliares tales como climáticas o topográficas sigue siendo relevante y válido (Chen *et al.*, 2007; Dapporto, 2010; Hongoh *et al.*, 2012).

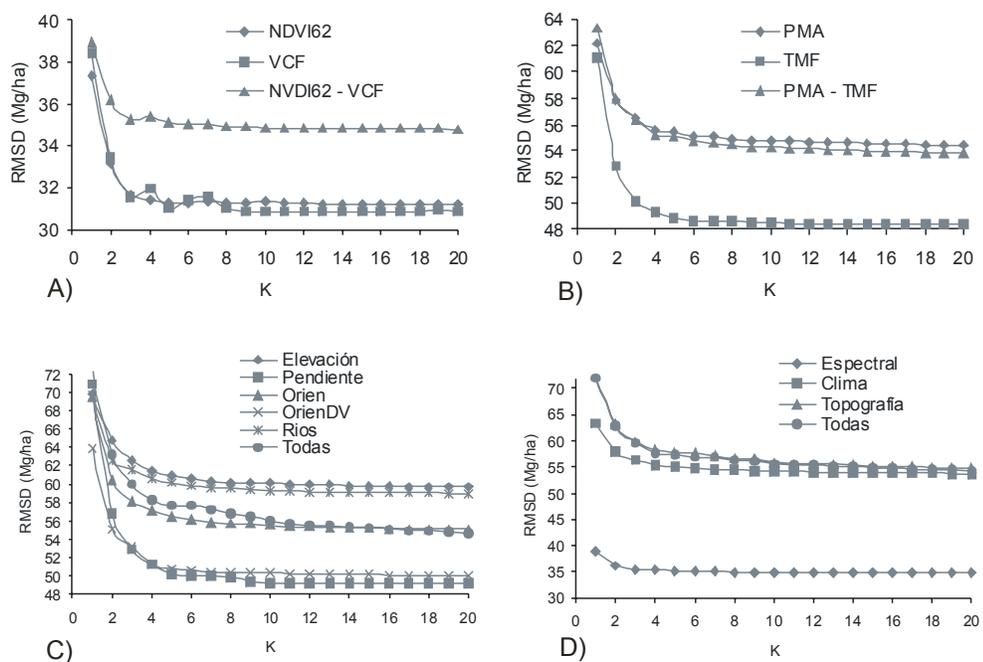


Figura 4.3. Selección de variables con base en la validación cruzada y su efecto en el RMSD para el algoritmo kNN basado en la distancia euclideana sin estandarizar. A) variables espectrales, B) variables climáticas, C) variables topográficas y D) todas las variables.

4.3.2. Comparación de Métodos

En la Figura 4.4 se puede observar una pequeña disminución adicional en el RMSD cuando se emplea alguna de las distancias estandarizadas (EUC, euclídeana; MAH, mahalanobis; y, MSN: vecino más similar) para la búsqueda de los vecinos más cercanos usando las variables satelitales. El comportamiento de las distancias estandarizadas fue similar; sin embargo, la distancia EUC, fue la que alcanzó el RMSD más bajo (RMSD=29.57 Mg/ha, k=10), le siguió la distancia MAH (RMSD=29.91 Mg/ha, k=6), posteriormente la distancia MSN (RMSD=30.56 Mg/ha, k=14), y finalmente las distancias euclídeanas sin estandarizar (RAW) (RMSD=31.69Mg/ha, k=11). Con el propósito de averiguar si existen diferencias estadísticamente significativas entre los RMSD obtenidos en la validación cruzada para cada método, se realizó una comparación de medias con la prueba de Duncan ($\alpha=0.05$) (García-Villalpando *et al.*, 2001). Esto mostró resultados interesantes, ya que no se encontraron diferencias estadísticamente significativas entre los métodos que emplean distancias estandarizadas (EUC, MAH y MSN); pero, sí se detectaron diferencias contra el método RAW que emplea las distancias sin estandarizar. Es decir, según la prueba de Duncan, los métodos empleados se pueden clasificar en dos grupos: Grupo 1, que incluye cualquiera de las distancias estandarizadas: EUC, MSN y MAH; y Grupo 2, donde queda el método que emplea las distancias sin estandarizar.

Los k óptimos seleccionados (k = 6, 10, 11 y 14) para este trabajo se encuentran dentro del rango que se reporta en la literatura ($5 \geq k \leq 20$) (Gjertsen *et al.*, 2007; Chirici *et al.*, 2008; Kajisa *et al.*, 2008; Aguirre-Salado *et al.*, 2009; Baffeta *et al.*, 2009; Fuchs *et al.*, 2009; McRoberts, 2009; Breidenbach *et al.*, 2011). Mientras que los valores de RMSD obtenidos en esta investigación, a nivel de píxel, durante la validación cruzada varían desde 29-32 Mg/ha (49-52%). Estos errores a nivel píxel coinciden con otros trabajos que también han estudiado la biomasa forestal (Mg/ha) recientemente. Por ejemplo, Blackard *et*

al. (2008), usaron técnica de árboles de regresión para mapear la biomasa forestal en Estados Unidos y emplearon varios tipos de datos del sensor MODIS (MOD09v003, MOD13v003, MOD44), datos climáticos, topográficos y datos de inventario forestal. Ellos obtuvieron un error relativo (%), a nivel píxel, que varió entre 0.51 (51%) en el territorio de Puerto Rico hasta 0.92 (92%) en la parte sur de Estados Unidos. Labrecque *et al.* (2006) estimaron la biomasa forestal en bosques de Canadá con el algoritmo kNN usando datos Landsat TM y datos de inventario. El RMSE que obtuvieron usando 5 vecinos más cercanos, fue de 54 Mg/ha. Reese *et al.* (2002) emplearon el algoritmo kNN para estimar la biomasa en bosques de Suecia usando datos espectrales del sensor Landsat 5 TM y SPOT3. El RMSE obtenido a nivel píxel, empleando 5 vecinos más cercanos, resultó entre 53 y 79%. Otros autores que trabajaron con áreas más pequeñas y sensores más detallados, encontraron menores valores de RMSE. Tian *et al.* (2012) mapearon la biomasa forestal con el método kNN (distancia de Mahalanobis) y datos SPOT 5 HRG; su mejor combinación de variables explicativas fue con los primeros dos componentes principales, el índice infrarrojo, y la elevación (información auxiliar); obtuvieron un RMSE bajo (20.7 Mg/ha). Estos autores disminuyeron aun más su RMSE, ya que usaron datos LIDAR, llegando a un RMSE de 15.2 Mg/ha. Estas precisiones se obtienen con sensores de muy alta resolución espacial pero con el inconveniente de que su aplicación resulta difícil para áreas considerables (>1 millón ha).

Para complementar el presente análisis se estimó la correlación (observados vs estimados) para conocer el comportamiento predictivo de cada método empleado. Las correlaciones obtenidas en los cuatro métodos fueron altamente significativas y son como sigue: MSN ($r = 0.90$), MAH ($r = 0.88$), EUC ($r = 0.86$) y RAW ($r = 0.84$). Cuando se analiza la línea 1:1 en la Figura 4.5 se puede observar una subestimación en las predicciones para el caso del método RAW; mientras que para las demás distancias se registra un mejor ajuste, alcanzando el máximo en el método MSN.

Revisando los métodos vale la pena notar que al emplear la distancia euclídeana (RAW), las variables no están estandarizadas, provocando que la variable explicatoria que tenga mayor varianza domine en la estimación. Cuando se emplea la distancia de mahalanobis (MAH), las variables ya se encuentran estandarizadas, tomando en cuenta la estructura de correlación entre las variables mediante la inversa de la matriz de covarianzas de las X's; mientras que cuando se emplea la distancia MSN, al usar la matriz de ponderación que se genera a partir del análisis de correlación canónica, además de que las variables se encuentran estandarizadas, se maximiza la correlación entre la variable de respuesta y las variables explicatorias (Loog *et al*, 2005). Esto explica porqué es más recomendable trabajar con variables estandarizadas y no correlacionadas, especialmente cuando lo que se calculan son distancias. Por lo que, la distancia MSN, al quitar estos dos inconvenientes en los datos, resulta una buena opción para implementar la estimación mediante el vecino más cercano y esto explica el buen comportamiento del método MSN (Figura 4.5 D).

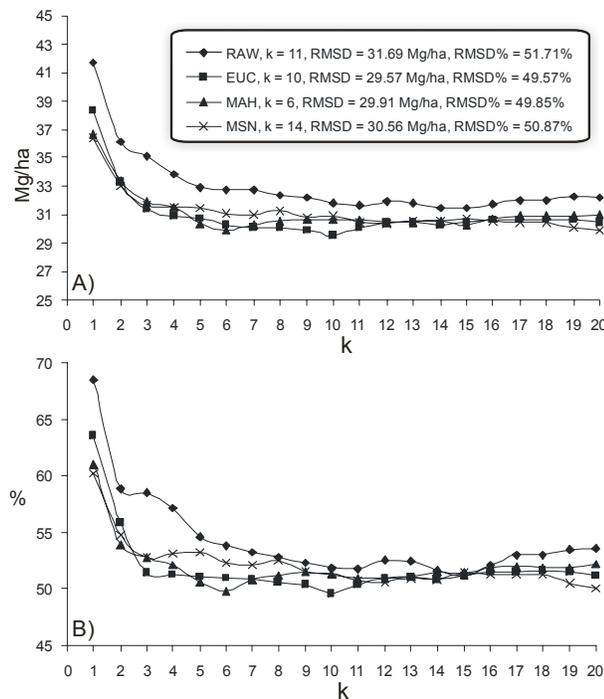


Figura 4.4. Comparación de métodos no paramétricos entrenados con las variables satelitales: NDVI62 y VCF. A) RMSD absoluto y B) RMSD relativo.

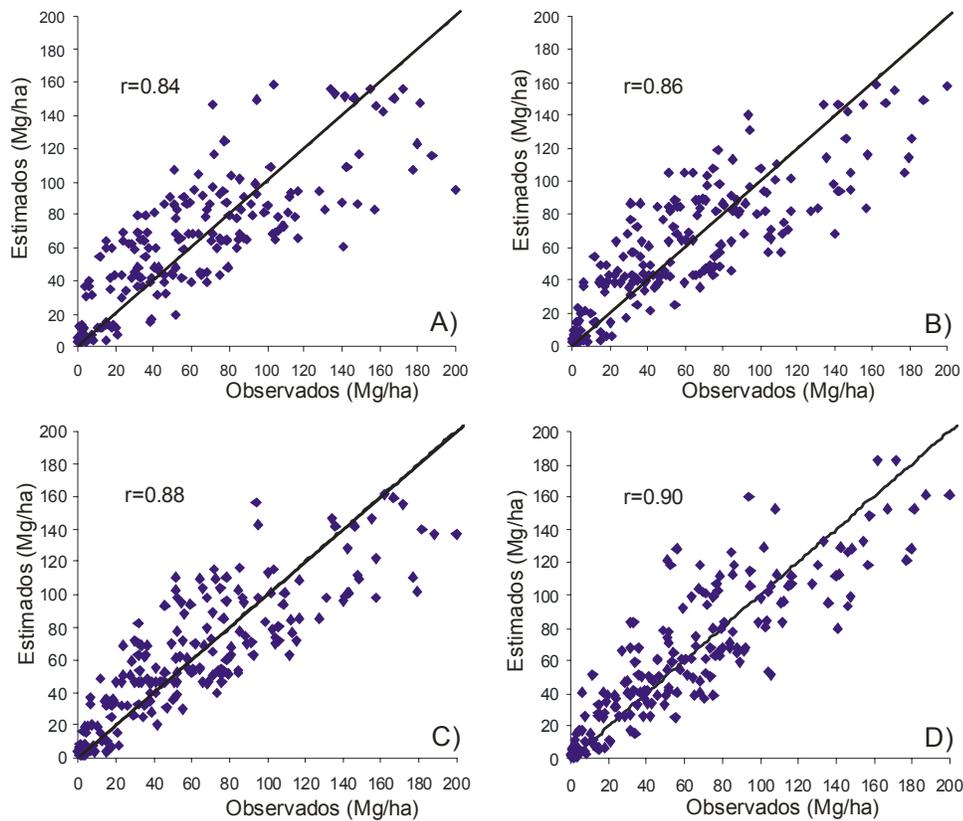


Figura 4.5. Desempeño de los diferentes métodos no paramétricos: A) Raw, B) Euclideana, C) Mahalanobis y D) MSN.

4.3.3. Estimación de la Biomasa

El mapa seleccionado para estimar la biomasa arbórea aérea (Mg/ha) fue el que se obtuvo usando la distancia MSN (Figura 4.6). Para comparar las estimaciones obtenidas con las reportadas en la literatura científica, se estimó el promedio de biomasa arbórea aérea para cada tipo de vegetación usando los píxeles de la imagen de satélite. La biomasa arbórea aérea promediada por formación vegetal es como sigue: bosques (78 Mg/ha), selvas (137.8 Mg/ha) y matorral (7.5 Mg/ha). En la Tabla 2 se presentan los diferentes promedios de biomasa arbórea aérea por tipo de vegetación y además una comparación con las estimaciones de biomasa realizada por De Jong *et al.* (2010) contra los tipos de vegetación más comparables con los del presente trabajo. Ellos determinaron contenidos de biomasa promedio para cada tipo de vegetación en

México usando datos de Inventario Nacional Forestal y la cartografía vectorial (1993-2002) (sin Teledetección). Aunque sus resultados son comparables con los de este trabajo, ellos presentan algunas subestimaciones que probablemente se deban a las diferentes metodologías empleadas por ambos estudios. De Jong *et al.* (2010) emplearon ecuaciones alométricas por default (agrupadas por niveles de precipitación) que arrojaran estimaciones más conservadoras para evitar problemas de sobreestimación de la biomasa. Mientras que en este trabajo se emplearon estimaciones de biomasa arbórea aérea individual que se realizaron a nivel de género y especie, empleando 25 ecuaciones reportadas por Návar *et al.* (2009). Ésta es una de las causas probables de tales discrepancias en la información.

El bosque de coníferas (*Juniperus spp*) encontrado en el área de estudio presentó contenidos bajos de biomasa arbórea aérea (42 Mg/ha). Estas especies son de porte bajo tienen hábitos de crecimiento propios de climas áridos (≈ 600 mm de precipitación). Powell *et al.* (2010) reportan valores de biomasa para bosques templados de Arizona y Minnessota que varían desde 45 Mg/ha (150-880 mm de precipitación) hasta 60 Mg/ha (490-1090 mm de precipitación). El bosque de encino (*Quercus spp*) presentó mayores contenidos promedio de biomasa arbórea aérea de 70.9 Mg/ha y se puede encontrar en las partes elevadas de las zonas montañosas, con una mayor tendencia hacia la dirección este, ya que desde ahí provienen vientos cargados de humedad del Golfo de México (≈ 1000 mm de precipitación). De la categoría de bosque, el bosque mesófilo de montaña fue la vegetación que mayor contenido de biomasa arbórea aérea registró (121.1 Mg/ha). Este tipo de vegetación presenta una diversidad de especies arbóreas de porte alto, principalmente de los géneros *Quercus spp* y *Liquidambar spp* (Valdez-Tamez *et al.*, 2003). Acosta Mireles *et al.* (2009) determinaron los contenidos de carbono en bosques mixtos de *Pinus patula* en Tlaxcala, México. Para la parte aérea de dichos ecosistemas encontraron valores entre 118.3 Mg/ha (236.6 Mg/ha de biomasa) y 77.2 Mg/ha (144.4 Mg/ha de biomasa).

El matorral xerófilo se compone de especies, cuya morfología y fisiología, se encuentra adaptada a largos periodos de escasez de agua (<600 mm de precipitación). Aquí se incluyen comunidades vegetales como el matorral desértico micrófilo (*Larrea spp*), matorral desértico rosetófilo (*Agave spp*), matorral desértico crassicaule (*Opuntia spp*, *Stenocereus spp*), mezquital (*Prosopis spp*). La biomasa acumulada en este tipo de ecosistemas fue en promedio de 7.5 Mg/ha.). Lasserre *et al.* (2011) estimaron los contenidos de biomasa para un matorral en Italia empleando imágenes de satélite y el estimador no paramétrico kNN. Los valores variaron entre 0.5 y 4 Mg/ha.

La selva perennifolia fue el tipo de vegetación que presentó los mayores contenidos de biomasa arbórea aérea, con un promedio de 152.7 Mg/ha, en un rango de 1.9 - 219.2 Mg/ha. La gran variación de los valores estimados a nivel píxel obedece a la considerable fragmentación del ecosistema. Normalmente, las selvas se encuentran bajo una fuerte presión antrópica, lo que determina altos índices de fragmentación del habitat. Para comprender mejor tales variaciones, es necesario estudiar con mayor profundidad el patron espacial de los claros en la selva y determinar una medida cuantitativa del nivel de fragmentación de estos ecosistemas. Dado que las estimaciones se realizan para cada píxel, con base en las variables satelitales, la desviación estandar es considerable (Gjertsen *et al.*, 2007); esto se debe a la variación natural del ecosistema captada de manera espacialmente explicita por el sensor remoto.

Las estimaciones de biomasa arbórea aérea obtenidas en esta investigación (Figura 4.6) se encuentran entre los rangos establecidos en la literatura y las variaciones detectadas obedecen principalmente a: 1) factores ambientales (precipitación y temperatura) y 2) factores antrópicos que determinan el grado de fragmentación de los diferentes tipos de vegetación (Silva-Arredondo y Návar-Chaidez, 2009). Los factores ambientales determinan la productividad de un sitio determinado a una escala general; mientras que los

factores antrópicos son evidentes a escalas espaciales más pequeñas, que se refleja en la variabilidad entre píxeles individuales.

Tabla 2. Estimaciones promedio de biomasa para cada tipo de vegetación obtenidas con el kNN (distancia MSN) (RMSD=30.56 Mg/ha, k=14, r = 0.90).

Tipo de vegetación	Area (ha)	Promedio (Mg/ha)	STD	Biomasa total (Mg)	Biomasa De Jong et al. (2010)
Bosque de coníferas	116,250	42.0	45.2	4,880,300.5	94
Bosque de encino	440,925	70.9	60.2	31,246,247.6	62
Bosque mesófilo de montaña	7,975	121.1	49.0	965,570.5	104
Matorral xerófilo	3,009,300	7.5	16.5	22,469,501.3	12-40
Selva caducifolia	185,825	122.9	64.2	22,844,075.3	38
Selva perennifolia	175,725	152.7	61.3	26,831,917.6	104
Total	3,936,000			109,237,612.8	

STD: Desviación estándar.

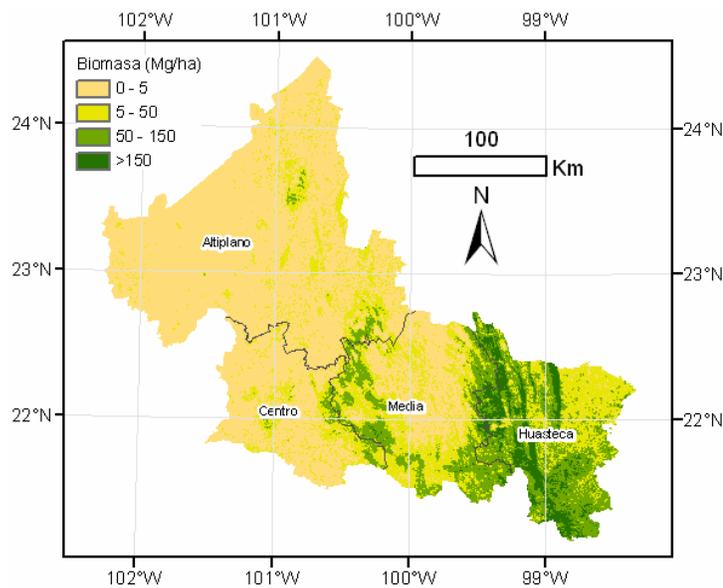


Figura 4.6. Mapa de biomasa arbórea aérea obtenido con el algoritmo kNN empleando la distancia MSN (RMSD=30.56 Mg/ha, k=14, r = 0.90): San Luis Potosí, México.

4.4. Conclusiones

En este capítulo se compararon diferentes métodos no paramétricos del vecino más cercano (RAW, EUC, MAH y MSN), así como diferentes combinaciones de variables explicatorias en la estimación de la biomasa forestal en una porción de la región Centro Norte de México. Los análisis realizados sugieren que las variables NDVI62 (construido mediante el cociente normalizado de las diferencias entre la banda sensible al infrarrojo de onda corta e infrarrojo cercano) y VCF (Campos Continuos de Vegetación), tuvieron el mejor desempeño predictivo. Se observó que agregar variables predictoras auxiliares no presentó mejoras en la capacidad predictiva del algoritmo. Después de aplicar la prueba de Duncan a los errores obtenidos para cada método, se encontró que basándose en su desempeño, los métodos estudiados se pueden agrupar en dos: 1) RAW (distancias sin estandarizar) y 2) EUC, MAH y MSN. Finalmente, el método MSN presentó la mayor correlación entre los valores observados y estimados de biomasa ($r = 0.9$), siendo evidente el efecto que tiene el análisis de correlación canónica en la maximización de la correlación entre las variables explicatorias y la variable de respuesta. Se puede concluir que las variaciones promedio de la biomasa arbórea aérea en los matorrales, bosques y selvas de San Luis Potosí se encuentran entre 5 y 150 Mg/ha, generando un total almacenado de 109.2 Millones Mg (109,237,612.8 Mg).

4.5. Literatura Citada

- Aguirre-Salado, C.A., Valdez-Lazalde, J.R., Ángeles-Pérez, G., De los Santos-Posadas, H.M., Happanen, R., Aguirre-Salado, A.I. 2009. Mapeo de carbono arbóreo aéreo en bosques manejados de *Pinus patula* en Hidalgo, México. *Agrociencia* 43:209-220.
- Aguirre-Salado, C.A., Valdez-Lazalde, J.R., Ángeles-Pérez, G., de los Santos-Posadas, H.M., Aguirre-Salado, A.I. 2011. Mapeo del índice de área foliar

- y cobertura arbórea mediante fotografía hemisférica y datos SPOT 5 HRG: Regresión y K-nn. *Agrociencia*. 45(1):105-119.
- Anaya, J.A., Chuvieco, E., and Palacios-Orueta, A. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management*. 257:1237-1246.
- Baffeta, F., Fattorini, L., Franceschi, S. and Corona, P. 2009. Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. *Remote Sensing of Environment*. 113:463–475.
- Barth, A., Wallerman, J. and Stahl, G. 2009. Spatially consistent nearest neighbor imputation of forest stand data. *Remote Sensing of Environment*. 113:546-553.
- Bhaduri, K., Das, K. and Votava, P. 2010. Distributed anomaly detection using satellite data from multiple modalities. *Proceedings of The 2010 Conference on Intelligent Data Understanding (CIDU-NASA)*. San Francisco Bay Area. Oct 5-7. pp.109-123.
- Blackard., J.A., Finco, M.V., Helmer, E.H., Holden, E.H., Hoppus, M.L., Jacobs, D.M., Lister, A.J., Moisen, G.G., Nelson, M.D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D.L., Winterberger, K.C., Brandeis, T.J., Czaplewski, R.L., McRoberts, R.E., Patterson, P.L. and Tymcio, R.P. 2008. Mapping U.S. forest biomass using nation wide forest inventory data and moderate resolution information. *Remote Sensing of Environment*. 112:1658-1677.
- Breidenbach, J., Naesset, E. and Gobakken, T. 2011. Improving k-nearest neighbor predictions in forest inventories by combining high and low density airborne laser scanning data. *Remote Sensing of Environment*. In press.
- Canisius, F., Fernandes, R. and Chen, J. 2010. Comparison and evaluation of Medium Resolution Imaging Spectrometer leaf area index products across a range of land use. *Remote Sensing of Environment*. 114:950–960.

- Chen, H., Chen, L. and Albright, T.P. 2007. Predicting the potential distribution of invasive exotic species using GIS and information-theoretic approaches: A case of ragweed (*Ambrosia artemisifolia* L.) distribution in China. Chinese Science Bulletin. 52(9):1223-1230.
- Chirici, G., Barbati, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F. and Bertini, R. 2008. Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. Remote Sensing of Environment 112: 2686-2700.
- CONAFOR. 2010. Manual de Procedimientos para el muestreo de campo. Remuestreo-2010. Inventario Nacional Forestal y de Suelos. Comisión Nacional Forestal. Secretaría del Medio Ambiente y Recursos Naturales. Zapopan, Jalisco, México. 140 p.
- Crookston, N.L. and Finley, A.O. 2008. yalImpute: An R Package for kNN Imputation. Journal of Statistical Software. 23(10):1-16.
- Crookston, N.L. and Finley, A.O. 2011. yalImpute: an R package for k-NN imputation. Package "yalImpute". The Comprehensive R Archive Network (CRAN). 40 p.
- Cruz-Leyva, I.A., Valdez-Lazalde, J.R., Ángeles-Pérez, G. y De los Santos-Posadas, H.M. 2010. Modelación espacial de área basal y volumen de madera en bosques manejados de *Pinus patula* y *P. teocote* en el Ejido Atopixco, Hidalgo. Madera y Bosques. 16(3):75-97.
- Dapporto, L. 2010. Predicting distribution of *Zygaena* moths on West Mediterranean islands. Implications for biogeography and conservation (Lepidoptera Zygaenidae). Journal of Insect Conservation. DOI: 10.1007/s10841-010-9318-y.
- Daucsavage, J. Kaminski, M., Ramachandran, B., Jenkerson, C., Sprenger, K., Faust, R. and Rockvam, T. 2010. ASTER and MODIS Land Data Management at the Land Processes, and National Snow and Ice Data Centers. In: Ramachandran, B., Justice, C.O., and Abrams, M.J. Land Remote Sensing and Global Environmental Change. NASA Earth

- Observing System and the Science of ASTER and MODIS. Springer. pp. 167-182.
- De Jong, B., Anaya, C., Masera, O., Olguin, M., Paz, F., Etchevers, J., Martínez, R.D., Guerrero, G., Balbontín, C. 2010. Greenhouse gas emissions between 1993 and 2002 from land-use change and forestry in Mexico. *Forest Ecology and Management*. 260(10): 1689-1701.
- De Leeuw, J., Georgiadou, Y., Kerle, N., De Gier, N., Inoue, Y., Ferwerda, J., Smies, M. and Narantuya, D. 2010. The Function of Remote Sensing in support of environmental policy. *Remote Sensing*. 2:1731-1750.
- FAO. 2009. State of The World's Forests. Food and Agriculture Organization of the United Nations. Roma, Italy. 152 p. ISBN: 978-92-5-106057-5.
- Franco-López, H., Alan, Ek, A.R. and Bauer, M.E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment* 77:251– 274.
- Finley, A.O. and McRoberts, R.E. 2008. Efficient k-nearest neighbor searches for multi-source forest attribute mapping. *Remote Sensing of Environment*. 112:2203–2211.
- Fuchs, H., Magdon, P., Klein, C. and Flessa, H. 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: combining satellite imagery and field inventory. *Remote Sensing of Environment*. 113:518– 531.
- Gallaun, H., Zanchi, G., Nabuurs, G.J., Hengeveld, G., Schardt, M. and Verkerk, P. 2009. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*. 260(3):252-261.
- García-Pedrajas, N. and Ortiz-Boyer, D. 2009. Boosting k-nearest neighbor classifier by means of input space projection. *Expert Systems with Applications*. 36:10570–10582.
- García-Villalpando, J., Castillo-Morales, A., Ramirez-Guzmán, M.E., Rendón-Sánchez, G. y Larqué-Saavedra, M.U. 2001. Comparación de los

- procedimientos de Tukey. Duncan, Dunnett, HSU y Bechhoffer para la selección de medias. *Agrociencia*. 35:79-86.
- Getirana, A.C.V. 2010. Integrating spatial altimetry data into the automatic calibration of hydrological models. *Journal of Hydrology*. 387:244–255.
- Gjertsen, A.K. 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment*. 110:420–430.
- GOFC-GOLD. 2010. A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals caused by deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. GOFC-GOLD Report version COP16-1. GOFC-GOLD Project Office, Natural Resources Canada, Alberta, Canada. 203 p.
- Hansen, M.C., DeFries, R.S., Townshend, J.R., Carroll, M., Dimiceli, C., Sohlberg, R., 2003. MOD44B: vegetation continuous fields collection 3, version 3.0.0. *Earth Interactions* 1–20.
- Havemann, T. 2009. Measuring and monitoring terrestrial carbon. The state of the science and implications for policy makers. The Terrestrial Carbon Group. The Heinz Center and UN-REDD programme. 67 p.
- Holmgren, P. 2008. Role of satellite remote sensing in REDD. UN-REDD Programme. MRV Working Paper 1. 11 p.
- Hongoh, V., Berrang-Ford, L., Scott, M.E. and Lindsay, L.R. 2012. Expanding geographical distribution of the mosquito, *Culex pipiens*, in Canada under climate change. *Applied Geography*. 33:53-62.
- Huggel, C., Scheider, D., Miranda, P.J., Delgado, G.H. and Kääh, A. 2008. Evaluation of ASTER and SRTM DEM data for lahar modeling: a case study on lahars from Popocatépetl Volcano, Mexico. *Journal of Volcanology and Geothermal Research*. 170:99–110.
- IMTA (Instituto Mexicano de Tecnología del Agua). 2006. Extractor Rápido de Información Climatológica III, v. 1.0. Información climatológica disponible en formato electrónico. Jiutepec, Morelos, México.

- INAFED (Instituto Nacional para el Federalismo y el Desarrollo Municipal). (2005). Estado de San Luis Potosí. Enciclopedia de los Municipios de México.
<http://www.e-local.gob.mx/work/templates/enciclo/sanluispotosi/regi.htm>.
- INEGI. 2009. Vegetación y Uso de Suelo Serie IV. Instituto Nacional de Estadística y Geografía e Informática. México.
- Kajisa, T., Murakami, T., Mizoue, N., Kitahara, F. and Yoshida, S. 2008. Estimation of stand volumes using the k-nearest neighbors method in Kyushu, J. Journal of Forest Research. 13:249-254.
- Kaul, M., Dadhwal, V.K. and Mohren, G.M.J. 2009. Land use change and net C flux in Indian forests. Forest Ecology and Management. 258:100–108.
- Köhl, M., Magnussen, S.S., Marchetti, M. 2006. Sampling Methods, Remote Sensing and GIS Multisource Forest Inventory. Springer. p. 190.
- Labrecque, S., Fournier, R., Luther, J. and Piercey, D. 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. Forest Ecology and Management. 226(1-3):129-144.
- Lasserre, B., Chirici, G., Chiavetta, U., Garfi, V., Tognetti, R., Drigo, R., DiMartino and Marchetti, M. 2011. Assessment of potential bioenergy from coppice forests through the integration of remote sensing and field surveys. Biomass and Bioenergy. 35(1):716-724.
- Loog, M., Ginneken, B, Duin, R.P. 2005. Dimensionality reduction of image features using the canonical contextual correlation projection. Pattern Recognition. 38(12):2409-2418.
- McRoberts, R.E. 2008. Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. Remote Sensing of Environment 112: 2212-2221.
- Mc-Roberts, R.E. 2009. A two-step nearest neighbors algorithm using satellite imagery for predicting forest structure within species composition classes. Remote Sensing of Environment. 113:532–545.

- Nakakaawa, C.A. Vedeld, P.O. and Aune, J.B. 2011. Spatial and temporal land use and carbon stock changes in Uganda: implications for a future REDD strategy. *Mitigation and Adaptation Strategies for Global Change*. 16:25-62.
- Návar, J. 2009. Biomass component equations for latin american species and groups of species. *Annals of Forest Science*. 66(2):208.
- Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B. and Ohmann, J.L. 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*. 114:1053–1068.
- SAS Institute Inc. 2004. *SAS/STAT 9.1 User's Guide*. SAS Institute Inc., Cary, NC, USA. 5121 p.
- SEMARNAT-INE 2009 Fourth National Communication of Mexico to the United Nations Framework Convention on Climate Change. Secretaria del Medio Ambiente y Recursos Naturales. Instituto Nacional de Ecología. México, D.F. pp. 128.
- Silva-Arredondo, F.M. y Návar-Cháidez, J.J. 2009. Estimación de factores de expansión de carbono en comunidades forestales templadas del norte de durango, México. *Revista Chapingo. Serie Ciencias Forestales y del Ambiente*. 15(2):155-163.
- Streck, C., O'Sullivan R., Janson-Smith, T. and Tarasofsky, R. 2008. *Climate change an forests. Emerging policy and market oportunities*. Brookings Institution Press. Baltimore, MD. 346 p. ISBN 978-0-8157-8192-9.
- Stümer, W., Kenter, B. and Köhl, M. 2010. Spatial interpolation of insitu data by self-organizing map algorithms (neural networks) for the assessment of carbon stosk in European forests. *Forest Ecology and Management*. 260:287–293.
- Tian, X., Su, Z., Chen, E., Li, Z., Van der Tol, C., Guo, J. and He, Q. 2012. Estimation of forest above-ground biomass using multi-parameter remote

- sensing data over a cold and arid area. *International Journal of Applied Earth Observation and Geoinformation*. 14:160–168.
- Valdez-Tamez, V., Rahum-Foroughbakhch, P. and Gláfiro-Alanís, F. 2011. Distribución relictual del bosque mesófilo de montaña en el noreste de México. *Ciencia UANL*. 6(3):360-365.
- Wolter, P.T., Townsend, P.A. and Sturtevant, B.R. 2009. Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data. *Remote Sensing of Environment*. 113:2019–2036.
- Zhou, P., Luukkanen, O., Tokola, T. and Nieminen, J. Effect of vegetation cover on soil erosion in a mountainous watershed. *Catena* 75:319–325.

5. CONCLUSIONES FINALES

La primera hipótesis planteó que “los índices espectrales de vegetación basados en humedad predicen mejor la densidad forestal con un comportamiento no lineal”. Se realizó el análisis de índices de vegetación, sensibles a la clorofila y al contenido de humedad, contruidos con datos espectrales obtenidos del producto combinado del sensor MODIS MCD43A4, contra estimaciones de biomasa forestal obtenidas del procesamiento de datos del Inventario Nacional Forestal de México, usando dos modelos uno lineal y otro exponencial. Los resultados obtenidos sugieren que los índices espectrales de vegetación basados en el contenido de humedad de la vegetación y los campos continuos de vegetación fueron la mejor combinación de variables explicatorias para estimar la variabilidad espacial de la densidad forestal.

En la segunda hipótesis se planteó que “la cobertura fraccional forestal puede ser estimada mediante métodos de clasificación difusa y datos espectrales”. Para ello se exploraron tres métodos de clasificación subpixel para la determinación de cobertura fraccional utilizando datos de resolución espacial moderada del producto combinado MODIS MCD43A4. Los métodos seleccionados incluyeron la probabilidad posterior bayesiana, el k vecino más cercano en su versión difusa (FkNN), y el análisis de mezcla espectral (LMA). Según el análisis realizado, el algoritmo bayesiano no se comportó adecuadamente, mientras que el algoritmo LMA se seleccionó por ser de los mejores métodos reportados para resolver el problema de píxel mezclado. Los resultados obtenidos indican que el método FkNN ($r = 0.78$) se comportó similarmente a LMA ($r = 0.81$) para la determinación de mezclas a nivel subpixel. Esto aumenta la lista de métodos probados que pueden ser empleados para el diseño de nuevas estrategias de monitoreo de cobertura fraccional a nivel nacional.

Para probar la tercer hipótesis planteada, "el uso de diferentes matrices de ponderación de la distancia espectral para obtener los vecinos más cercanos puede mejorar las estimaciones de densidad forestal mediante el uso de datos espectrales y datos de inventario forestal", se estudiaron cuatro variantes de la matriz de ponderación de la distancia espectral empleada para la determinación de los vecinos más cercanos en problemas de estimación de variables de densidad forestal mediante el uso de imágenes de satélite. Las variaciones incluyen 1) RAW que emplea la matriz identidad, es decir, sin ponderación, 2) EUC que emplea la inversa de la suma directa de la matriz de covarianzas de las X, 3) MAH o distancia de Mahalanobis que emplea la inversa de la matriz de covarianza de las X, y 4) MSN o vecino más similar que emplea una matriz $W = \Gamma \Lambda^2 \Gamma^T$, donde Γ es la matriz de vectores canónicos correspondientes a las X's y es obtenida mediante el análisis de correlación canónica entre las X y las Y, y Λ es la matriz de correlación canónica. Después de aplicar la prueba de Duncan, con el propósito de averiguar si existieron diferencias estadísticamente significativas entre los errores obtenidos en la validación cruzada para cada método, se determinó que los métodos empleados se pueden clasificar en dos grupos: Grupo 1, que incluye cualquiera de las distancias estandarizadas: EUC, MSN y MAH; y Grupo 2, donde queda el método que emplea las distancias sin estandarizar. Los mejores resultados ($r = 0.9$, valores estimados vs valores de referencia) fueron obtenidos con datos eminentemente espectrales y método del método MSN.

Los objetivos del segundo y cuarto capítulo, además de la comparación del desempeño de los métodos estadísticos empleados, incluyeron el generar estimaciones de biomasa forestal. Las estimaciones totales realizadas para las 3.9 millones de arboladas distribuidas en el estado de San Luis Potosí son como sigue: 1) modelo lineal estimó 93.8 millones de Mg (93,808,669.39 Mg), 2) modelo exponencial estimó 116.2 millones de Mg (116,255,822.93 Mg) y, 3) el modelo KNN ponderado con la distancia MSN estimó 109.2 millones de Mg (109,237,612.8 Mg). Las diferencias de las estimaciones pueden ser explicadas

por la naturaleza de cada uno de los métodos. El modelo exponencial se ajustó mejor a la distribución de los datos. No obstante tiene una tendencia de rápido crecimiento hacia el infinito y es más proclive a generar estimaciones elevadas. El promedio de estas estimaciones es de 106.4 millones. La estimación realizada por el método KNN ponderado con la distancia MSN fue el más adecuado ya que se encuentra más cercano al promedio.

Las estimaciones más bajas son más recomendables para el estudio de recursos forestales, para evitar caer en los problemas propios de la sobreestimación. Es decir, planear sobre la base de estimaciones pesimistas (bajos) para asegurar la sustentabilidad en el manejo de los recursos forestales. No obstante, en este caso, se sugiere el uso de la estimación del algoritmo KNN ya que representa una manera dinámica cuasi-irrepetible para la estimación del valor de cada píxel con base en las distancias espectrales a cada una de las parcelas de referencia, lo que le confiere un poder superior al de los métodos paramétricos. Por lo que esta tesis concluye que el KNN estandarizado con MSN es el método más adecuado para la realización de estimaciones de biomasa forestal con datos espectrales.

RESUMEN AUTOBIOGRÁFICO

Carlos Arturo Aguirre Salado

Candidato para el Grado de

Doctor en Ciencias con Especialidad en Manejo de Recursos Naturales

Tesis: MODELACIÓN GEOESPACIAL DE LA DENSIDAD FORESTAL

Campo de estudio: Geomática Aplicada a los Recursos Naturales

Biografía

Datos Personales: Nacido en Texcoco, Estado de México el 7 de Agosto de 1980, hijo de Margarito Aguirre Bravo y Bertha Salado Morales.

Educación: Ingeniero Forestal egresado de la Universidad Autónoma Chapingo en 2001 y Maestro en Ciencias con Mención Honorífica egresado del Colegio de Postgraduados en 2007.

Experiencia Profesional: Desde 2001 a la fecha se ha desempeñado en diversas posiciones que incluyen asistente técnico, investigador, consultor privado y profesor investigador de tiempo completo en el área de Geomática Aplicada al Estudio de Recursos Naturales.