



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Delimitación de la frontera agrícola
en el año 2021, a través de
metodologías de interpretación,
clasificación y procesamiento de
imágenes satelitales en México**

TESIS

Que para obtener el título de

Ingeniero Geomático

P R E S E N T A

Baltazar Itzcuahtzin Cortés Barrera

DIRECTORA DE TESIS

Dra. Griselda Berenice Hernández



Ciudad Universitaria, Cd. Mx., 2024

Agradecimientos

Esta investigación comenzó como parte de un proyecto de servicio social en la que se buscaba mejorar las metodologías y procesos para el uso geoespacial en la agricultura mediante métodos de sistemas de información geográfica (SIG) y percepción remota, fue gracias a la institución gubernamental del Servicio de Información Agroalimentaria y Pesquera (SIAP) de la Secretaría de Agricultura y Desarrollo Rural (SADER) que confió en los conocimientos aprendidos en mi formación profesional como egresado de la Facultad de Ingeniería de la UNAM que me asignaron la tarea de buscar nuevas alternativas para la teledetección y registro de la “Frontera agrícola” de México con las herramientas que ellos me proporcionaron.

En especial quiero mencionar al Ing. Julián Roberto Domínguez Rodríguez, el Ing. José de Jesús Ornelas Quezada y la Mtra. Noemi López González quienes me abrieron las puertas para participar con ellos y fueron una guía en mi proceso de investigación de este tema.

También quisiera agradecer a mi Facultad de Ingeniería de la UNAM y a los profesores que tuve durante la carrera por los años de aprendizaje y principalmente a los que formaron parte de mi jurado sinodal para el proceso de mi titulación; el Ing. José Benito Gómez Daza, el Ing. Marcos Mauricio Sandoval Salazar y a la M. en I. María Elena Osorio Tai, también agradecer al Dr. Luis Antonio García Villanueva profesor de Ingeniería Civil por la revisión de esta investigación, a su vez le agradezco de manera especial a la Dra. Griselda Berenice Hernández Cruz, quien fue mi asesora de esta tesis y que estuvo a lo largo del proceso de mi titulación, le doy mis más sincera gratitud y respeto por su entusiasmo y paciencia, así como por su tiempo y dedicación a la revisión de este trabajo.

Por último, quiero agradecer a mi familia quien siempre me apoyó todos estos años, no solamente en mi proceso de titulación, también a lo largo de mi formación profesional, principalmente mi querida madre la Mtra. María de los Ángeles Barrera Hernández, mi padre el Médico Veterinario Zootecnista, Baltazar Cortés García, mis abuelos el Sr. René Barrera Rendon y la Sra. Araceli Hernández Muñoz y a mi novia Sheila Stefany Barriga.

Espero que este trabajo pueda aportar nuevos conocimientos a nuestra sociedad y a la ciencia Geomática, y a su vez que ayude a dar nuevos pasos a las investigaciones en el ámbito de los sistemas de información geográfica y la percepción remota, así como a la geomática en general, puesto que son herramientas para obtener un mayor conocimiento sobre el mundo que nos rodea y comprender mejor sus necesidades, sus carencias y sobre todo, cómo poder ayudar a que la sociedad y al medio ambiente y la agricultura mantengan un equilibrio ideal para poder alcanzar una sustentabilidad para la actividad económica y el bienestar de la sociedad mexicana.

Resumen

La actualización en la delimitación de la frontera agrícola es esencial para distintos sectores de investigación, saber dónde comienza el uso de suelo agrícola y dónde termina puede arrojar datos geospaciales utilizados para desarrollar modelos estadísticos que facilitan el estudio en diversos ámbitos, ya sea económicos, ecológicos, poblacional, uso de suelo, y producción agrícola. La presente investigación se enfocó en la delimitación del suelo agrícola durante el año 2021 en 2 zonas geográficas del territorio mexicano con características diferentes en clima, relieve y distribución de cultivos, los cuales son el municipio de Tlacoachistlahuaca y Xochistlahuaca en el estado de Guerrero y el municipio de Cajeme y sus alrededores en el estado de Sonora lo que comúnmente se conoce como el Valle del Yaqui. Los resultados de esta investigación se muestran mediante mapas fisiográficos de las zonas analizadas, divididas a su vez en distintas clases dependiendo de los elementos encontrados en el terreno (cultivos, ríos, zonas urbanas, vegetación natural, etc.). El mapa se obtuvo mediante una clasificación supervisada acompañado del método de clasificación por árboles de decisión “*Random forest*”. Los resultados de los mapas realizados muestran las diferencias territoriales entre el entorno agrícola y el ambiente que lo rodea lo que se conoce como “frontera agrícola”, la cual mostró que durante el año 2021 existió un 25.23% (31,521.98 ha.) de la cobertura total analizada (124,922.22 ha.) en los municipios del estado de Guerrero (Tlacoachistlahuaca y Xochistlahuaca), lo cual indica que un poco más del cuarto del área total de ambos municipios es destinado al enfoque agrícola y está hecho para tales condiciones. Por otro lado, en la región de Valle del Yaqui la frontera agrícola final mostró un 38.68% (163,674.93 ha.) de la cobertura total analizada (423,085.3107 ha.), lo cual indica que más de un tercio del territorio se destinó a actividades agrícolas durante el año 2021.

Palabras clave:

- *Ingeniería agrícola*
- *Teledetección*
- *Aprendizaje automático*
- *Clasificación supervisada*
- *Índice de vegetación*

Abstract

Updating the delimitation of the agricultural frontier is essential for different research sectors. Knowing where the use of agricultural land begins and where it ends can provide geospatial data used to develop statistical models that facilitate study in various areas, whether economic or ecological, population, land use, and agricultural production. The present investigation focused on the delimitation of agricultural land during the year 2021 in 2 geographical areas of the Mexican territory with different characteristics in climate, relief and distribution of crops, which are the municipality of Tlacoachistlahuaca and Xochistlahuaca in state of Guerrero and the municipality of Cajeme and its surroundings in the state of Sonora what is commonly known as the Yaqui Valley. The results of this research are shown through physiographic maps of the analyzed areas, divided into different classes depending on the elements found in the terrain (crops, rivers, urban areas, natural vegetation, etc.). The map is obtained through a supervised classification accompanied by the “Random forest” decision tree classification method. The results of the maps made show the territorial differences between the agricultural environment and the environment that surrounds it, which is known as the “agricultural frontier”, which showed that during the year 2021 there was 25.23% (31,521.98 ha.) of the coverage total analyzed

(124,922.22 ha.), in the municipalities of Guerrero (Tlacoachistlahuaca and Xochistlahuaca), which indicates that a little more than a quarter of the total area of both municipalities is destined for agricultural focus and is made for such conditions. On the other hand, in the Yaqui Valley region, the final agricultural frontier showed 38.68% (163,674.93 ha.) of the total coverage analyzed (423,085.31 ha.), which indicates that more than a third of the territory was used for agricultural activities during the year 2021.

Keywords:

- *Agricultural engineering*
- *Remote sensing*
- *Machine learning*
- *Supervised classification*
- *Vegetation Index*

Introducción

Percepción remota como herramienta para el análisis de la agricultura

La teledetección es una herramienta aplicable para muchas ramas, aplicaciones para esta se han encontrado a lo largo de los años y gracias al avance tecnológico de los satélites y nuevos sensores se han podido realizar estudios de diversos tipos. Varios de estos estudios se han aplicado a cuestiones agrícolas, para detectar o evaluar factores limitantes, discriminación de cultivos, inventarios, etapas de crecimiento, vigor vegetal, necesidades nutrimentales, variación espacial de la productividad entre otros.

Diferentes relaciones o razones de datos multispectrales y esquemas de clasificación se han aplicado dependiendo del área geográfica, diversidad de cultivos, tamaño del campo, fenología del cultivo, y condiciones del suelo. Relacionado a las condiciones del cultivo, algunos estudios de sensores remotos se han enfocado en parámetros físicos individuales de los sistemas agrícolas, tales como el estrés de nutrientes o de disponibilidad de agua, que permiten analizar la sanidad y rendimiento de coberturas vegetales.¹

Los campos de cultivo, en la mayoría de los casos, son idóneos para su estudio gracias a su vasta extensión territorial ya que pueden ser fáciles de detectar para los satélites y a su vez llegan a seguir patrones concretos para posteriores predicciones y estudios estadísticos. Sin embargo, los datos derivados de esta técnica, no se han utilizado a su máximo potencial para la gestión de los cultivos, en gran parte, debido a que estos datos no están disponibles para los investigadores y agricultores, y a la comprensión de las características de los sensores como resolución espacial, espectral y radiométrica, además de la fecha de adquisición al integrarla con información convencional la cual deriva del estudio de campo con datos de percepción remota.²

El SIG para el estudio de fronteras agrícolas.

Algunos investigadores se han dedicado a utilizar sistemas de información geográfica y técnicas de percepción en la vegetación para enfocar el estudio a la determinación de la composición de la fisiografía de la vegetación y su uso, de allí nace el concepto de fronteras agrícolas, las cuales se denominan como la suma de todas aquellas áreas incorporadas o susceptibles de incorporarse a la producción agrícola mediante diversos métodos y con distintos plazos y costos en un área determinada.³

Se considera que existen en el país grandes extensiones de tierra natural susceptibles de ser dedicadas a actividades agrícolas, pero se desconoce su localización, extensión y cuál es su grado de productividad.

Las fronteras agrícolas representan entornos dinámicos caracterizados por cambios generalizados en el uso de la tierra y la cobertura de esta, generalmente de bosques y agricultura, pero también siguiendo patrones complejos no lineales. Además de lo anterior, las fronteras comparten características comunes que incluyen abundancia de tierra sin colonizar, bajas densidades de población humana y lejanía de los principales centros urbanos e industriales.

Sin embargo, las tierras ocupadas por cultivos no son las mismas a través del tiempo, ya que nada asegura que una misma superficie es sembrada año tras año. Por lo anterior, la frontera agrícola es un concepto que en materia de superficie cambia y no es fijo.⁴

Estudios previos

Algunos investigadores ya han planteado anteriormente métodos eficaces para determinar el espacio fisiográfico segmentado de imágenes satelitales, tales como *Erika Danaé López Espinoza*, con su metodología para mejorar la clasificación no supervisada mediante métodos algorítmicos como el MRF-FLAT y el TS-MRF/TEF o *Juan Pablo Ardila López* su metodología propuesta de clasificación orientada a objetos, sin embargo, la clasificación de áreas se da de manera muy general en la mayoría de investigaciones y se ha dado un enfoque deficiente a la segmentación agrícola, mencionada en esta investigación como frontera agrícola, ya que la mayoría de las metodologías de clasificación no suelen enfocarse en la composición del espectro en las plantas, sino más bien en algoritmos que mejoren la precisión de los píxeles de una imagen, por lo que deja a un lado las variables que pueden existir para que un software clasifique y diferencie correctamente la vegetación natural y la vegetación agrícola, que estrictamente hablando, pueden contener los mismos valores digitales. Es por ello, que técnicas de clasificación supervisada suelen ser mejores opciones para este tipo de casos, aunque la mayoría de las veces dichas técnicas se utilizan para el caso de un estudio en específico y no se profundiza demasiado acerca del proceso de selección humano.^{5 6}

Algunos casos como el de la investigación *Satellite Remote Sensing of Forest Degradation Using NDFI and the BFAST Algorithm*, utilizan la clasificación supervisada y algoritmos de clasificación para el monitoreo de la deforestación y degradación forestal a través de imágenes satelitales utilizando técnicas de análisis de series temporales de NDVI usando BFAST el cual es un algoritmo que permite detectar quiebres en series temporales incluyendo en el análisis la tendencia, variaciones por estacionalidades, y residuos.⁷

Además, el algoritmo BFAST en el análisis de series temporales proporciona la ubicación y fecha del cambio en la cobertura de la tierra a partir de datos espacio- temporales. De esta forma, la implementación de este algoritmo en el análisis de series temporales de NDVI, permite explorar las capacidades del índice para proporcionar información relacionada con las dinámicas de cambio asociadas a la degradación forestal.

Otro trabajo es el de “Comparación de métodos de clasificación de imágenes de satélite en la Cuenca del río Argos (Región de Murcia)” creado por investigadores del Departamento de Ciencias del Mar y Biología aplicada de la Universidad de Alicante, el cual utiliza métodos de aprendizaje automático como el de Random Forest y un método de clasificación contextual: SMAP, el cual combina métodos de clasificación supervisada por regiones.⁶

Dichos métodos algorítmicos se complementan con el método de máxima verosimilitud, que consiste en una clasificación tradicional supervisada con asignación de píxeles clasificados con base en el criterio propio del ingeniero o técnico a cargo.

Otro método utilizado en la investigación anteriormente mencionada fue el de mínima distancia, que consiste en asignar cada celdilla a la clase respecto a cuyo vector de medias está más próxima en el espacio de variables.⁸

Este método es más rápido de llevar a cabo, pero mucho menos preciso que el anterior. Sin embargo, en dicha investigación no se profundiza en los métodos para una mejor identificación del terreno basándose en técnicas de percepción remota ni índices de vegetación.

Objetivo de la investigación

En la presente investigación se pretende crear una nueva metodología enfocada a la identificación de potenciales campos agrícolas utilizando conocimientos en análisis de percepción remota, algoritmos de clasificación y técnicas de identificación de uso de suelo a través de imágenes satelitales. Los resultados finales darán como resultado características detalladas sobre la fisiografía de los sitios estudiados, los cuales podrían servir para investigaciones futuras en distintos ámbitos de la ciencia.

Índice General

Resumen	1
Introducción	2
CAPÍTULO 1. Técnicas de percepción remota aplicadas a la agricultura	
I.1. Concepto de percepción remota	7
I.2. Espectro electromagnético	8
I.3. Interacción de la energía con la superficie terrestre	9
I.3.1 Interacción de la reflectividad del suelo (Reflexión).....	9
I.3.2 Interacción de la radiación electromagnética con la atmósfera	11
I.4. Respuestas espectrales en la vegetación	14
I.4.1. Porción de espectro electromagnético para el estudio de la vegetación	15
I.4.2. Principales satélites y sensores útiles en el estudio de la vegetación ...	16
I.4.3. Antecedentes de la teledetección para el estudio de la vegetación	17
CAPÍTULO II. Análisis digital de imágenes	
II.1. Área de estudio	19
II.2. Obtención de imágenes de satélite	21
II.3. Variables a considerar para el estudio de los sitios de evaluación	22
II.3.1. Clima	22
II.3.2. Suelo	23
II.3.3. Vegetación	23
II.3.4. Orografía	24
II.4. Características de los cultivos en los sitios de evaluación	25
II.4.1. Formas de cultivo	25
II.4.2. Rotación de cultivos	26
II.5. Realces espectrales para la vegetación	27
II.5.1. Rangos de valores para índices de vegetación	29

II.6. Algoritmo <i>Random forest</i> como clasificador por aleatoriedad	31
II.6.1. Polígonos de entrenamiento como muestras de arranque	33
CAPÍTULO III. Preprocesamientos para la clasificación de los sitios evaluados	
III.1. Reclasificación de variables	35
III.2. Polígonos de entrenamiento	39
III.3. Procesamiento del algoritmo de <i>Random forest</i>	42
III.3.1. Matriz de confusión	44
CAPÍTULO IV. Resultados y discusión	
IV.1 Resultados de la reclasificación de variables	46
IV.2 Clasificación resultante del algoritmo <i>Random forest</i>	52
IV.3 Matriz de confusión resultante	58
IV.4 Proceso de validación y delimitación de la frontera agrícola	61
CAPÍTULO V. Discusión	69
V.1. Conclusiones	70
Anexo	72
Bibliografías	75
Bibliografías de figuras	80

CAPÍTULO I. Técnicas de percepción remota aplicadas a la agricultura

I.1. Concepto de percepción remota

Cualquier área que se dedique al análisis, exploración y monitoreo del planeta Tierra conoce la importancia de la percepción remota como ciencia y disciplina. La teledetección o percepción remota se define como la adquisición y medición de información sobre ciertas propiedades de fenómenos, objetos o materiales mediante un dispositivo de registro que no está en contacto físico con las características bajo vigilancia.⁹

Con ayuda de los satélites se ha facilitado la obtención de imágenes de grandes porciones de la Tierra con menor tiempo y costo, además que, a través de bandas multispectrales, los satélites ofrecen información valiosa acerca de la composición del terreno ya que mediante la combinación de dichas bandas se obtienen datos que de otra forma no es posible observar a simple vista.

Dicha percepción remota se realiza a través de procesos de teledetección que empiezan desde diversos satélites artificiales. Estos satélites contienen diferentes sensores que captan la energía en forma de radiación que emite la superficie terrestre al reflejar la energía solar que llega a la Tierra; Posteriormente, una estación de recepción capta gracias a una antena, la información que recopiló el sensor dentro del satélite y la digitaliza en forma de datos e imágenes, de forma que sea posible manipular la información y procesarla para distintos fines.

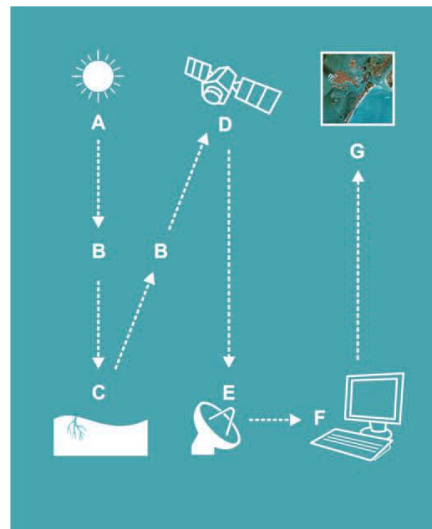


Figura (1) El sol (A) emana energía (B) que es llevada a la Tierra en forma de radiación (C), que a su vez es reflejada al espacio donde los satélites con sensores (D) retrasmíten la información obtenida a antenas(E) donde los investigadores pueden procesarlas en softwares especializados (F) y así obtener imágenes espectrales (G). Referencia [A]

Para poder realizar un estudio a través de la teledetección es necesario contar con imágenes proporcionadas por distintos satélites geospaciales los cuales cuentan con sensores remotos montados en plataformas que capten imágenes de la superficie con la mejor resolución posible.

Los sensores captan porciones de radiación del espectro electromagnético reflejadas de la superficie terrestre a través de las denominadas ventanas atmosféricas, las cuales son las franjas de longitudes de onda que forman parte de la radiación que logra escapar de la atmosfera, dichas longitudes son agrupadas en regiones conocidas como bandas. La información de dichas bandas se recopila en distintas variables como amplitud espectral, resolución radiométrica, resolución espacial y resolución temporal, los cuales aportan información fundamental para entender las características de cada imagen obtenida.

I.2. Espectro electromagnético

Para poder comprender mejor cómo funciona la percepción remota es necesario entender las interacciones físicas entre la superficie y la forma en la que la radiación interactúa con esta. Para empezar, hay que comprender que la principal fuente de energía proviene del Sol en forma de radiación electromagnética, conformada por luz visible, infrarroja y ultravioleta por mencionar algunos. Existen diferentes formas en la que la energía puede ser transferida como por ejemplo la conducción, convección o radiación.

La energía radiada se comporta básicamente acorde con la teoría ondulatoria de la luz, la cual dice que es un continuo de ondas que se caracterizan por su tamaño (medida en nanómetros, nm) y frecuencia el número de ondas por unidad de tiempo. Ambas propiedades se encuentran relacionadas, la longitud de onda es la distancia entre cresta y cresta de la onda y la frecuencia es el número de ciclos, mientras más pequeña es la longitud de onda más alta es la frecuencia. Las longitudes de onda de importancia para la percepción remota son aquellas ubicadas en el intervalo de la luz visible (390 - 780 nm.), el infrarrojo cercano (700 - 900 nm), infrarrojo térmico (3000 - 14000 nm) y la radiación en microondas (arriba de 15000 nm).¹⁰

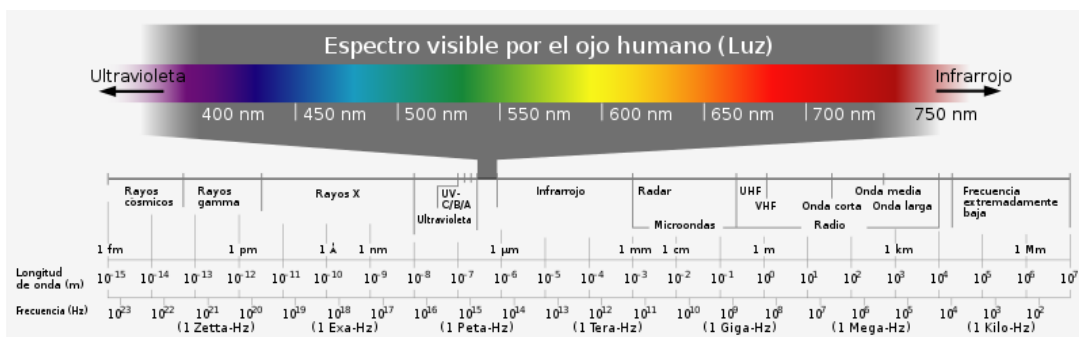


Figura (2) Rangos que componen el espectro electromagnético. Referencia [B]

La región del espectro “visible” es muy pequeña, debido a que la sensibilidad espectral del ojo humano se extiende sólo desde los 400 nm hasta aproximadamente los 700 nm. El color "azul" corresponde al rango de 400 a 500 nm, el "verde" de 500 a 600 nm y el "rojo" de 600

a 700 nm. Estas radiaciones son una porción muy pequeña del espectro electromagnético, son registradas por emulsiones fotográficas y por radiómetros que operan en estas bandas. En esta región del visible, se produce el máximo de energía solar que llega a la Tierra. Es por ello que se necesitan instrumentos especializados para poder identificar rangos del espectro que los seres humanos no pueden ver más allá del espectro visible. ¹¹

En el caso de la percepción remota, el rango que más será utilizado para el estudio de superficies terrestres será el rango visible y el infrarrojo corto, medio y lejano, ya que la radiación que emite la energía solar en forma de luz infrarroja es suficientemente fuerte como para hacer vibrar las moléculas de cuerpos terrestres y logren reflejar dicha energía. Es por esa razón que muchos sistemas de sensores remotos operan en las regiones de longitud de onda donde predominan la energía reflejada.

I.3. Interacción de la energía con la superficie terrestre

Cuando la radiación impacta con un objeto, dicha radiación puede ser reflejada, absorbida o transmitida, por lo tanto, cada objeto posee una respuesta, “firma o signatura espectral”, que es propia de cada material y dependerá de su energía reflejada o emitida, cuando esto sucede un sensor puede identificar la cantidad de flujo radiante emitido por el objeto y esto permite identificar las características de dicho objeto ya que la cantidad de energía que se refleja transmite o absorbe en cada cuerpo es diferente para cada longitud de onda. Estas diferencias permiten distinguir diferentes rasgos y/o elementos en una imagen.

El color que se observa a diario se debe a que los objetos que son golpeados por la radiación en forma de luz tienen distintas composiciones y reflejan de forma distinta la longitud de onda de dicha radiación; Por ejemplo, las hojas de los árboles son de color verde por que la composición molecular de éstas hace que refleje la parte del espectro visible de color verde (500 nm aprox.) y el resto de los colores sean absorbidos por el material.

I.3.1 Interacción de la reflectividad del suelo (Reflexión)

Tener en cuenta la composición de los objetos es importante a la hora de hacer teledetección, ya que la reflectancia de éste también cambiará y estos cambios permitirán mostrar patrones de relación entre objetos de distinta índole, como suelo desnudo, suelo con vegetación, cuerpos de agua, zonas urbanas entre otros. Cuando los objetos de la superficie terrestre absorben la energía, se calientan y pueden emitir parte de su energía a la atmósfera. La proporción de energía total incidente que interactúa con la superficie terrestre depende de la longitud de onda, del tipo de material y de la condición de los rasgos sobre la superficie.

En percepción remota, el interés básico se centra en la radiación que es reflejada por los blancos de interacción. A esto se le denomina reflexión y puede ser de dos tipos:

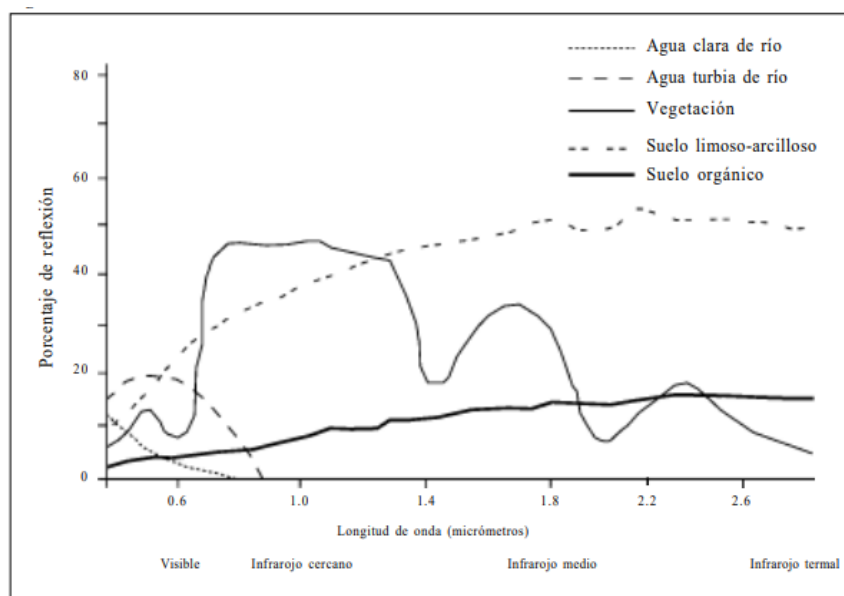
a) Reflexión especular. Ocurre en superficies que reflejan la radiación solar con un ángulo igual al ángulo incidente

b) Reflexión difusa. Ocurre en aquellas superficies que reflejan la radiación hacia todas las direcciones (superficies lambertianas).¹²

Mediante un sistema de teledetección es posible registrar tres tipos de variaciones en los objetos: espectrales, espaciales y temporales. Todo objeto de la superficie terrestre refleja o emite un campo de energía electromagnética (EMM) según una función que depende de su longitud de onda.

Si se dispone de un sensor capaz de captar y medir la energía procedente de los objetos, cada uno de ellos aparecerá en la pantalla o imagen con su firma específica, de un valor sólo suyo, para cada longitud de onda registrada por el sensor. De este modo se puede contar con datos espectrales numéricos que irán trazando una curva en función de la longitud de onda.¹³

En la *Figura (3)* se presentan los patrones característicos de reflectancia de distintos tipos de clases de cobertura del suelo y las características que los distinguen según su porcentaje de reflexión en el espectro electromagnético.



Patrones característicos de reflectancia de distintas clases de cobertura del suelo

Figura (3) Esquema de las distintas reflectancias de diferentes superficies. El eje X representa la longitud de onda y el eje Y el porcentaje de reflexión. Cada superficie tiene distintos niveles de reflectancia dependiendo de la longitud de onda a la que sea observada. Referencia [C]

Las características de algunas de las distintas superficies importantes que representa la gráfica anterior son las siguientes:

Suelo desnudo: la reflectancia del suelo está determinada por el contenido de humedad superficial, la composición mineral, el contenido de materia orgánica, la textura y la rugosidad.

- Los suelos con más humedad absorben más radiación visible (VIS) y de manera particular, más radiación en el infrarrojo cercano (IRC) que los suelos secos.
- Los suelos con alto contenido de carbonato de calcio y los suelos arenosos con contenidos altos de cuarzo reflejan fuertemente en el intervalo del VIS y el IRC.
- Los suelos con alto contenido de hierro reflejan fuertemente en el VIS, en particular en la porción roja del espectro.
- Los suelos con textura más fina retienen más agua que los suelos de textura gruesa por lo que los primeros, debido al mayor contenido de agua, reflejarán menos que los suelos de textura gruesa. Sin embargo, los suelos de textura fina al desecarse pueden formar una capa impermeable que aumente la reflectancia.

Agua: las características espectrales varían de acuerdo con varios factores. El contenido de partículas o sedimentos, materia orgánica, material vegetal y profundidad son algunos de los factores más importantes que afectan su comportamiento espectral.

Vegetación: el comportamiento espectral de la vegetación depende fuertemente de las propiedades de las hojas y la estructura del dosel de la vegetación (número de capas de hojas) y la orientación de las hojas.¹²

1.3.2 Interacción de la radiación electromagnética con la atmósfera

Existen diferentes fenómenos y procesos, en los cuales la energía solar interactúa con la atmósfera de la Tierra al atravesarla con luz y calor los cuales se conocen como radiación. Dicha radiación al interactuar con la atmósfera, (*Figura (4)*) crea un sistema de transferencia de energía en donde los gases pueden retener, reflejar o simplemente dejar pasar dicha energía a la superficie terrestre. Esto depende mayormente de la longitud de onda de la energía transmitida y del tamaño y naturaleza de la sustancia que modifica la radiación.

Dichos procesos de atenuación que sufre la radiación solar en su paso hacia la atmósfera se conocen como **dispersión, absorción y emisión**.

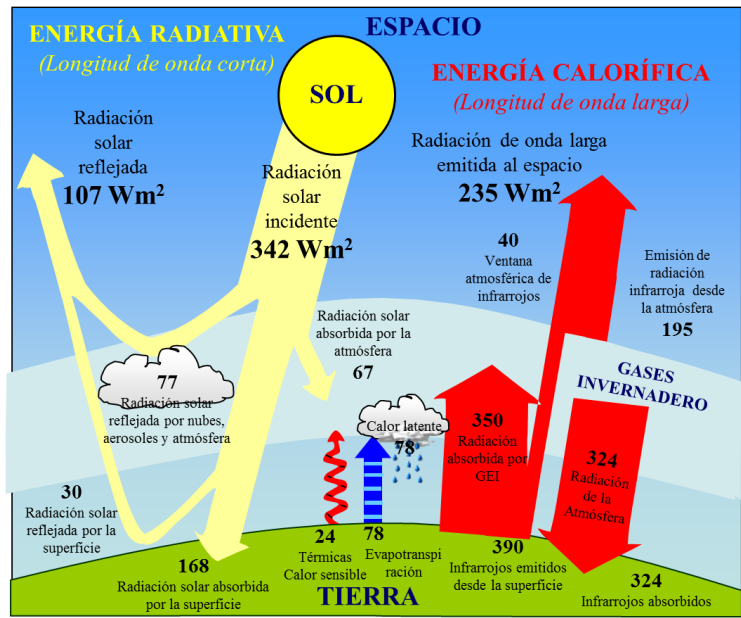


Figura (4) En esta imagen se aprecia como superficie de la tierra y la atmósfera, emanan, absorben y transmiten la energía radiante. Referencia [D]

Absorción

Proceso por el cual las moléculas y partículas de la atmósfera absorben la energía radiante (65 Wm^{-2}) y la transforman en energía interna, que posteriormente se emite en el infrarrojo térmico. El ozono, el bióxido de carbono y el vapor de agua son los constituyentes atmosféricos que en mayor medida absorben la radiación solar.¹²

En la teledetección los sensores están diseñados para captar la radiación en los lugares donde la absorción atmosférica es más alta. Esto se debe a que la absorción reduce la observación de bandas del espectro electromagnético creando así las llamadas **ventanas atmosféricas** las cuales son espacios entre el espectro donde la radiación no es completamente absorbida. Dichas regiones son las ideales para que opere un sistema de sensor remoto.

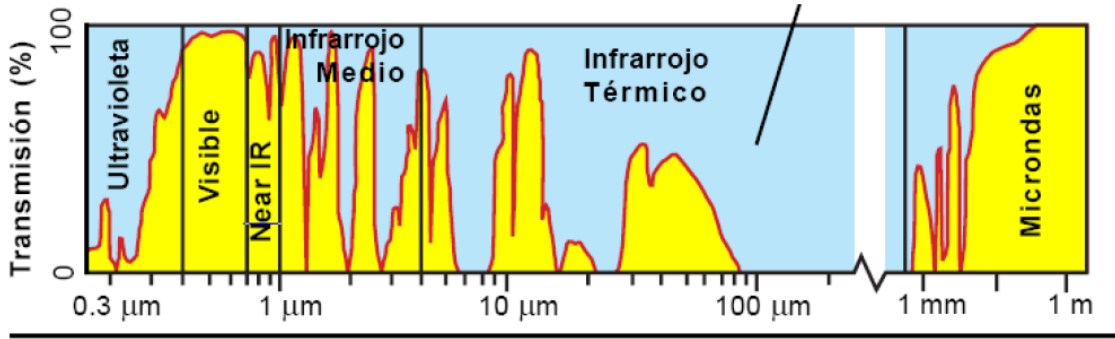


Figura (5). Variación en la transmisión atmosférica por absorción de las longitudes de onda de los gases atmosféricos. Referencia [E]

Cada uno de los gases atmosféricos tiene capacidad para absorber radiación en diferentes longitudes de onda, pero principalmente son el ozono, el dióxido de carbono y el vapor de agua los que mayor relevancia tienen.

Dispersión

La dispersión es el fenómeno más común y es lo que causa que la luz se difumine a lo largo de la atmósfera y la dispersa en todas direcciones; Un ejemplo de este efecto es que el cielo tenga un color azulado, esto se debe a que los gases de la atmósfera dispersan más efectivamente las longitudes de onda más cortas (violeta y azul) que en longitudes de onda más largas (naranja y rojo).

Existen diversos tipos de dispersión atmosférica, cada uno se diferencia por su reflectancia por parte de partículas con características diversas como el tamaño, su composición y su cantidad. Algunas de estas dispersiones son las siguientes:

Dispersión de Rayleigh

La dispersión Rayleigh se produce cuando las partículas atmosféricas tienen un diámetro muy pequeño en relación con la longitud de onda de la radiación. Estas partículas pueden ser aerosoles de pequeño diámetro, pero más normalmente son algunas moléculas de gases atmosféricos, como N_2 u O_2 , cuyos diámetros son mucho más pequeños que la longitud de onda de la radiación visible e infrarroja próxima. Esta dispersión se produce incluso en atmósferas totalmente limpias, sin impurezas, por lo que a veces se hace referencia a ella como "dispersión de atmósfera clara".¹⁴

Dispersión de Mie

La dispersión Mie es causada por partículas de mayor tamaño presentes en la atmósfera, incluyendo polvo, polen, humo y gotas de agua. Se produce cuando el diámetro de las partículas es similar a la longitud de onda de la radiación. Este tipo de dispersión también afecta a un rango amplio de longitudes de onda, y también es hasta cierto punto dependiente de la longitud de onda, pero la relación no es tan directa como en el caso de la dispersión Rayleigh. Frente a la dispersión Rayleigh, la Mie tiende a ser máxima en los primeros 5 km de atmósfera, donde abundan más las partículas grandes, aunque depende mucho más del estado de la atmósfera.¹⁴

Dispersión no selectiva

La dispersión no selectiva es la causada por partículas cuyo diámetro es mucho mayor que la longitud de onda de la radiación que es dispersada. Como su nombre indica, no es dependiente de la longitud de onda y afecta por igual a todas ellas. Por esto, las nubes o nieblas se ven blancas o grisáceas, ya que todas las longitudes de onda visibles son dispersadas por igual. Estas partículas de gran diámetro son generalmente polvo atmosférico o, más frecuentemente, gotas de agua.¹⁴

Desde la perspectiva de la percepción remota, la dispersión es útil en el sentido que los sensores satelitales pueden registrar la luminosidad que refleja la atmósfera y la que llega a la superficie terrestre lo que ayuda a que aumente el brillo de la imagen tomada, además que disminuye el contraste y difumina los bordes de los objetos captados por la cámara.

Emisión

Cualquier cuerpo con temperatura mayor que $-273\text{ }^{\circ}\text{K}$ emite su propia radiación, que lógicamente tiene mayor importancia en el infrarrojo térmico. Por tanto, su efecto es fundamental si se pretende trabajar en la banda del infrarrojo térmico. Cuando se trabaja en el infrarrojo térmico, el objetivo fundamental es estimar la temperatura de la superficie terrestre. A su vez hay que tomar en cuenta que los datos que llegan al satélite incluyen la emisión de la superficie y de las diferentes capas de la atmósfera, resulta difícil distinguir un efecto del otro, por lo que existen diversas técnicas para solventar este problema a través de la corrección de imágenes como la corrección atmosférica.¹²

I.4. Respuestas espectrales en la vegetación

Como todo objeto reflectante, la vegetación tiene sus particulares firmas espectrales, dichas firmas dependerán del tipo de vegetación, las condiciones donde se encuentra, su interacción con otras especies vegetales, la época de germinación y su interacción con la acción humana ya sea por medio de fertilizantes, forma de cultivos, entre otros.

Diversas firmas espectrales se relacionan también con pigmentos fotosintéticos y la cantidad de agua que almacenan las hojas. La baja reflectividad que se presenta en la porción visible se debe a la absorción de los pigmentos de la hoja, principalmente clorofila, xantofilas y carotenos. La reflectancia más baja se da en las bandas azul y roja de la región visible del espectro. El efecto absorbente es menor en la banda verde, fenómeno que genera el color con el que percibimos la vegetación "vigorosa" en plena foliación.¹⁵

La vegetación en general suele tener una reflectividad baja en el espectro visible, sin embargo, las plantas siguen un patrón espectral en particular debido a la clorofila que contienen las hojas de las plantas, la cual es la responsable de reflejar el color verde del espectro visible.

Según la biología toda planta que produzca clorofila producirá moléculas que serán las responsables de absorber la energía radiante proveniente del sol, dichas moléculas tienen distintas reflectancias en diferentes zonas del espectro electromagnético. Como la luz solar es la fuente primaria de energía para las plantas, la radiación de esta es determinante para las respuestas sensoriales de las plantas en muchos ámbitos de su crecimiento y desarrollo. Otra característica distintiva de la vegetación, se encuentra en la banda del rojo, donde se observan valores altos de absorción de la radiación por parte de la clorofila en la banda del rojo y valores altos de dispersión de la radiación por la estructura celular en la banda del infrarrojo cercano.¹⁶

Siendo las plantas organismos de carácter sésil, los mecanismos encaminados hacia lograr la adaptación oportuna a los cambiantes factores del entorno son de un gran valor selectivo. La adaptación a los cambios ambientales depende de toda una serie de ajustes moleculares, fisiológicos y morfológicos que se presentan como respuesta a las modificaciones de la irradiancia, la calidad espectral, el fotoperíodo, la temperatura, etc. ¹⁷

Las hojas de las plantas son las responsables de la reflectancia en la superficie de estas y la absorción es responsable principalmente por la estructura interna, que a su vez provoca la composición, concentración y distribución de los pigmentos (la clorofila, el caroteno y la xantofila).

1.4.1. Porción de espectro electromagnético para el estudio de la vegetación

Los pigmentos fotosintéticos de las plantas crean una respuesta fuerte en las longitudes de onda de 400 nm a 700 nm, de forma tal, que cuando aparecen condiciones anormales en el follaje se produce insuficiente clorofila lo que provoca un incremento en la reflectancia del espectro visible, por lo que la planta adquiere un tono verde pálido, amarillo o amarillo blanquecino (*Figura (6)*). En esta región del espectro se produce la fotosíntesis, proceso que implica la absorción de energía. En resumen, esta región se caracteriza por una escasa reflectancia y una fuerte absorción debida a los pigmentos foliares de las hojas, esto se traduce en un máximo de reflectancia en torno a los 550 nm. ¹⁸

En las longitudes de onda de 700 nm a 1100 nm las absorciones electroquímicas están limitadas por los componentes encontrados en plantas secas, celulosa, lignina y otros carbohidratos estructurales. Sin embargo, es afectada por la reflexión por múltiple esparcimiento de la radiación dentro de la hoja, relacionados con la estructura interna, fracción de espacios de aire e interfaces agua-aire que refractan la luz dentro de las hojas.

En el rango comprendido entre 1100 nm y 2500 nm, el infrarrojo medio o infrarrojo de onda corta presenta fuerte absorción, principalmente por el agua en hojas verdes, existiendo picos característicos en la vegetación a 1450, 1940 y 2500 nm, correspondiente a absorciones primarias de agua en la hoja y absorciones secundarias en 980 nm y 1240 nm.

En hojas secas, la reflectancia y transmisión están influenciados por componentes de carbón seco como celulosa y lignina, nitrógeno, azúcares y otros componentes. Las uniones de nitrógeno e hidrógeno tienen su primer armónico a 1510 nm y una serie de combinaciones de bandas para 1980, 2060 y 2180 nm. ⁹

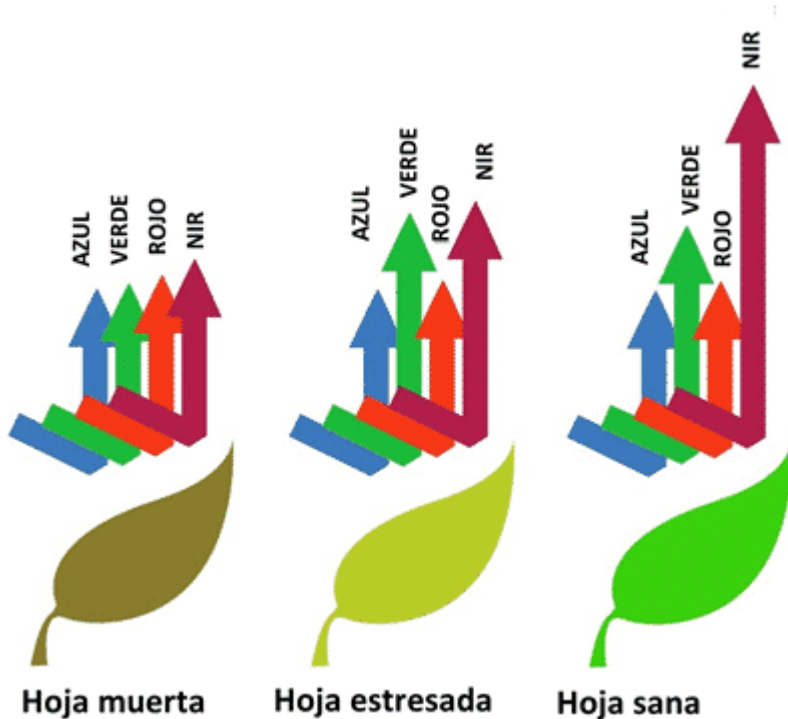


Figura (6) En esta imagen se ve como el color es afectado por el estado de la clorofila, en buen estado las hojas logran reflejar el color verde y el infrarrojo cercano, con forme la hoja pierde clorofila empieza a absorber más radiación lo que ocasiona que color tienda a verse más amarillento y opaco. Referencia [F]

I.4.2. Principales satélites y sensores útiles en el estudio de la vegetación

La cobertura vegetal fue uno de los primeros focos de la investigación de la evaluación y manejo de recursos naturales, usando imágenes de satélite, especialmente a partir del lanzamiento de la serie LANDSAT en 1972. Los sistemas satelitales de observación de la Tierra, LANDSAT, SPOT y NOAA entre otros, ofrecen imágenes multitemporales que son usadas ampliamente, para evaluar y monitorear el estado de la vegetación, en los niveles global, regional, nacional y local. ¹⁹

Para lograr esto, la información satelital entregada a través de una imagen multibanda debe ser categorizada y agrupada, para permitir discriminar un área con características particulares de otra. Una forma de expresar esta categorización o agrupamiento es mediante la elaboración de índices, en este caso exclusivamente para la vegetación.

Existen diversos satélites geoespaciales que permiten la obtención de imágenes espectrales, la mayoría cuentan con sensores remotos que comparten ciertas funciones fundamentales para la detección de vegetación en masa, debido a que al trabajar con imágenes espaciales se debe tener en cuenta la resolución espectral, espacial y frecuencia de cobertura del área seleccionada.

Para el estudio de la vegetación se requieren imágenes con gran cobertura y una resolución de menos de 60 metros ya que varios de estos estudios requieren una observación precisa de la fenología y la distribución de la planta y sus diferencias entre sí, por lo tanto, se requieren satélites con sensores idóneos para la tarea. Algunos satélites que cumplen con estas características y con la resolución adecuada son; Sentinel 2 y 3, SPOT, Envisat, Terra, Landsat por mencionar algunos.

Una vez que se establecen las características que determinan la naturaleza de la información que puede extraerse de una imagen, de acuerdo con las diferentes porciones del espectro electromagnético, se pueden explorar procesamientos que permitan satisfacer los requerimientos necesarios de información. Esencialmente se utilizan las bandas del espectro visible y del infrarrojo para obtener información de coberturas vegetales mediante sensores remotos, las cuales pueden combinarse algorítmicamente en un índice de vegetación, para resaltar elementos de discriminación de estado y variedad vegetal en la superficie terrestre.

1

1.4.3. Antecedentes de la teledetección para el estudio de la vegetación en México

Varios investigadores, universidades e incluso dependencias del gobierno de México y del mundo han utilizado diversas técnicas de teledetección con el objetivo de mejorar la claridad y precisión en los estudios de la vegetación y cultivos agrícolas con fines estadísticos y de análisis foliar.

En México la principal dependencia que se encarga de utilizar técnicas modernas de teledetección para el monitoreo de la vegetación es la SADER (Secretaría de Agricultura y Desarrollo Rural) la cual se caracteriza por la implementación de una plataforma tecnológica robusta y escalable aunada a la experiencia de un equipo de especialistas en materia geoespacial coordinados por el SIAP utilizando un SIG, realizan la actualización de la frontera agrícola del país tomando como base las coberturas nacionales de imágenes satelitales con resolución espacial de 1.5 metros de la constelación SPOT que proporciona la antena ERMEX cada año; se realiza la interpretación visual con apoyo de puntos verificados en campo a través de una aplicación regionalizada multipropósito para navegadores GPS.²⁰

Dicha información se utiliza en informes estadísticos cada año para dar a conocer al público las condiciones en las que se encuentra el campo agrícola mexicano y sus características a nivel nacional.

Además de las investigaciones que se llevan a cabo en el gobierno del país, distintas universidades han utilizado la teledetección en un uso agrícola analizando la “**Estimación de la evapotranspiración de un viñedo de uva de mesa**”.²¹

Como indican los analistas de este proyecto la razón de esta investigación es que la uva es un cultivo con un requerimiento alto de agua y por lo tanto la precipitación escasa y la demanda evaporativa elevada de la zona dificulta su crecimiento. Por ello su producción depende de la aplicación de riego y es importante estimar con precisión y de manera extensiva la evapotranspiración de este cultivo. Para mejorar la eficiencia del riego a corto plazo, se realizaron investigaciones y metodologías con el objetivo de desarrollar y evaluar un modelo

entre el índice de vegetación de diferencias normalizadas (NDVI) calculado con imágenes Sentinel-2, y el coeficiente de cultivo (KC) determinado con el sistema de covarianza de vórtices (Eddy Covariance, EC) como opción para estimar la ETC (evapotranspiración del cultivo) de un viñedo de uva de mesa en la Costa de Hermosillo, Sonora (*Figura (7)*).²¹

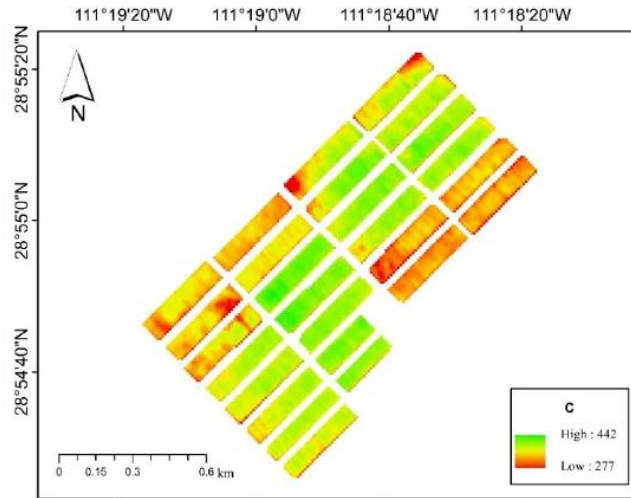


Figura (7) Promedio de NDVI (A), promedio de KCNDVI (B) y ETCEST (C) del periodo marzo – junio del año 2019, en un viñedo de uva de mesa en Sonora, México. Referencia [G]

Otra investigación llevada a cabo por parte del Centro de Investigaciones Tropicales CITRO en la Universidad Veracruzana, estudio la deforestación de las selvas de la región Uxpanapa y los cambios del uso del suelo suscitados desde el año 1976 al 2009.

Se utilizaron técnicas de clasificación supervisada respaldados por puntos de apoyo *in situ* para la clasificación del suelo, además de utilizar imágenes satelitales LANDSAT y SPOT para poder realizar un *timelapse* de la zona. Los resultados mostraron que hubo una tasa de deforestación de 2,02% anual, además que la vegetación selvática en 1975 predominaba con un 75% del paisaje, mientras que para el año 2009 esta clase solo ocupaba el 37%. (*Figura (8)*).²²

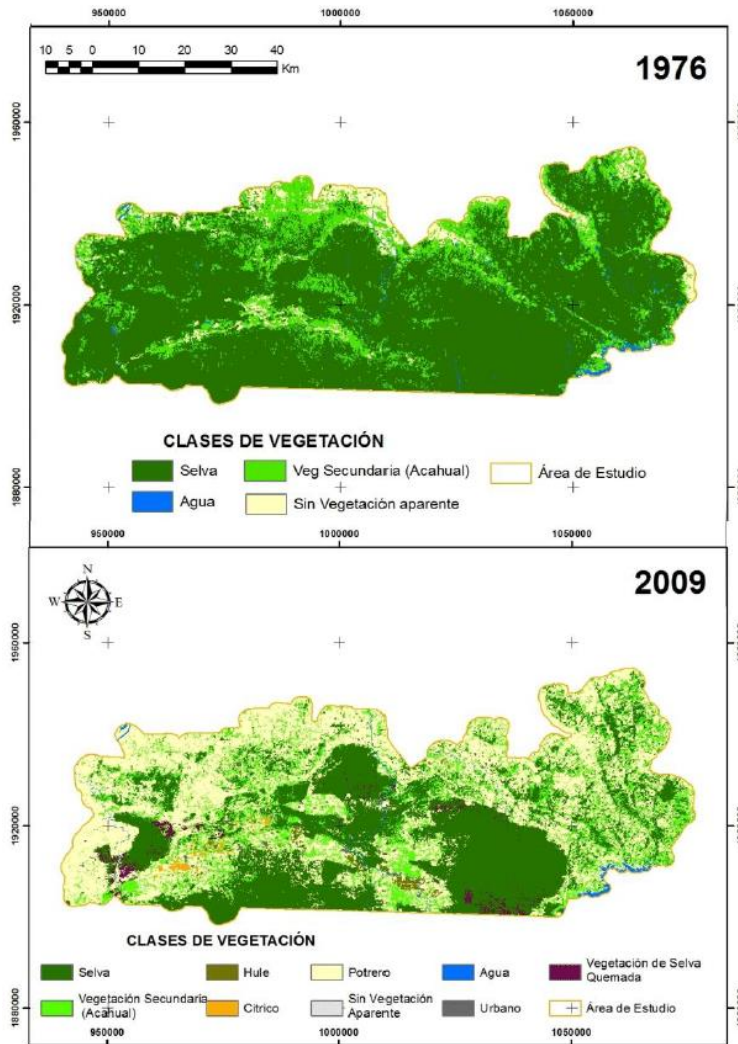


Figura (8) Comparativa del uso de suelo en la selva de Uxpanapa, Veracruz entre el año de 1976 y 2009.
Referencia [H]

CAPÍTULO II. Análisis digital de imágenes

II.1. Área de estudio

Las áreas de estudio utilizadas difieren entre sí por sus características fisiográficas climatológicas y de uso de suelo, sin embargo, la razón principal por la que ambos sitios fueron seleccionados, fue por las marcadas diferencias entre las fronteras agrícolas que contienen, ya que se pretende demostrar que las metodologías funcionan tanto para sitios

cuya frontera agrícola es claramente visible y de los que tienen una frontera agrícola sumamente difusa.

Los sitios de evaluación seleccionados destacan por ser regiones completamente diferentes en un sentido fisiográfico, las características como suelo, relieve, vegetación o clima influyen directamente en los resultados que se esperan obtener en el tratamiento de imágenes satelitales y en la metodología aplicada a la clasificación de zonas agrícolas. El contraste de los diferentes factores físico-ambientales que existen en cada sitio, permiten comparar resultados de las distintas variables obtenidas en el proceso y de esta manera demostrar que las metodologías utilizadas pueden aplicarse para el análisis y estudio de zonas completamente diversas.

El primer sitio (*Figura (8)*) se encuentra en el municipio de Tlacoachistlahuaca y Xochistlahuaca en el Estado de Guerrero los cuales se encuentran en las coordenadas 16°48' N, 98°18' O y 16°47' N, 98°14' O respectivamente. Mientras que el segundo sitio (*Figura (9)*) se encuentra en el estado de Sonora donde los campos agrícolas a estudiar abarcan parte de los municipios de Cajeme, Bácum, San Ignacio Río Muerto y Guaymas, a su vez estos municipios se encuentran cerca de Ciudad Obregón, punto principal de la región cuyas coordenadas son en el paralelo 27°29' N y el meridiano 109°59' O.²³

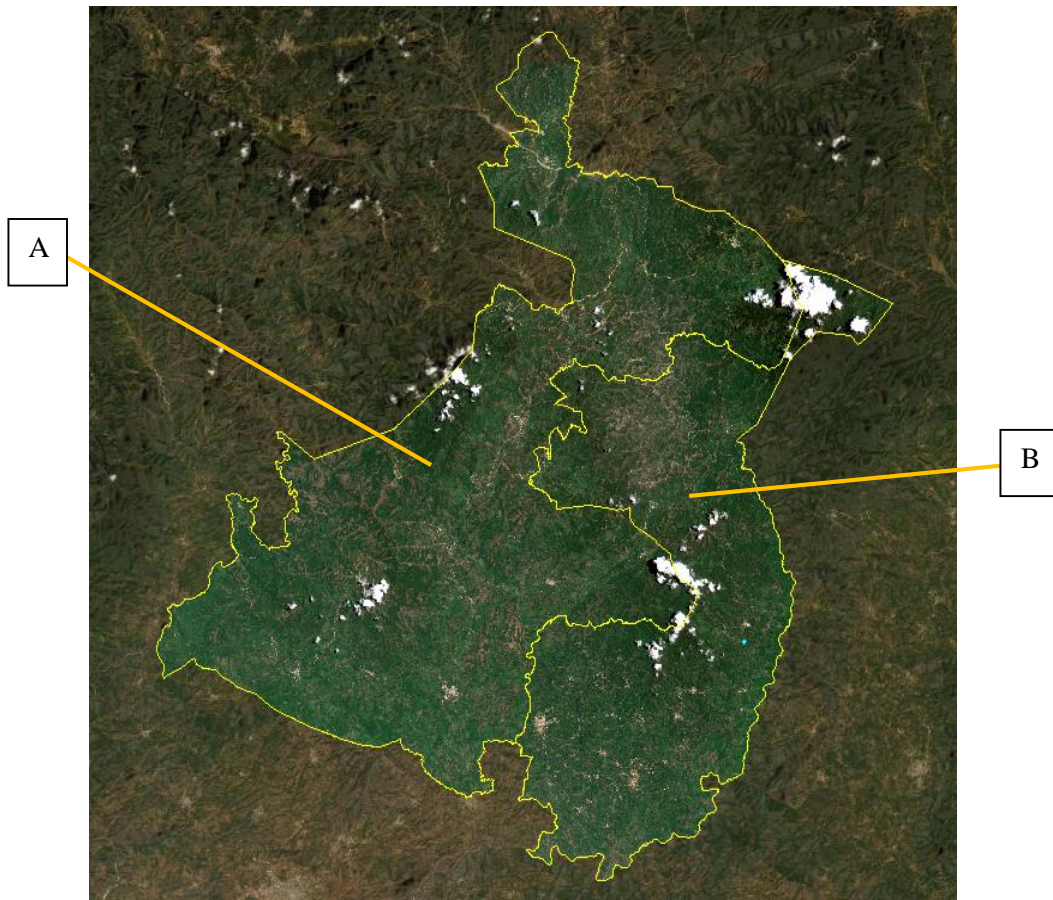


Figura (8) Municipio de Tlacoachistlahuaca (A) y Xochistlahuaca (B).

Esta zona fue elegida por sus campos de cultivos irregulares y por su escarpado suelo, además de tener un clima templado-tropical. El área de estudio que abarca casi en su totalidad ambos municipios es de 1255.20 Km²



Figura (9) Ciudad obregón (A) en Cajeme (B) y municipios colindantes: Bácum (C), San Ignacio Río Muerto (D) y Guaymas (E)

Al contrario de la región anterior, en la (figura (9)) se observa un sitio plano. En el centro, sur y oeste se encuentra el Valle del Yaqui con 106 200 hectáreas de superficie agrícola regular y planificada, mientras que al norte se encuentra la zona serrana y al este el mar de Cortés.²³

El área de estudio tiene una superficie de 4478.55 Km²

II.2. Obtención de imágenes de satélite

Las imágenes de la región de Ciudad Obregón fueron obtenidas del repertorio de imágenes de la compañía *EOS Data Analytics* en la plataforma de *Land Viewer*. Dichas imágenes fueron obtenidas del satélite *Sentinel 2*. Las imágenes fueron tomadas el 5 de febrero del 2022.

Por otra parte, las imágenes obtenidas de la región que incluye los municipios de Tlacoachistlahuaca y Xochistlahuaca fueron proporcionadas por parte del Servicio de Información Agroalimentaria y Pesquera (SIAP) y obtenidas del mismo satélite *Sentinel 2*, cuyas imágenes fueron tomadas el 1 de agosto del 2020.

Las imágenes obtenidas cuentan con una corrección realizada previamente como la detección de nubes o la calibración radiométrica y geométrica además de ortorectificación. Además, incluyen todas las bandas que pueden obtenerse del satélite *Sentinel 2* es decir 12 bandas, las cuales conforman regiones desde el espectro visible en el azul (443 nm) hasta el infrarrojo de onda corta (2190 nm) con una resolución de 10, 20 y 60 metros por píxel de acuerdo con sus respectivas bandas.

Para la selección de imágenes, se tomó en cuenta una nubosidad menor al 10% de cobertura y que contarán con cultivos anuales y perennes.

Para la compilación de variables también fue necesario incluir imágenes de temperatura, y precipitación en cada zona, las cuales fueron obtenidas del repositorio institucional de imágenes satelitales del Instituto de Ciencias de la Atmósfera y Cambio Climático, UNAM. Además, se obtuvo el ráster de pendiente por medio de un modelo de elevaciones digitales obtenido del Instituto de Nacional de Estadística y Geografía (INEGI) ^{24,25,26}

II.3. Variables a considerar para el estudio de los sitios de evaluación

II.3.1. *Clima*

Sitio 1: Municipios de Tlacoachistlahuaca y Xochistlahuaca, estado de Guerrero

Esta zona al situarse debajo del Trópico de Cáncer y formar parte de la Sierra Madre del Sur, cuentan con un clima subtropical el cual, según la clasificación de climas de Wladimir (1936) tiene las denominaciones de $Aw2(w)$ en la mayor parte del territorio y $Aw1(w)$ en los valles al norte de la región, los cuales indican que tiene un clima cálido con lluvias la mayor parte del verano, con un cociente de precipitación anual de 55.3 y temporada de sequía en invierno con un 5 a 10% de precipitación. Las partes montañosas que conforman aproximadamente un 10% de la región cuentan con un clima $A(C)m(w)$ y $A(C)w2(w)$ que simboliza un clima semi cálido muy húmedo con intensas lluvias en verano y con precipitaciones menores al 10% en el mes más seco de invierno.²⁷

Sitio 2: Municipio de Cajeme y colindancias, estado de Sonora

Este sitio al igual que la mayor parte del norte de la República Mexicana se encuentra en una zona desértica y se caracteriza por tener un clima seco en toda el área. Según la clasificación Köppen, esta cuenta principalmente con 2 tipos de climas, uno tipo $BW(h)hw$ en el interior y $BW(h)hw(x)$ cerca de la costa, se caracterizan principalmente por pertenecer a un clima muy seco con temperatura cálida que no baja de 18°C en el mes más frío y con temperaturas

mayores a 22°C el resto del año, además suele haber lluvias en verano que alcanzan del 5 al 10% de precipitación anual. ²⁷

II.3.2. Suelo

Sitio 1: Municipios de Tlacoachistlahuaca y Xochistlahuaca, estado de Guerrero

El sitio se compone principalmente de 2 tipos de suelos, el Regosol eutrico y el Litosol, ambos encontrados en gran parte del país, principalmente en la Sierra Madre Occidental y del sur. El Regosol cubre la mayor parte de la región de ambos municipios y está compuesto por una capa conocida como “ócrica”, que, al ser retirada la vegetación, se vuelve dura y costrosa impidiendo la penetración de agua hacia el subsuelo. La consecuente sequedad y dureza del suelo es desfavorable para la germinación y el establecimiento de las plantas. El agua, al no poder penetrar al suelo, corre por la superficie provocando erosión.

Por otra parte, el Litosol que se encuentra mayormente en el norte montañoso de la región se caracteriza por ser un suelo muy delgado, pedregoso y poco desarrollado que puede contener una gran cantidad de material calcáreo. No son muy aptos para la agricultura, pero pueden servir para zonas de pastoreo. ²⁸

Sitio 2: Municipio de Cajeme y colindancias, estado de Sonora

Los tipos de suelo que componen esta región son principalmente 3: el Solonchak, el Xerosol y el Vertisol, todos típicos de regiones desérticas y arenosas. El Solonchak, se encuentra presente mayormente cerca de la costa y se caracteriza por presentar acumulación de sales solubles en alguna parte del suelo o en todo su espesor, debido a la fuerte evapotranspiración a que están sujetos.

El Xerosol se compone superficialmente de varias combinaciones de materia orgánica, arcillas y sales, y es un suelo de transición entre el Solonchak y el Vertisol, además que suele ser un suelo apto para la vegetación de matorral.

El Vertisol se encuentra principalmente en las zonas cultivables de esta región y está compuesto de material arcilloso y sedimentos, se caracterizan principalmente por su drenaje deficiente de agua lo que ocasiona una impermeabilidad natural. Esta misma ayuda a que su suelo sea beneficioso para el cultivo excepto cuando es época de sequía donde el suelo se vuelve demasiado duro. ²⁹

II.3.3. Vegetación

Sitio 1: Municipios de Tlacoachistlahuaca y Xochistlahuaca, estado de Guerrero

La vegetación en estos municipios es abundante y se compone principalmente de pastizales y bosques. Los bosques de encino y otros árboles caducifolios abundan cerca de los cerros, mientras que el bosque mesófilo con árboles como el pino suele encontrarse en la parte más alta de las montañas que se encuentran al noreste y oeste de la región. Hacia el sur y algunas partes del norte y centro de la región se pueden encontrar pastizales inducidos y cultivados, mientras que las zonas agrícolas de temporal suelen concentrarse en el sur y centro-sur de la región caracterizados por tener una pendiente menos pronunciada que en el resto de la región.

30

Sitio 2: Municipio de Cajeme y colindancias, estado de Sonora

Como es característico de esta región, el Valle del Yaqui se caracteriza por la típica vegetación desértica compuesta en distintas zonas. Al oeste de la región cercana a la costa se encuentran manglares además de vegetación halófila xerófila la cual se desarrolla en suelos con alto contenido de sales, en las cuencas cerradas de las zonas áridas y semiáridas, como lagunas costeras, marismas y litorales. En la parte norte se encuentran mezquites y matorrales subtropicales y al Este y centro la zona ha sido modificada para dar paso al cultivo en su mayoría con riego permanente.³⁰

II.3.4. Orografía

Sitio 1: Municipios de Tlacoachistlahuaca y Xochistlahuaca, estado de Guerrero

La región se encuentra en la denominada región montañosa de Guerrero al Este del mismo estado. Como su nombre lo indica dentro de estos territorios municipales, existen elevaciones montañosas que alcanzan hasta 2,000 metros sobre el nivel del mar, de las cuales destacan la Sierra de San Jerónimo y los cerros del Águila, Alcapixtla y Tres Cruces los cuales a su vez forman parte de la Sierra Madre del Sur en la región conocida como Nudo Mixteco. Las zonas semiplanas se localizan principalmente en el área central del municipio y zonas planas se localizan al sur, donde también se concentran la mayoría de las zonas cultivables, las cuales se alimentan de riachuelos y ríos como el de Santa Catarina que sirve como frontera natural con el estado de Oaxaca.³¹

Sitio 2: Municipio de Cajeme y colindancias, estado de Sonora

La característica geográfica principal que resalta en esta región es el Valle del Yaqui el cual tiene una extensión de 450,000 hectáreas y abarca desde el municipio de Guaymas hasta el municipio de Huatabampo al sur de Sonora, que comprende el Valle del Yaqui el cual es atravesado por el río del mismo nombre y alimenta las zonas agrícolas que a su paso se encuentran a lo largo del mismo valle. A su vez el río atraviesa las faldas de la Sierra del Bacatete que se encuentra al norte de esta región y termina desembocando en el Golfo de California.

II.4. Características de los cultivos en los sitios de evaluación

La geografía de estas regiones, así como su clima y su aprovechamiento de recursos hídricos influyen directamente en la forma y distribución, pero también en los distintos tipos de cultivos, forma de riego y distribución de parcelas, en conjunto todas estas características están diseñadas para sacar el máximo provecho en los sitios donde se encuentran. Sin embargo, la diferencia de condiciones socioeconómicas por parte de la población también es un factor muy influyente para la forma y distribución de parcelas.

II.4.1. Formas de cultivo

Algunos de los factores importantes que influyen en el nivel y la calidad de la producción agrícola son la disponibilidad de recursos principalmente hídricos, el tipo de suelo, y la variabilidad natural de las lluvias. En el caso de los municipios de Tlacoachistlahuaca y Xochistlahuaca, estos factores son muy importantes ya que al no contar con una infraestructura de riego suelen depender directamente del clima y los recursos naturales que puedan obtener.

En general, la agricultura de esos municipios de Guerrero y colindantes es de temporal y de riego, de cultivos cíclicos y perennes, donde la mayoría de parcelados cuentan con pequeñas unidades de riego diseminadas por todo el estado (*Figura (10)*). Los productores de estas zonas agrícolas aprovechan escurrimientos de agua y espacios en torno a ríos y barrancas o mediante pozos para utilizarlos como suministro de agua para sus riegos. Los parcelados que ahí se encuentran mayoritariamente están distribuidos de forma irregular, ocupando el espacio que se tenga disponible, habiendo cultivos individuales que no alcanzan ni un cuarto de hectárea.³²



(Figura 10) Serie de cultivos típicos de la región visto desde una imagen satelital (Izquierda) y una fotografía representando el paisaje de la zona agrícola (Derecha) Referencia [1]

Por otra parte, en el valle del Yaqui se tiene una región agraria más organizada y distribuida siendo consecuencia de la gran infraestructura de riego con la que cuenta. En el Valle del

Yaqui se construyó una enorme cuadrícula por medio de calles rectas de norte a sur y de oriente a poniente en donde la cuadrícula está regida por dos grandes ejes que forman una T (*Figura (11)*). El diseño de este sistema de cultivo y su precisa organización para el uso del agua que proviene del río Yaqui y sus canales artificiales hacen que sea de las formas más eficientes de agricultura en clima desértico y aprovechar el máximo la siembra de distintos alimentos a pesar de las sequías y otros factores adversos.³³



(Figura 11) Serie de cultivos del valle del Yaqui visto desde una imagen satelital (Izquierda) y una fotografía representando el paisaje de cultivos en la zona agrícola (Derecha) Referencia [J]

II.4.2. Rotación de cultivos

La rotación de cultivos es algo común en las diferentes zonas agrícolas del país, se hace principalmente para reponer los nutrientes que faltan en el suelo de forma natural, sin productos artificiales, lo que viene especialmente bien para lugares donde no se cuenta con fertilizantes ni otros métodos para mantener cultivos perennes en un mismo sitio. Guerrero se caracteriza por tener una compleja orografía y topografía que lo pone en ventaja con otros estados, debido a que cuenta con una amplia riqueza en biodiversidad y con más opciones productivas, por estas condiciones agroecológicas se ha logrado desarrollar en condiciones de riego y temporal la producción de 52 cultivos cíclicos y 40 perennes, colocando al estado entre los cinco primeros lugares a nivel nacional como productor de mango, copra, jamaica, melón, ajonjolí y café.³⁴

Es por ello por lo que, a pesar de las condiciones geográficas y de difícil acceso la zona es especialmente fértil para sembrar alimentos dependiendo de su temporada ya que se reutiliza el mismo terreno para diferentes cosechas en distintos periodos de tiempo. Se estima que en el año 2021 los municipios de Tlacoachistlahuaca y Xochistlahuaca cosecharon distintos alimentos en aproximadamente 17,914.41 hectáreas de las distintas parcelas distribuidas en toda la región, algunos de los alimentos más producidos fueron: maíz grano (15,074.45 ha.), frijol (1,522.53 ha.), aguacate (296.5 ha.), ajonjolí (366.85 ha.), mango (239 ha.), entre otros.³⁵

En la región del valle del Yaqui no se suele depender de la rotación de cultivos debido a que en muchas regiones hay un tipo de alimento que por su rentabilidad prevalece, y resulta difícil encontrar otros que se puedan incorporar a la rotación. Además, no todos los cultivos generarán beneficios en una rotación; incluso es posible que algunos tengan efectos negativos

en la cosecha subsecuente, ya que una gran variabilidad de alimentos tiene diferentes tipos de raíces y, en consecuencia, distintas necesidades de nutrientes.³⁶

La región también cuenta con una economía sostenible que permite a los agricultores disponer de fertilizantes y otros métodos que reducen la degradación de nutrientes y evitan las plagas que se concentran por tener un mismo tipo de cosecha. Sin embargo, algunas zonas del valle se caracterizan por presentar problemas de salinidad y altas concentraciones de sodio intercambiable, lo que provoca disminución en la productividad de alimentos como la nuez pecana. Al terminar la temporada de riego con agua del sistema de presas del Río Yaqui, el uso de agua de pozos profundos es utilizada generalmente en los meses de mayo a agosto. Esto puede ocasionar acumulación de sales en el suelo cuando el agua es de mala calidad, por ello uno de los métodos para contrarrestar este efecto es el uso yeso agrícola como enmienda para rehabilitar el suelo mejorando las propiedades químicas del mismo.³⁷

Gracias a los planes agrícolas e infraestructura realizados para el valle se ha podido contar con una amplia distribución de alimentos de todo tipo, incluyendo algunos que por sus características no podrían germinar en condiciones naturales en la región. Algunos de los alimentos más producidos por hectáreas en el valle conformado mayormente por los municipios de Cajeme, BÁCUM y San Ignacio Río muerto son: trigo grano (77,776.28 ha.) maíz grano (49,442 ha.), alfalfa (5,565 ha.), papa (4,265 ha.) y nuez (3,602 ha.).³⁵

II.5. Realces espectrales para la vegetación

Los índices de vegetación incluidos en esta investigación fueron seleccionados por sus valores de reflectancia, los cuales, ayudan a diferenciar entre vegetación sana y con estrés hídrico cuando se cuenta con la observación de un área con gran cobertura vegetal como es en caso de los cultivos que se observan desde las imágenes satelitales. Dichos índices son el **NDVI, DNBR, NDMI, PVI y SAVI**.

Los índices incluidos se describen a continuación:

Índice NDVI

El Índice de Vegetación de Diferencia Normalizada (NDVI por sus siglas en inglés) es el más conocido y utilizado cuando se trata de diferenciar cobertura vegetal sana y marchita, en este caso, se utiliza para diferenciar cobertura vegetal densa y poco densa la cual sería un indicador de presencia de cultivo. A su vez, el índice marca una diferencia importante entre parcelas con cultivos, cultivos en descanso y cerro (suelo natural con escasa vegetación).

Sin embargo, muchas de las áreas, comúnmente estudiadas, con baja densidad o cobertura de vegetación, están en zonas áridas y semi-áridas, donde muchas de las plantas que crecen en estos climas, poseen una variedad de adaptaciones para hacer frente a la falta de agua y a las altas temperaturas. Estas cualidades, las hacen difíciles de detectar, a menos que se observen durante los períodos de relativa abundancia de agua, cuando toda una nueva serie de adaptaciones para maximizar la productividad de las plantas entre en funcionamiento.

Índice PVI

El índice PVI es un índice de vegetación basado en la distancia, de manera que el espacio perpendicular del píxel en la línea es proporcionalmente dependiente de la reflectancia de la vegetación. Al tomar en cuenta la pendiente y el gradiente de la reflectividad del suelo, los valores más altos del índice PVI crean un efecto de línea de suelo más brillante, por lo cual valores relativamente bajos de reflectancia en la vegetación indicarían vegetación pobre o con escasa cobertura, por lo cual el índice no tomaría en cuenta estas zonas como parte de una cobertura vegetal importante y tendrían un valor bajo en el rango de reflectancia.³⁸

Cabe destacar que este índice solo funciona si se han realizados correcciones atmosféricas correspondientes ya que se ve afectado por las fluctuaciones atmosféricas. Además, es recomendable utilizarlo mayormente en zonas áridas o semi desérticas, debido a que el tipo de vegetación natural suele tener índices de vegetación bajos debido a su composición fisiológica obteniendo así poca reflectancia para estos índices.³⁹

Índice DNBR

Utilizado comúnmente para la identificación de vegetación carbonizada el *Difference Normalized Burn Ratio (DNBR)* es la relación normalizada de la diferencia del infrarrojo cercano y el infrarrojo de onda corta. La combinación de estas últimas bandas permite analizar el fenómeno en condiciones previas y posteriores al incendio. En la zona post-incendio, las áreas quemadas recientemente muestran una reflectividad relativamente baja en la banda infrarroja cercana y una alta reflectancia en la banda infrarroja de onda corta.⁴⁰

Este índice es eficiente al detectar un déficit de humedad en la vegetación, así como vegetación recién quemada, pero en este caso sirve para marcar diferencia de valores entre las regiones de cultivos que colindan con suelo árido o semi-árido ya que cuentan con condiciones parecidas a la reflectancia que emite la vegetación carbonizada o semi-carbonizada.

Índice NDMI y SAVI

El índice *NDMI (Normalized Difference Moisture Index)* expresa el contenido de humedad diferenciado a nivel de elemento del paisaje, especialmente en lo que respecta al suelo y la vegetación. El índice *SAVI (Soil-Adjusted Vegetation Index)* viene a reducir el efecto causado por las condiciones de iluminación del suelo, además que muestra la cobertura vegetal a través de la corrección de la reflectividad del suelo. Estas condiciones juegan un papel importante en el área analizada debido principalmente a que el tipo de suelo dominante es de textura arenosa, característica dominante en la región árida analizada.⁴¹

El índice SAVI también tiene en cuenta la reflectividad del suelo y permite aislar la información que aporta la vegetación de la que procede del suelo que está bajo ella.

La siguiente tabla describe las fórmulas de los índices de vegetación y su aplicación para las bandas obtenidas del satélite Sentinel 2:

Tabla (1) Fórmulas de los índices de vegetación utilizados

Fórmula de los índices de vegetación en el espectro electromagnético	Fórmula aplicada a bandas Sentinel 2
$NDVI = ((NIR - RED) / (NIR + RED))$	$NDVI = ((B8A - B4) / (B8A + B4))$
$DNBR = ((NIR - SWIR 2) / (NIR + SWIR 2))$	$DNBR = ((B8 - I2) / (B8 + B12))$
$NDMI = ((NIR - SWIR 1) / (NIR + SWIR 1))$	$NDMI = ((B8 - I1) / (B8 + B11))$
$SAVI = ((1.5 * (NIR - RED)) / (NIR + RED + 0.5))$	$SAVI = ((1.5 * (B8A - B4)) / (B8A + B4 + 0.5))$
$PVI = (NIR - a * RED - b) / (\sqrt{1 + a^2})$ <i>a = pendiente de la línea del suelo</i> <i>b = gradiente de la línea del suelo</i>	$PVI = (B8 - a * B4 - b) / (\sqrt{1 + a^2})$ <i>a = pendiente de la línea del suelo</i> <i>b = gradiente de la línea del suelo</i>

II.5.1. Rangos de valores para índices de vegetación

Los valores de los índices de vegetación se expresan como una cuantificación de los valores de reflectancias obtenidas como resultado de las fórmulas en cada una de estas, por lo que en cada muestra de imagen satelital el rango de los distintos índices de vegetación variará dependiendo de distintos factores como la calidad de la imagen, el tipo de suelo, factores ambientales y el tipo de índice de vegetación que se utiliza, así como su fórmula y normalización.

En general, todos los índices expuestos anteriormente, exceptuando al DNBR y PVI siguen un mismo patrón para la identificación de vegetación. Como se puede observar en la (figura (12)) a menor valor se considerará como suelo sin vegetación o muy poca, aunque también pueden representar áreas urbanas, cuerpos de agua o nubes y sus sombras. Por el contrario, mayor valor del índice representará cultivos con siembra, cultivos con pasto y zonas con mayor vegetación y humedad como bosques y hierba densa.

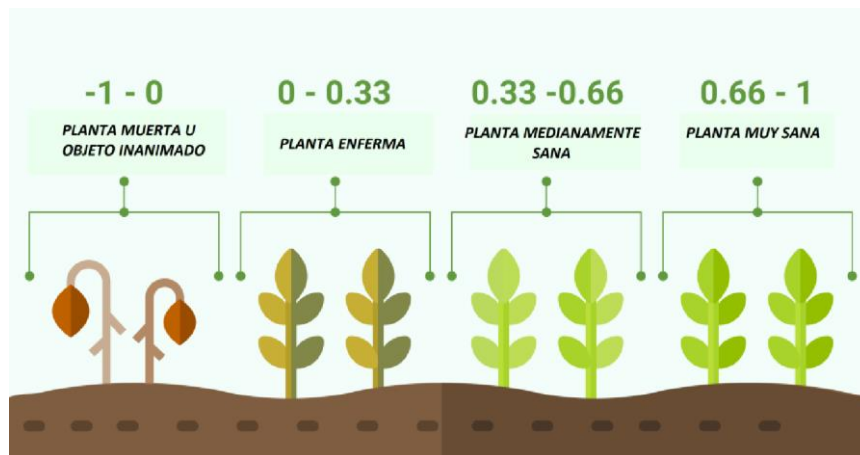


Figura (12) Diferencia de valores en el NDVI sobre vegetación sana y marchita. Referencia [K]

Identificación de parámetros

Existen ciertos parámetros que indican los rangos en los que se pueden encontrar cultivos con siembra y pasto dependiendo de la zona, el tipo de suelo y relieve, pero para obtener mayor exactitud se encuentran los valores propios de los distintos índices de vegetación que se encuentran en el ráster con vegetación genérico. (Tabla (2)) y (Tabla (3))

Los rangos de reflectancia para cultivos y vegetación de los distintos índices son los siguientes:

Tabla (2) Valores comparativos de índices de vegetación NDVI y SAVI según el sensor OLI de septiembre de 2015. ⁴²

Zona	NDVI media	SAVI media
Cultivo	0.265	21.80
Cultivo en pivot	0.420	55.60
Bosque nativo	0.412	36.30
Arbustiva, pie de montaña	0.239	19.92
Áreas degradadas	0.038	12.75
Barreales	0.151	11.95

Nota: Los valores de reflectancia en el índice SAVI no están normalizados.

Tabla (3) Interpretación del NDMI según el proveedor de imágenes satelitales EOS data analytics. ⁴³

Niveles de reflectancia del NDMI	Interpretación
-1 – -0.8	Suelo desnudo
-0.8 – -0.6	Cobertura del dosel casi inexistente
-0.6 – -0.4	Cobertura del dosel muy baja
-0.4 – -0.2	Cobertura del dosel baja y seca o muy baja y húmeda
-0.2 – 0	Cobertura del dosel media-baja y estrés hídrico alto o cobertura del dosel baja y estrés hídrico bajo
0 – 0.2	Cobertura media del dosel y alto estrés hídrico o cobertura media-baja del dosel y bajo estrés hídrico
0.2 – 0.4	Cobertura del dosel media-alta y estrés hídrico alto o cobertura media y estrés hídrico bajo
0.4 – 0.6	Cobertura del dosel alta, sin estrés hídrico
0.6 – 0.8	Cobertura del dosel muy alta, sin estrés hídrico
0.8 – 1	Cobertura total del dosel, sin estrés hídrico/anegamiento

Por otra parte, dado que el índice PVI no tiene oficialmente un rango de índice de vegetación definido, se suele dar la libertad al analista crear un rango de observación dependiendo de los resultados y características que se hayan obtenido en una imagen.

Los valores de los índices de vegetación dependen de las características que se observen en el análisis de la imagen del sitio de estudio, pero usualmente se considera que un valor de 0 o cercano a este, supondrá una escasa vegetación mientras que un valor de 0 a -1 se considerará nula y por último una reflectancia entre rangos de 0.1 a 1 se considerará vegetación con niveles de clorofila consistentes, es decir, vegetación sana y moderadamente sana.

II.6. Algoritmo *Random forest* como clasificador por aleatoriedad

Random forest es un algoritmo de aprendizaje de toma de decisiones y que es utilizado en distintos ámbitos de las matemáticas como modelo predictor y clasificador de variables. La manera en la que el algoritmo toma decisiones es mediante árboles de decisión en forma de diagrama de flujo, donde cada nodo interno representa un valor del vector de características, las ramas representan decisiones, y los nodos representan un resultado de la clasificación.

Random forest utiliza múltiples de estos árboles de decisión aleatorios donde posteriormente se usa el voto suave (*soft-voting*), donde se elige el mejor árbol de decisión con más importancia de todo el bosque de árboles de decisión aleatorios.⁴⁴

Random forest también funciona como un algoritmo de aprendizaje supervisado que puede utilizarse tanto para tareas de clasificación como de regresión. El algoritmo está en los denominados *ensemble methods*, métodos que combinan las predicciones de varios estimadores con el fin de mejorar la generalización con respecto a un estimador.⁴⁵

Aunque *Random forest* es una agrupación de árboles de decisión, este mismo genera reglas de agrupamiento y genera árboles de decisión de forma aleatoria lo que permite tener más opciones de selección y no se sobreajusta a un solo árbol de decisión. Además, mientras que los árboles de decisión profundos pueden sufrir sobreajuste, ***Random forest evita el sobreajuste en la mayoría de los casos, ya que trabaja con subconjuntos aleatorios de características y construyen árboles más pequeños a partir de dichos subconjuntos.***^{46,47}

La aleatoriedad que se le otorga al algoritmo al realizar los árboles de decisión es para disminuir la varianza del estimador conjunto. Así como los árboles de decisión individuales suelen presentar una alta varianza y tienden a sobreajustarse *Random forest* consigue reducir la varianza combinando diversos árboles, sin embargo, esto puede aumentar el sesgo estadístico. Aunque para usos clasificatorios, la reducción de la varianza suele ser significativa, lo que conlleva a obtener un modelo más preciso.

Es por ello que *Random forest* es el método de *machine learning* más utilizado en las implementaciones de mantenimiento predictivo ya que los árboles de decisión permiten que un gran número de observaciones formen parte de las predicciones, aunque también suele tener la desventaja de requerir de un tiempo de procesamiento más largo que otros algoritmos de *machine learning* debido a la gran cantidad de árboles de decisión que debe crear.⁴⁸

Una de las mejores aplicaciones para el algoritmo de *Random Forest* es la clasificación de variables, consiste en la utilización de múltiples árboles de decisión en los que varía de forma aleatoria las variables que se utilizan, creando un listado de posibles resultados, clasificando los menos útiles y los más aptos. Además, *Random Forest* es el algoritmo de *Machine Learning* que más se utiliza para la clasificación supervisada basada en píxeles o polígonos de entrenamiento, además de obtener una buena exactitud en la temática gráfica.⁴⁹

Para realizar el procesamiento de clasificación de variables, *Random forest* utiliza el principio de embolsado (*Bagging*), también conocido como *Bootstrap Aggregation*. El embolsado genera una muestra aleatoria del conjunto de datos proporcionados, por lo tanto, cada modelo se genera a partir de las muestras de las variables llamadas *Bootstrap Samples* que son proporcionadas por los datos originales con un reemplazo ya conocido, en un proceso conocido como muestreo por filas.⁵⁰

Después del proceso de embolsado, el modelo se entrena de forma independiente lo que genera resultados diferentes en cada interacción aleatoria. El resultado final se basa en la votación por mayoría después de combinar los resultados de todos los modelos. Este paso, que consiste en combinar todos los resultados y generar nuevos resultados basados en la votación por mayoría, a esto se le conoce como agregación.

En la (figura (13)) se muestra el diagrama del funcionamiento de *random forest* de la siguiente manera:

- 1.- Se genera un conjunto de árboles de decisión que juntos forman un bosque aleatorio y se toman n números de registros aleatorios del conjunto de datos de cada árbol de decisión.
- 2.- Se construyen árboles de decisión individuales para cada una de las variables/clases.
- 3.- Cada árbol de decisión generará una salida donde se clasificarán.
- 4.- El resultado final se considera en función de la votación mayoritaria en el caso de una clasificación o el promedio en caso de una regresión.

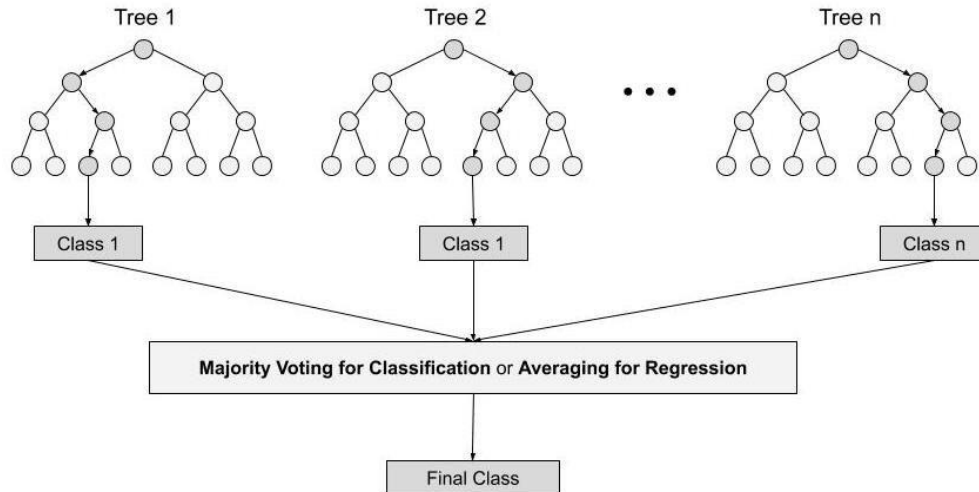


Figura (13) Diagrama del funcionamiento de clasificación por votación de random forest. Referencia [L]

Sin embargo, aunque sea un algoritmo de aprendizaje supervisado como las redes neuronales, no es posible controlar la topología de los estimadores generados por el algoritmo, esto hace que los parámetros que realice no puedan ser controlados, ya que son escogidos aleatoriamente. Por esta razón mejorar los resultados de este algoritmo realizando un cambio en sus parámetros no garantiza obtener mejores resultados.⁵¹

II.6.1. Polígonos de entrenamiento como muestras de arranque

Para que el algoritmo de random forest pueda crear una clasificación funcional, es indispensable tener una muestra de arranque (*bootstrap sample*), esta muestra de datos se utiliza para que el modelo algorítmico tenga un “entrenamiento” que le ayude correlacionar el conjunto de datos muestra con los datos a clasificar. Para el caso de una clasificación supervisada de una imagen, se utilizan **polígonos de entrenamiento**, figuras vectoriales que contienen la información de una imagen ráster, generando así una descripción en forma numérica de los atributos espectrales de cada cobertura de las imágenes utilizadas.⁵²

La forma más simple para categorizar la información temática en imágenes satelitales es la clasificación por píxeles. El algoritmo de *random forest* escoge cada uno de los píxeles de una imagen a alguna clase determinada, de acuerdo con su mayor semejanza con los datos de entrenamiento. Este agrupamiento de píxeles se realiza usualmente mediante un proceso de datos estadísticos (*clustering*) en el que cada píxel es asignado a una categoría de acuerdo con su correspondiente nivel digital (ND), también propiamente dicho como respuesta espectral del píxel.⁵³

Los datos espectrales que se encuentran dentro de cada píxel se usan como una base numérica para la categorización de cada uno, lo que provoca que las imágenes presenten información

espectral distinta con base en su radiancia en el caso de bandas espectrales u otro tipo de variables.

Para generar la información temática que se requiere para poder visualizar la clasificación de forma gráfica, los píxeles individuales que conforman la imagen satelital deben ser agrupados en categorías definidas o clases.

La clasificación se realiza a partir de los polígonos de entrenamiento realizados de forma supervisada para cada clase, donde posteriormente el algoritmo de clasificación, en este caso *random forest*, agrupara los píxeles que conformen cada polígono de forma aleatoria formando varios árboles de decisión, como se muestra a continuación (*figura (14)*).

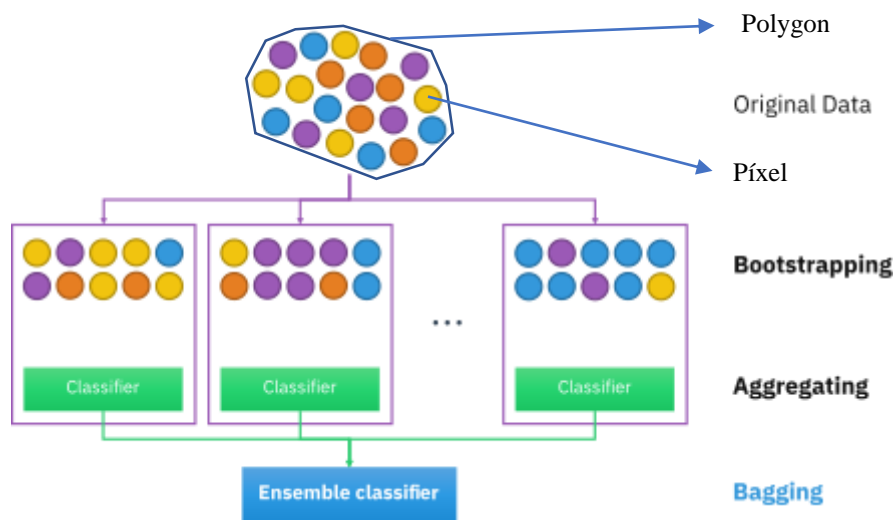


Figura (14) Cada píxel conformado por el polígono de entrenamiento (bootstrap sample) es analizado y clasificado de diferentes formas aleatorias en distintos árboles de decisión por el algoritmo de random forest. Referencia [M]

Cuando las entidades de entrenamiento de entrada son polígonos, la variable para la predicción de categorías de polígonos debe convertirse en ráster para poder utilizar la información espectral contenida en cada píxel. Una vez realizada la transformación el polígono se divide en puntos de cada celda ráster cuyo centroide quede dentro del polígono, posteriormente el polígono se trata como un *dataset* de puntos, por lo tanto, el modelo ya no se entrena en el polígono, sino en los valores de ráster extraídos para cada centroide de celda. Aquellos píxeles que contengan variables numéricas cuyo centroide no quede contenido dentro del polígono deben pasar por un proceso de interpolación bilineal, el cual consiste en utilizar una media ponderada de los valores de los píxeles más cercanos para darle un nuevo valor al píxel faltante.⁵⁴ (*Figura (15)*)

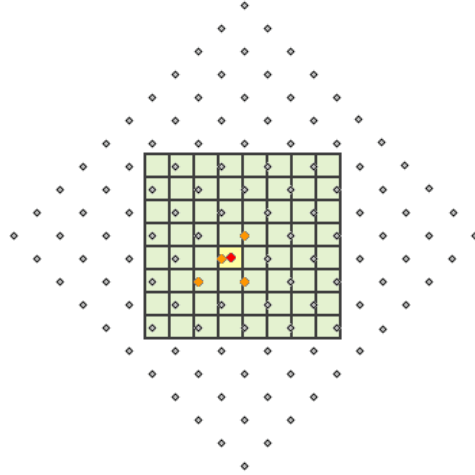


Figura (15) Ejemplo de interpolación bilineal, el punto rojo representa el pixel con valor faltante, y los puntos naranjas representan los pixeles vecinos. Referencia [N]

CAPÍTULO III. Preprocesamientos para la clasificación de los sitios evaluados

III.1 Reclasificación de variables para la elaboración de un ráster multicriterio

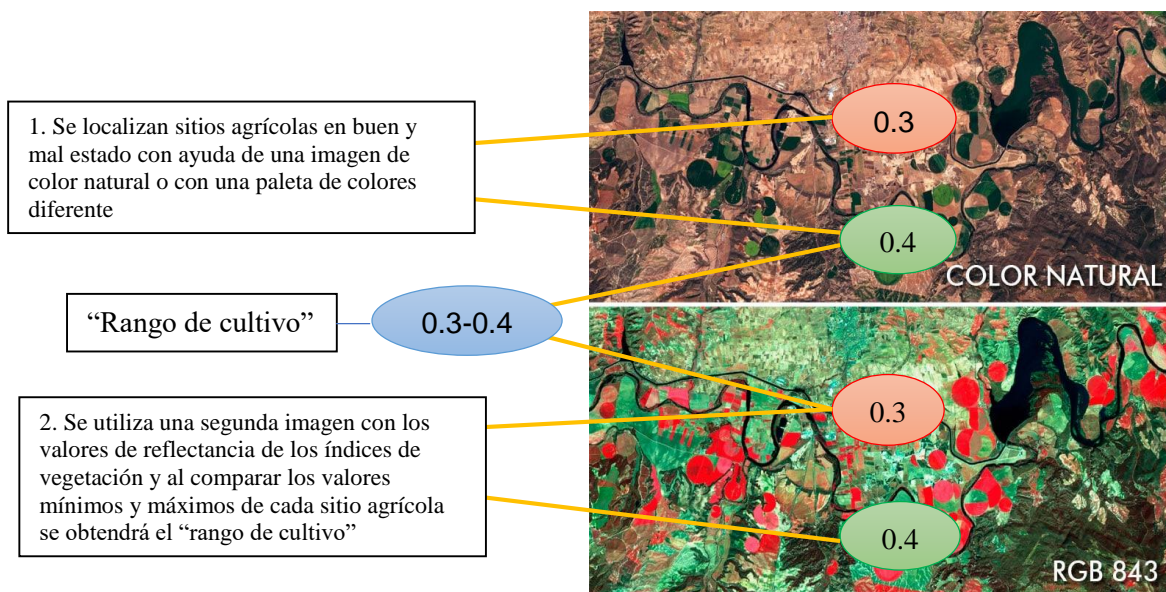
Las variables a considerar en esta investigación son los niveles de precipitación, temperatura, las pendientes del terreno en los sitios seleccionados y distintos índices de vegetación (NDVI, NDMI y SAVI). Este proceso se lleva a cabo, otorgando rangos a los valores de las variables anteriormente mencionadas y asignando el valor que más se compare con los niveles óptimos de la fisiografía del terreno que correspondería a cultivos.

De las variables anteriormente mencionadas solo se toman en cuenta las variables de pendiente y los índices de vegetación para formar parte del ráster multicriterio, ya que la distribución de valores en los índices de temperatura y precipitación suele ser muy general en la imagen y no suelen tener diferencias destacadas en la imagen. Por otro lado, las variables de los distintos índices de vegetación son utilizadas para determinar el tipo de cobertura, evaluar su variación temporal o establecer el estado de salud de cultivos a partir de estimaciones de características como vigor vegetal, contenido de clorofila, estado nutricional o estado hídrico. En cuanto la variable de pendiente, esta fue seleccionada debido a que diferentes tipos de cultivos requieren una pendiente suavizada, esto debido a que, a

mayor intensidad de pendiente, la velocidad que toma el agua al escurrir del terreno es mayor y por lo tanto es mayor la erosión del suelo.⁵⁵

Los índices de vegetación anteriormente mencionados deben pasar por un proceso de identificación de coberturas para cultivos, ya que con esto se identifican los valores de reflectancia idóneos de cada índice, esto con el objetivo de crear rangos de valores que puedan ser fácilmente identificados con un ráster multicriterio. Los valores adecuados variarían dependiendo de la clase de cultivos que contenga la imagen, por ello se pueden obtener valores de cada índice de vegetación comparando 2 sitios agrícolas de una misma imagen, seleccionando un campo con abundancia de cultivos en buen estado y otro con escasos cultivos o que sufra de estrés hídrico. Los sitios son seleccionados por medio de una imagen en color verdadero o falso color de manera que sean fácilmente identificables ambos tipos de cultivo (*Figura (13)*). Los valores de índice de vegetación que muestre cada sitio en un segundo ráster con dichos valores, mostrarán el rango límite que tendrá la clase deseada, ya que la clase se debe definir por valores idóneos que contengan vegetación alta, pero sin mostrar valores altos como se vería en bosques densos, ni muy baja como para llegar a los valores de suelo desnudo. El valor de reflectancia intermedia resultante de cada índice es la que se utilizará como base para definir a un campo de cultivo en una imagen en particular.

Se le ha denominado a este rango de valor intermedio **“Rango de cultivo”** que es un punto de referencia de los distintos rasters para la identificación de zonas agrícolas (*Figura (16)*). Cabe mencionar que el procedimiento es igual para todos los índices de vegetación, además el rango de cultivo solo corresponde a las parcelas que sean exclusivas para campos de cultivo que contengan al menos un mínimo porcentaje de vegetación, el cultivo en descanso puede entrar en esta categoría si respeta el rango mínimo que se le da al rango de cultivo.



(Figura 16) Imagen comparativa de ejemplo. Los sitios seleccionados pertenecientes a campos agrícolas se diferencian por el estado de salud de los cultivos que contengan (1) y posteriormente se obtiene esa diferenciación por sus valores de reflectancia de algún índice de vegetación (2). Referencia [Ñ]

La identificación de los cultivos en buen y mal estado puede depender de varios factores, el analista debe de consultar capas del sitio sobre el tipo de cultivo, la temporalidad de las cosechas, la fecha de toma de la imagen, entre otros, ya que se deben elegir las mejores opciones de comparación para que el rango de cultivo sea lo más preciso posible. Es por ello que, en el caso de cultivos en buen estado se pueden tomar casos como cosechas en penúltimas etapas de maduración, y que no cuenten con gran cantidad de follaje para no ser confundidas con vegetación natural (*figura (17)*) y en el caso de cultivos en mal estado, pueden ser parcelas recién cultivadas o con un estrés hídrico excesivo, sin llegar a la marchitación. (*figura (18)*).



Figura (17) Vegetación suficientemente densa en cultivos. Referencia [O]



Figura (18) Vegetación pobre en cultivos sin llegar a suelo desnudo. Referencia [P]

Cabe destacar que estos valores también pueden ser afectados por la cantidad de superficie que abarca la vegetación en un píxel (*figura (19)*), ya que, aunque la planta u árbol cuenten con buena salud, si la cobertura de vegetación no es suficientemente densa, los valores medidos de reflectancia pueden deberse, no sólo a la vegetación, sino también al suelo desnudo subyacente.¹⁹



Figura (19) En la imagen de la izquierda los pixeles tienen cobertura vegetal del 100%, en cambio en la imagen de la derecha, el pixel del centro, posee una cobertura del 15% aproximadamente, donde el valor registrado por sensor está fuertemente influenciado por la reflectancia del suelo. Referencia [Q]

Por lo anterior se puede interpretar que una alta densidad de vegetación siempre llevará a valores altos de reflectancia en cualquier índice, de esta manera se observa que lugares como bosques o selvas pueden diferenciarse de los cultivos gracias a su nivel de reflectancia, aunque también existe la posibilidad de haber cultivos con alta densidad como árboles con frutos, por lo que se requiere un mayor análisis de la zona de estudio para ajustar los valores de reflectancia en la clasificación.

Por otra parte, para el caso de la variable de pendiente, se recomiendan buscar sitios potenciales para cultivos que tengan un **rango de pendiente idónea** de entre 0.0% a 4.0% de porcentaje (o 2.29° grados de pendiente) pero en caso de analizar un sitio donde se cultive en altas elevaciones como montañas y cerros se le puede asignar un valor aceptable de 8° a 10° grados de pendiente, aunque en algunos casos se han llegado a ver cultivos en pendientes de hasta 22° grados.⁵⁶

Para mejorar la diferenciación de las coberturas de interés empleando algoritmos de clasificación supervisada se utiliza información disponible de la topografía de la zona de estudio, en este caso se emplea un Modelo Digital de Terreno (MDT).⁵⁷

Las variables anteriormente vistas (pendiente e índices de vegetación) serán utilizadas para la elaboración de un ráster multicriterio, donde se realizan gráficas de distribución de píxeles para cada variable en su correspondiente imagen ráster. Las gráficas muestran la cantidad de reflectancia del índice y la cantidad de píxeles que se detectaron en cada valor de reflectancia. Posteriormente mediante una distribución de datos utilizando el modelo de clasificación de rupturas naturales de Jenks (*natural breaks*) se busca minimizar las diferencias entre valores de datos en la misma clase y maximizar las diferencias entre clases, dichas clases tendrán una distribución en la que pueda ser fácilmente diferenciada entre la cantidad de píxeles en una imagen y el valor que le corresponde en cada variable vista.⁵⁸ (figura (20))

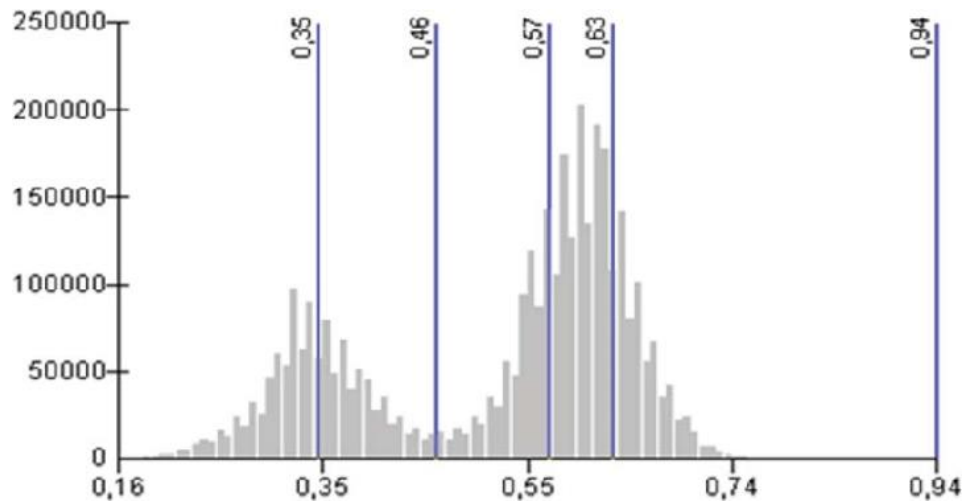


Figura (20) Ejemplo de distribución de valores mediante Natural Jenks dividida en 5 clases. Referencia [R]

Una vez obtenida la gráfica, las clases en las que fue dividida serán utilizadas para la asignación de pesos en el ráster multicriterio, donde la clase más importante y por lo tanto que tendrá más peso, será la que tenga el rango de valores asignados como **rango de cultivo y rango de pendiente idónea** obtenidos de la reclasificación de variables mencionado en párrafos anteriores. También es posible modificar el rango de las clases si la distribución con *Natural Jenks* no muestra la distribución deseada ya que los pesos se ven afectados por distintas características del terreno y la vegetación, por ello es importante tener en cuenta que los porcentajes variarán según estas condiciones o incluso otros factores.

Después de asignar los pesos que tendrá cada clase en cada variable, se obtiene un **ráster multicriterio** combinando todas las características de todas las variables divididas en clases, donde la clase con más importancia, será la que se relacionó con el rango de cultivo y pendiente idónea, y por lo tanto la clase con más probabilidad de hallar un cultivo en óptimas condiciones.

III.2 Polígonos de entrenamiento

Los polígonos de entrenamiento son creados de las clases en las que se dividió el ráster multicriterio, principalmente en la clase que se obtuvo a partir de las variables vistas anteriormente (índices de vegetación y pendiente), las regiones de la imagen en la que se encuentra dicha clase principal serán donde los polígonos de entrenamiento tendrán prioridad de colocación, ya que son sitios idóneos donde el rango de cultivo y el rango de pendiente idónea, conforman condiciones propicias para hallar cultivos, o sitios potenciales para ser cultivados.

Además del raster multicriterio, el analista deberá de basarse también en imágenes en color verdadero para asegurarse que las zonas seleccionadas para la colocación de los polígonos de entrenamiento sean las adecuadas (*Figura (21)*). Si se requiere un análisis del terreno con más profundidad, también se puede recurrir a capas de información vectorial o rasterizadas, de distintas investigaciones o proyectos previos en el sitio de estudio y realizar un *time-lapse* que ayude a visualizar los cambios temporales del sitio.

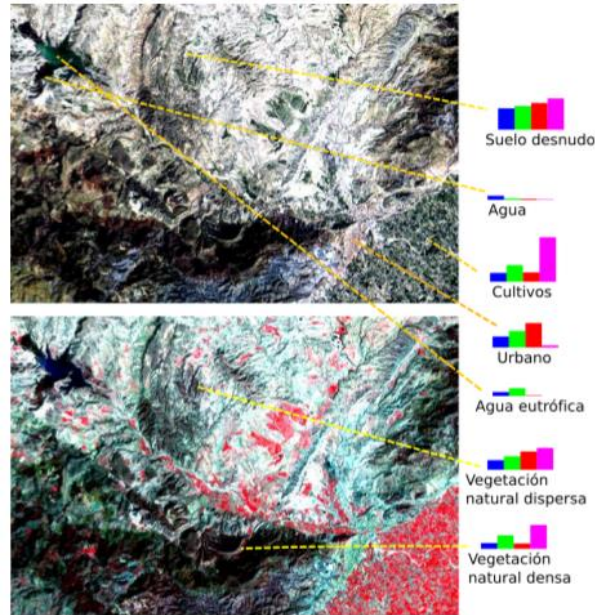


Figura (21) Ejemplo de la identificación de clases, en distintos sitios de la imagen con ayuda de combinación de bandas e imágenes en color verdadero. Referencia [S]

Aunque el polígono de entrenamiento principal a colocar será el de la **clase cultivo**, de igual manera los polígonos de entrenamiento que conforman otro tipo de clases serán importantes para diferenciar el objeto de estudio (en este caso cultivos) de otro tipo de objetos que pueden encontrarse en la imagen satelital del sitio de estudio.

Las clases en las que se deberán dividir los polígonos de entrenamiento deberán tener en cuenta 2 factores principales, los cuales son:

- El número de clases
- El tipo de clases que contendrá cada sitio de estudio

El número de clases es importante en cada clasificación ya que determinaran el rango de exactitud al decidir qué clase pertenece a cada pixel en una imagen. Si existen numerosas clases el algoritmo puede confundir clases entre sí, dando como resultado pequeños pixeles de distintas clases un lugar reducido, minimizando así el área de los polígonos que conforman cada clase (Figura (22)). Ocurre lo mismo si el número de clases es muy bajo (normalmente menos de 4 ó 3) ya que éstas pueden confundirse entre sí causando que el algoritmo aumente el riesgo de una mala clasificación al tener un sesgo mayor entre los valores de cada pixel, causando así un porcentaje alto de error en la matriz del algoritmo al no reconocer que se trata de un nuevo tipo de objeto, ocasionando que se clasifique como la clase que tenga el valor digital más próximo aunque esta no represente a dicha sección de la imagen.



Figura (22) Ejemplo de una clasificación supervisada con exceso de clases. Referencia [T]

Por otra parte el tipo de clases utilizadas para la clasificación es igual de vital importancia, estas deberán dividirse en clases que representen objetos del terreno que tengan grandes diferencias espectrales entre sí, la razón de esto, es que *random forest* suele tener errores en la clasificación de una imagen satelital, si las clases comparten un rango de reflectancia similar, los valores de los píxeles de cada ráster también serán similares, por lo tanto el algoritmo puede colocarlos en la misma clasificación, aunque sean objetos completamente diferentes. Un ejemplo de esto, son los cuerpos de agua que contienen grandes cantidades de algas u otro tipo de vegetación, por lo que si no se crea una clase que represente cuerpos de agua, dicha clase podría formar parte de la clase de vegetación natural o cultivo, aunque tenga valores espectrales diferentes.

Para la clasificación estándar, se suele buscar clases principales que estén presentes en todas aquellas imágenes satelitales que se utilicen para este tema de esta investigación o similares, algunas de estas clases son:

- Área Urbana
- Cerro o suelo desnudo
- Campos de cultivo
- Cuerpos de agua
- Vegetación natural
- Nubes

- Sombras de nubes

No es descartable introducir clases que tengan una presencia importante en la imagen, como pueden ser; carreteras de concreto, caminos de tierra, nieve, arena o incluso campos de flores. A su vez también se pueden descartar clases como nubes o sombras de nubes si su presencia es muy poca o nula en la imagen.

Aunque las clases antes mencionadas son importantes para que el algoritmo de *random forest* tenga una mínima confusión en la clasificación, para esta investigación, la colocación de la clase que representa campos de cultivo es fundamental para la delimitación de la frontera agrícola. Es por ello que al colocar los polígonos de entrenamiento de la clase mencionada se deben tener en cuenta el **ráster multicriterio** y los valores espectrales que se establecieron en los rangos de cultivo y pendiente, un ejemplo de esto es no colocar polígonos de la clase cultivo en pendientes abruptas, aunque las imágenes en color verdadero y ráster con índices de vegetación indiquen presencia de cubierta vegetal.

En la mayoría de casos, las imágenes tendrán 2 tipos de cultivos que se agrupan en la misma clase; por un lado, existen cultivos con siembra o pasto y por otro los cultivos con suelo en descanso o abandonados. Dado que estos últimos contienen poca o nula materia vegetal, serán incluidos en otra clase o en su defecto se incluirán con la clase cultivo tomando en cuenta no exceder del rango de cultivo establecido.

Tomar en cuenta el contorno de la parcela que conforman los campos de cultivo también es importante para la colocación de esta clase, ya que usualmente los campos de cultivos tienen un margen que delimita su espacio del exterior y que es relativamente fácil de identificar en una imagen satelital.

III.3 Procesamiento del algoritmo de *random forest*

Una vez analizadas las características del terreno con la metodología vista anteriormente, se colocan los polígonos de entrenamiento en formato vectorial de manera que coincidan de forma geoespacial con las imágenes satelitales del sitio de estudio y posteriormente son almacenados en un solo archivo.

Para que el proceso de clasificación en *radom forest* se realice, es necesario que el algoritmo cuente con 2 archivos: una pila de imágenes ráster y una capa vectorial que contenga los polígonos de entrenamiento.

La pila de imágenes raster será una compilación de todas las imágenes que contengan valores digitales de distintas características del terreno, así como de las bandas espectrales e índices de vegetación. Las imágenes introducidas para esta investigación son las siguientes:

- 13 imágenes satelitales del sitio de estudio que conforman cada una de las bandas espectrales de la constelación *Sentinel 2*
- Una imagen ráster del sitio de estudio que contenga el modelo digital de pendientes (MDP)
- Una imagen ráster que contenga valores de precipitación anual del sitio
- Una imagen ráster que contenga la temperatura anual del sitio
- Una imagen ráster que contenga una clasificación previa de la cobertura vegetal del sitio (opcional)

Además de las imágenes mencionadas anteriormente, se incluirán rústers con los valores de los índices de vegetación creados a partir de las bandas espectrales del sitio. Los índices incluidos son: NDVI, NDBR, NDMI, SAVI y PVI.

Cabe destacar que los rústers con variables de precipitación, temperatura, pendiente entre otros sirven de apoyo para que el algoritmo de *random forest* se base en otro tipo de características del terreno además de los índices de vegetación, esto ayuda a que el algoritmo tenga otros valores de selección al momento del embolsado (*Bagging*) y así obtenga más variedad de datos al momento de obtener muestras por medio de los árboles de decisión.

Por último, la pila de imágenes será comprimida en un solo ráster el cual será la muestra base donde se creará una nueva clasificación también en forma de imagen ráster.

Una vez que se cuentan con ambos archivos (La pila de imágenes y los polígonos de entrenamiento) el algoritmo de *random forest* se procesa en un software que pueda correr algoritmos utilizando lenguaje de programación compatibles con archivos de índole estadístico y gráficos. Para esta investigación se utilizó **RStudio** que es un conjunto de programas integrados para el manejo de datos, simulaciones y cálculos.⁵⁹

El lenguaje de programación de **R** se enfoca principalmente a análisis estadísticos y base de datos, además cuenta con librerías que permiten importar automáticamente herramientas que ahorran tiempo de código. El algoritmo diseñado (*Anexo 1*) utiliza librerías de rasterización, estadística y por supuesto un modelo de *random forest* que haga la clasificación.

Los archivos principales (pila de imágenes y polígonos de entrenamiento) se añaden al algoritmo donde estos últimos pasan por un proceso de rasterización donde los valores vectoriales son transformados en píxeles, posteriormente las clases rasterizadas de los polígonos de entrenamiento se les asigna un nivel digital de la pila de imágenes por lo que ahora serán considerados ráster de entrenamiento.

Una vez que se les asigno un conjunto de valores digitales al ráster de entrenamiento, los valores son reagrupados por filas y columnas donde cada columna representara un conjunto de datos auxiliares ráster, es decir las bandas y otros ráster de la pila de imágenes creada, por lo tanto, las filas representaran los valores digitales de cada ráster de entrenamiento.

Por último, se añade el paquete de *random forest* donde se ajustan los datos por defecto para el modelo, tales como el número de árboles de decisión, semillas de aleatoriedad y la cantidad mínima de muestras. Una vez realizado el proceso, se obtendrá una matriz de confusión del conjunto de datos y se le dará la indicación al algoritmo de generar también una imagen ráster de la clasificación final de la pila de datos con forme a las clases que le asignó *random forest* basándose en los ráster de entrenamiento. ⁶⁰

III.3.1 Matriz de confusión

Los resultados obtenidos en la clasificación de *random forest* son puestos en una matriz de confusión, la cual permite visualizar el desempeño del algoritmo en cuanto a los aciertos y errores del proceso.

Esta matriz contiene los valores de las clases en las que hubo aciertos y discrepancias, también conocidas como verdaderos positivos/negativos y falsos positivos/negativos. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real (*Figura (23)*). ⁶¹



Figura (23) Representación de una matriz de confusión, el número de columnas y filas será equivalente al número de clases que se hayan propuesto para el algoritmo. Referencia [U]

Random forest generará una matriz muestra de entre todos los árboles de decisión que haya creado (generalmente 500). Dicha matriz será elegida por el algoritmo como la más óptima por el menor porcentaje de error obtenido, es decir el menor número de falsos positivos y negativos en los valores de las clases.

La matriz se conforma por filas y columnas, de las cuales los valores que se encuentran en la diagonal de la matriz serán conformados por los verdaderos positivos (es decir clases que

corresponden a la predicción creada por el algoritmo), mientras que el resto de los números que no corresponden a la diagonal se consideraran falsos positivos y falsos negativos (clases que el algoritmo confundió con otras clases). (Figura (24))

		Predicted			
		On time	Late	Very late	
Actual	On time	40	7	3	50
	Late	5	25	5	35
	Very late	3	5	8	15

Figura (24) Los valores mostrados en verde conforman la diagonal de la matriz en la cual se encuentran los valores coincidentes entre el algoritmo y el valor real, mientras que los números en rojo representan el número de falsas predicciones. Referencia [V]

Una vez obtenida la muestra de matriz de confusión es tarea del analista buscar los errores o falsos positivos/negativos que se obtuvieron en las clases, por lo tanto, existen parámetros para saber si la matriz es aceptable o se requieren reclasificaciones y distintas pruebas en el algoritmo que arrojen matrices más exactas.

Existen diferentes formas de medir la asertividad de la matriz, una de ellas es la precisión general, la cual se calcula sumando el número de valores correctamente clasificados y dividiéndolo por el número total de valores. Aun que usualmente la más utilizada es el coeficiente *kappa* que mide la concordancia entre la clasificación y los valores de verdad. Un valor *kappa* de 1 representa un acuerdo perfecto, mientras que un valor de 0 representa ningún acuerdo.⁶²

En el caso de esta investigación, el propio algoritmo *random forest* proporciona una media para medir el *error rate* (índice de error) llamado OOB (Error fuera de bolsa por sus siglas en ingles), también llamado estimación fuera de bolsa, es un método para medir el error de predicción de bosques aleatorios, árboles de decisión potenciados y otros modelos de aprendizaje automático que utilizan agregación de arranque (*bagging*).⁶³

Esta estimación del error es ciertamente el más preciso para el tipo de clasificación que realiza *random forest*, sin embargo, un estudio realizado por investigadores alemanes Silke Janitza y Roman Hornung, ha demostrado que el error fuera de bolsa se sobreestima en entornos que incluyen un número igual de observaciones de todas las clases de respuesta, por lo que también es recomendable optar por otros índices de error.⁶⁴

Para el caso de esta investigación los parámetros de error aceptables que debe de arrojar la matriz de confusión son de menos de 2% para el índice OOB y no menos de 0.95 para la índice *kappa*. A su vez el valor indicado en las clases que están dentro del grupo de falsos positivos/negativos, no debe exceder el número 20 en cualquier clase.

Si el porcentaje de error excediera estos límites, se debe reconsiderar volver a ejecutar el algoritmo de *random forest* para crear una nueva matriz de confusión o modificar las clases ya sea disminuyendo o aumentando el número de polígonos de entrenamiento dependiendo en qué clase se encuentren los falsos valores.

Capítulo IV. Resultados y discusión

IV.1 Resultados de la reclasificación de variables

Rangos de cultivos

Después de analizar los ráster de índice de vegetación de cada sitio, los valores obtenidos para el rango de cultivo de cada variable son los siguientes:

Tabla (4) Tabla del rango de cultivo resultante para los distintos índices de vegetación en la región del Valle del Yaqui.

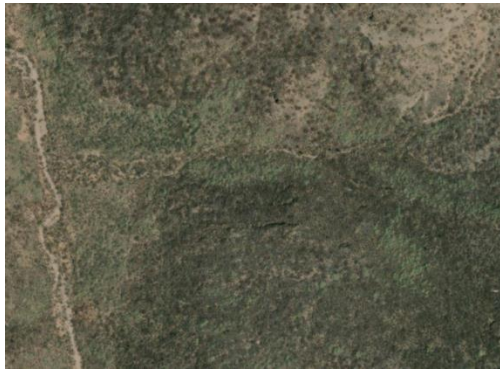
Índices de vegetación	Rango de cultivo según las observaciones
NDVI	0.25 – 0.72
NDMI	0.05 – 0.67
SAVI	0.4 – 1.05
PVI	0.39 – 1
NDBR	0.15 – 0.68

Autoría propia

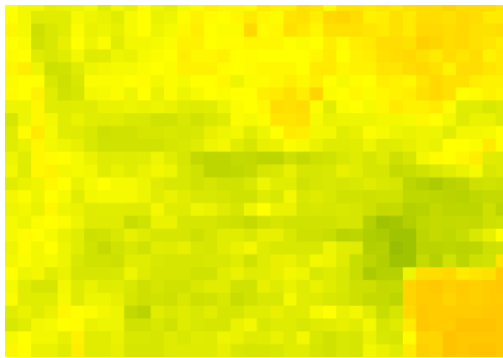
Observaciones:

- Los rangos de cultivo en esta región alcanzaron valores cercanos a 1 en los valores positivos de los índices de vegetación, ya que la vegetación más sana y con mayor reflectividad se encontraba principalmente en los cultivos y no en el follaje natural.
- El índice DNBR no se tomó en cuenta en la reclasificación de índices ya que sus características de reflectancia eran similares a las del NDMI y el resultado final se comprueba obteniendo un rango de cultivo similar.
- El índice PVI jugó un papel importante en la delimitación del rango de cultivo que permitió observar en el ráster como la mayoría de las zonas geográficas con “pendiente elevada” y con escasa vegetación son excluidas como parte de una cubierta vegetal importante, tal como son cultivos, los cuales no se verían afectados, y tendrían un índice PVI alto mientras tengan una cobertura vegetal uniforme y

consistente. En el área de ejemplo que muestran las (figuras 25)) se observa como el indicador de vegetación mostrado en el índice NDVI (B) es remplazado por un indicador de vegetación bajo o nulo que pertenece al raster resultante del índice normalizado PVI (C), lo que aportó una ayuda importante al observador al obtener un indicador de reflectancia más claro que una imagen en color verdadero (A) para la delimitación de las áreas de vegetación que favorece la asignación de sitios potenciales para la colocación de polígonos de entrenamiento.

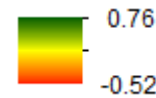


(Figura 25 A)

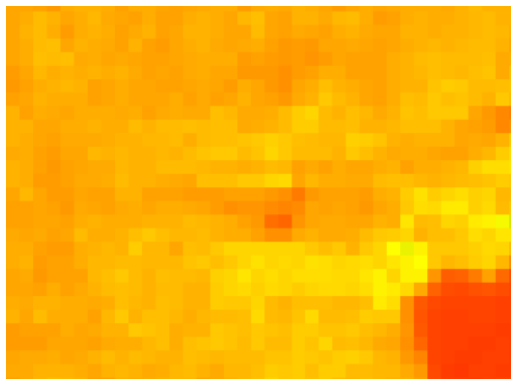


Índice de reflectancia

NDVI

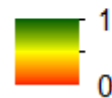


(Figura 25 B)



Índice de reflectancia

PVI



(Figura 25 C)

Figuras (25 A, 25 B y 25C) Comparación de rasters para el descarte de posibles zonas de vegetación y/o cultivo utilizando una imagen en color verdadero (25 A), y los índices de reflectancia del NDVI (25 B) y PVI (25 C). Un índice alto simboliza vegetación mientras que un índice bajo simboliza poca o nula vegetación. Autoría propia

Tabla (5) Tabla del rango de cultivo resultante para los distintos índices de vegetación en la región que conforman los municipios de Xochistlahuaca y Tlacoachistlahuaca.

Índices de vegetación	Rango de cultivo según la reflectancia
NDVI	0.53 – 0.79
NDMI	0.02 – 0.33
SAVI	1 – 1.2
NDBR	0.21 – 0.67

Autoría propia

Observaciones:

- Al tener una vegetación abundante, el rango de cultivo de esta región es menos amplio. Las características fisiológicas que comparten los cultivos sanos y la vegetación natural resultan similares, por lo cual un rango de cultivo reducido ayudó a destacar las características de un cultivo más eficientemente.
- Nuevamente el índice DNBR no se tomó en cuenta en la reclasificación de índices ya que sus características de reflectancia eran similares a las del NDMI.

Rangos de pendiente idónea

El rango de cultivo que se le otorgó a la pendiente en la región del Valle del Yaqui fue de 0 a 2.4 grados debido a que la gran mayoría de los cultivos no sobrepasaba este límite por las condiciones planas del terreno.

Por otra parte, la superficie del terreno en los municipios de Tlacoachistlahuaca y Xochistlahuaca supera los 8° de pendiente permitidos en un 93%, por lo tanto, el rango de pendiente también fue de 0 a 8° que es el límite establecido para el crecimiento adecuado de un cultivo.

Análisis del ráster multicriterio

El análisis de los índices de vegetación se llegó a la conclusión de asignar un porcentaje aproximado de pesos para cada índice de vegetación según las características del terreno. La (tabla (4)) muestra algunas opciones de porcentaje para tener en cuenta dependiendo la zona en que se realice la evaluación.

Tabla (6) PORCENTAJE DE PESOS RECOMENDADOS PARA LA EVALUACIÓN MULTICRITERIO

	Pendiente	NDMI	NDVI	SAVI	PVI
Para zonas con poca pendiente y alta vegetación natural	Igual o Menos de 10%	25% - 30%	25% - 30%	30% - 40%	Igual o Menos de 10%
Para zonas con poca pendiente y baja vegetación natural	30% - 40%	10% - 15%	15% - 20%	20% - 25%	30% - 40%
Para zonas con alta pendiente y alta vegetación natural	15% - 20%	15% - 20%	20% - 25%	30% - 40%	15% - 20%
Para zonas con alta pendiente y poca vegetación natural	20% - 30%	15% - 25%	15% - 25%	15% - 25%	20% - 30%

Autoría propia

En el ráster multicriterio las variables utilizadas fueron asignadas de acuerdo con los pesos asignados del ráster multicriterio se encuentran resumidos en la siguiente tabla (*Tabla (7)*). En ambos casos se conservaron los índices de vegetación NDVI, NDMI y SAVI además de la variable de pendiente.

Tabla (7) Porcentaje de pesos para la evaluación multicriterio en el Valle del Yaqui

Índices de vegetación y pendiente	Porcentaje de pesos asignados
NDVI	20%
NDMI	10%
SAVI	15%
PVI	25%
Pendiente	30%

Autoría propia

Observaciones:

- En la región desértica del Valle del Yaqui los índices de vegetación cobraron poca relevancia debido a que la mayor parte de vegetación es conformada por cultivos cuya reflectancia es muy alta a diferencia de la vegetación natural que cuenta con valores bajos, lo que ocasiona una diferencia grande entre ambas clases, por lo que los valores de los índices de vegetación NDVI, NDMI y SAVI no obtuvieron un peso importante ya que se consideraba que la mayor parte de porción del terreno que tuviera índices

de vegetación con niveles altos, se consideraría cultivo. Por lo tanto, no era necesario resaltar mucho en el ráster multicriterio las diferencias espectrales entre ambos tipos de vegetación (cultivo y vegetación natural).

- La variable de pendiente y el índice PVI ocupan los pesos más altos con un 30% y 25% de relevancia respectivamente, ya que, se enfocan en desestimar la poca vegetación natural que se encuentra en terrenos elevados y/o con estrés hídrico.
- El resultado mostró que en el sitio existe un 31% de cobertura susceptible a pertenecer a algún cultivo o a ser cultivados, principalmente en zonas húmedas, en constante riego o de forma natural (*Figura (26)*).
- La pendiente toma poca relevancia ya que sólo se registró un valor mayor a 8° en un 6% de la región, que es mayor a lo que usualmente necesita una tierra cultivable. Esto permitió que el ráster se enfocara en remarcar como lugares no validos aquellos que sufran de una pendiente alta, lo cual es un indicativo de poca probabilidad de cultivo en la región.

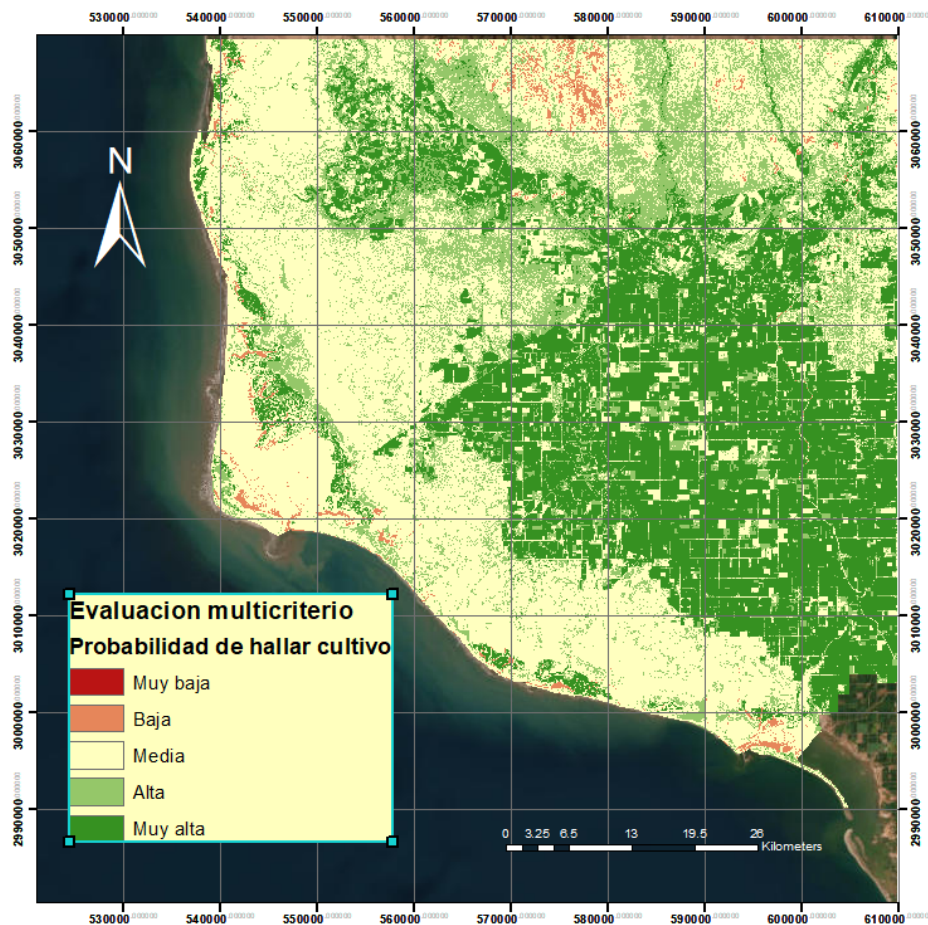


Figura (26) Mapa del ráster multicriterio resultante que muestra la probabilidad de hallar cultivo en el Valle del Yaqui. Autoría propia

Tabla (8) Porcentaje de pesos para la evaluación multicriterio en los municipios de Xochistlahuaca y Tlacoachistlahuaca

Índices de vegetación y pendiente	Porcentaje de pesos asignados
NDVI	25%
NDMI	25%
SAVI	30%
PVI	N/A
Pendiente	20%

Autoría propia

Observaciones:

- Para la región que conforma los municipios del estado de Guerrero estudiados, los índices de vegetación tuvieron más importancia ya que el 91% de la región se conforma por vegetación natural y cultivos.
- Por otro lado, el ráster multicriterio mostró que el 41% de la región es ideal para la presencia de cultivos, principalmente en las regiones bajas y cerca de poblaciones. *(Figura 27)*
- La pendiente toma una relevancia mayor en este caso ya que 29% de la región corresponde a una pendiente menor a 8°, por lo que estos lugares son los más idóneos para la presencia de cultivos aun cuando los resultados mostraron que el 36.7% de los cultivos se encontraban en pendientes superiores a esta medida, de los cuales el 6% supera los 20° de pendiente.

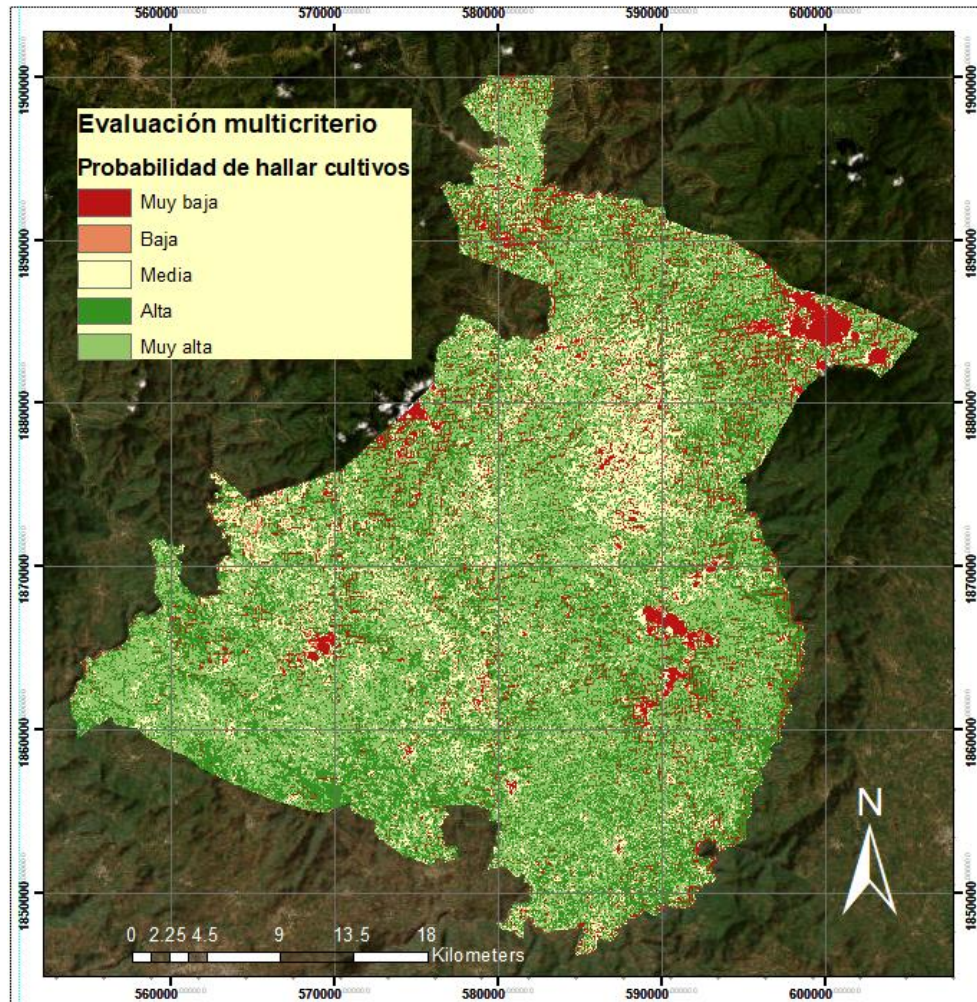


Figura (27) Mapa del ráster multicriterio resultante que muestra la probabilidad de hallar cultivo en los municipios de Tlacoachistlahuaca y Xochistlahuaca. Autoría propia

IV.2 Clasificación resultante del algoritmo *Random forest*

El resultado de la clasificación de *random forest* dio como resultado 2 raster de los sitios evaluados clasificados de acuerdo a los elementos más importantes fisiográficamente hallados en las zonas.

Municipios de Xochistlahuaca y Tlacoachistlahuaca.

A continuación, se describe el porcentaje de cobertura resultante de la clasificación final para cada una de las clases que asignaron en el sitio de estudio de los municipios de Xochistlahuaca y Tlacoachistlahuaca.

Tabla (11) Porcentaje de cobertura de clases para los municipios de Xochistlahuaca y Tlacochistlahuaca.

Clases	Porcentaje de cobertura
Cultivo	29.63%
Camino de asfalto	0.18%
Camino de terracería	0.53%
Área urbana	1.43%
Vegetación natural	50.71%
Cuerpos de agua	0.46%
Cerros	14.56%
Sombra de nubes	0.55%
Nubes	1.92%

Autoría propia

A partir del raster de clasificación se obtuvo un mapa (*Figura (28 B)*) en donde se destaca una cobertura de vegetación natural importante, debido a las condiciones climatológicas de la región, los resultados muestran que los campos de cultivo, ganaderos, pastoreo u otros semejantes representa aproximadamente el 29.63% de la región concentrando su presencia principalmente en el sur y centro de ambos municipios, contando también con pocos caminos y áreas pobladas lo que representa un 2% de la región.

Por otro lado, debido a que se observan nubes y sus sombras en la imagen satelital analizada, no fue posible obtener información de aproximadamente el 2.5% del territorio.

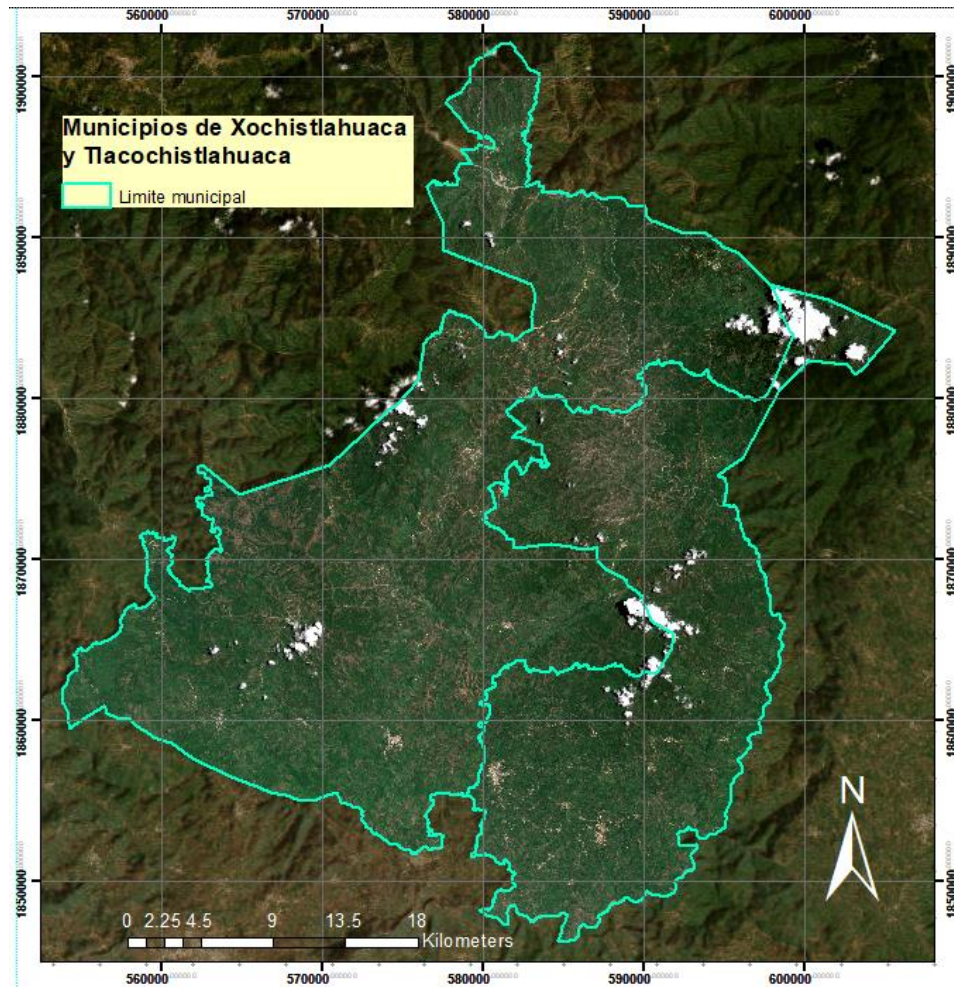
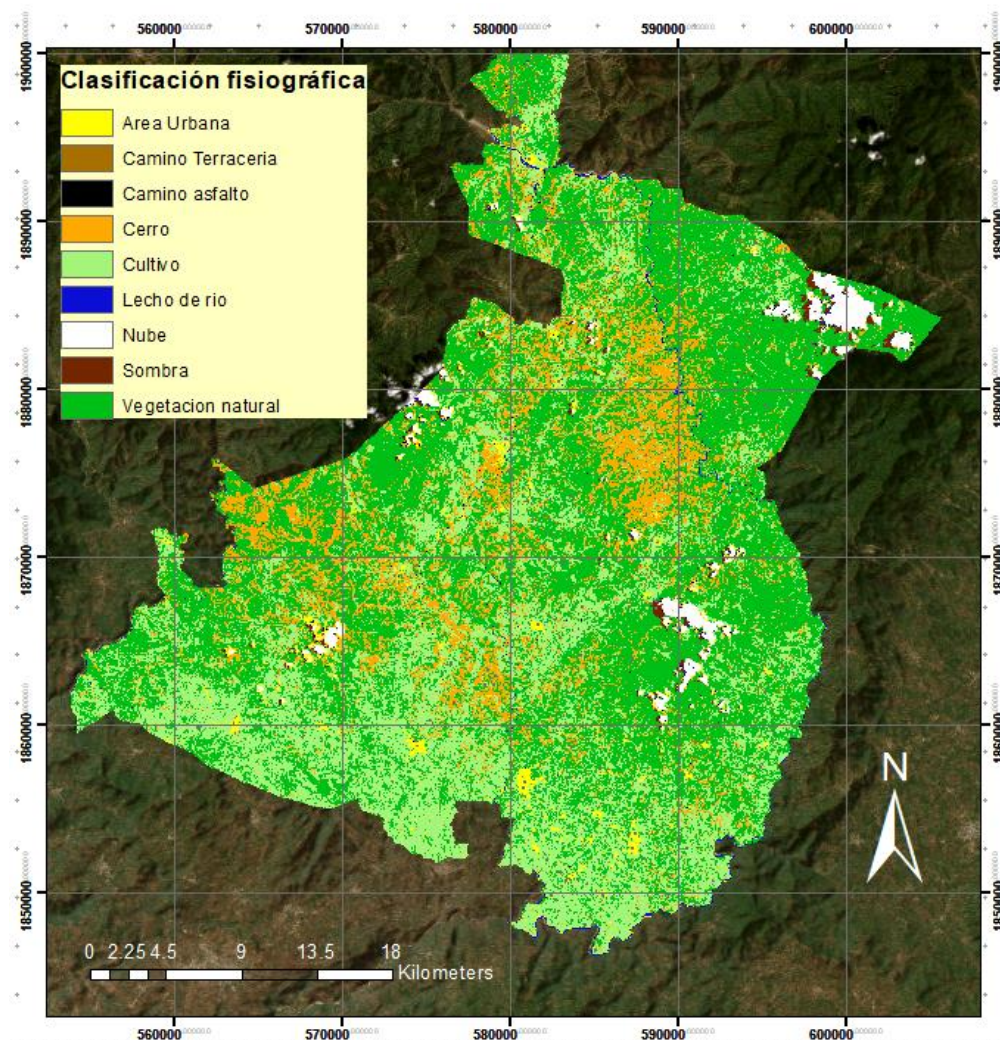


Figura (28 A)

Mapa en color verdadero del sitio. Autoría propia



(Figura (28 B)

Mapa fisiográfico resultante de la clasificación de random forest para los municipios de Tlachostlahuaca y Xochistlahuaca. Autoría propia

Valle del Yaqui.

La clasificación fisiografía del Valle del Yaqui dio como resultado un ráster donde se muestran las clases de cobertura de una manera gráfica. A continuación, se muestran los porcentajes de cobertura de cada clase establecida:

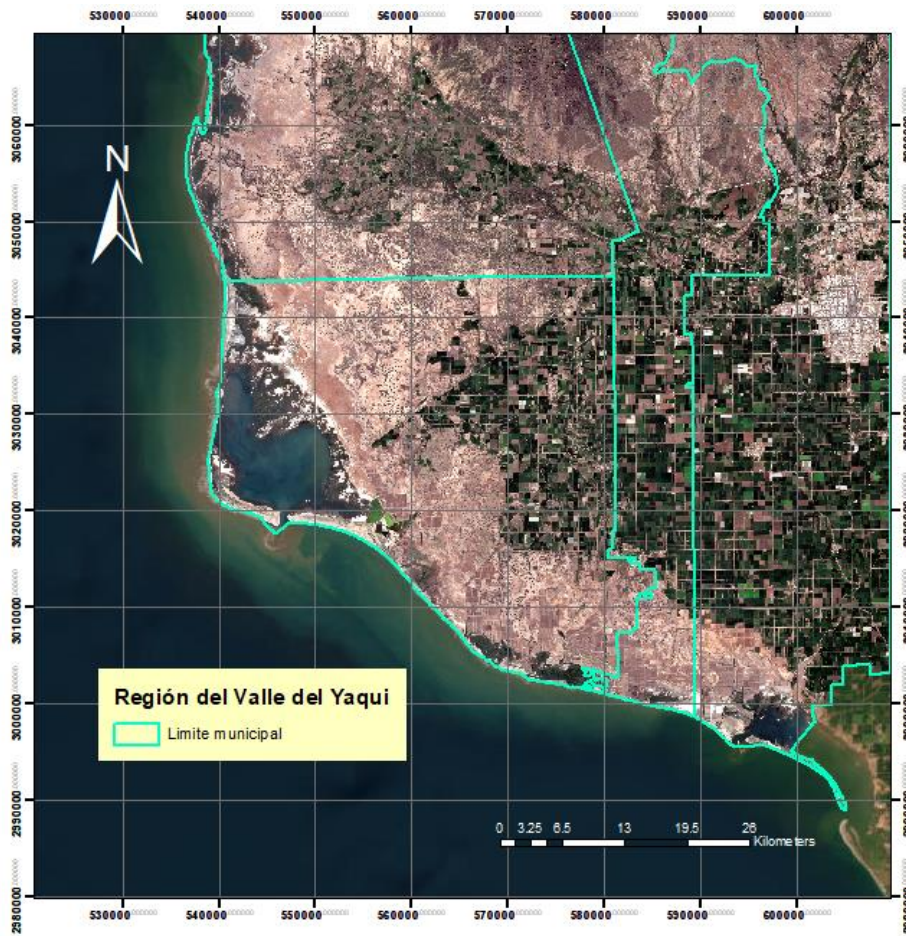
Tabla (12) Porcentaje de cobertura de clases para la región del Valle del Yaqui.

Clases	Porcentaje de cobertura
Arena y sedimentos	24.63%
Matorrales	16.98%
Cuerpos de agua	1.23%
Área urbana	3.08%
Cerros	9.52%
Cultivos	43.52%
Manglares	1.01%

Autoría propia

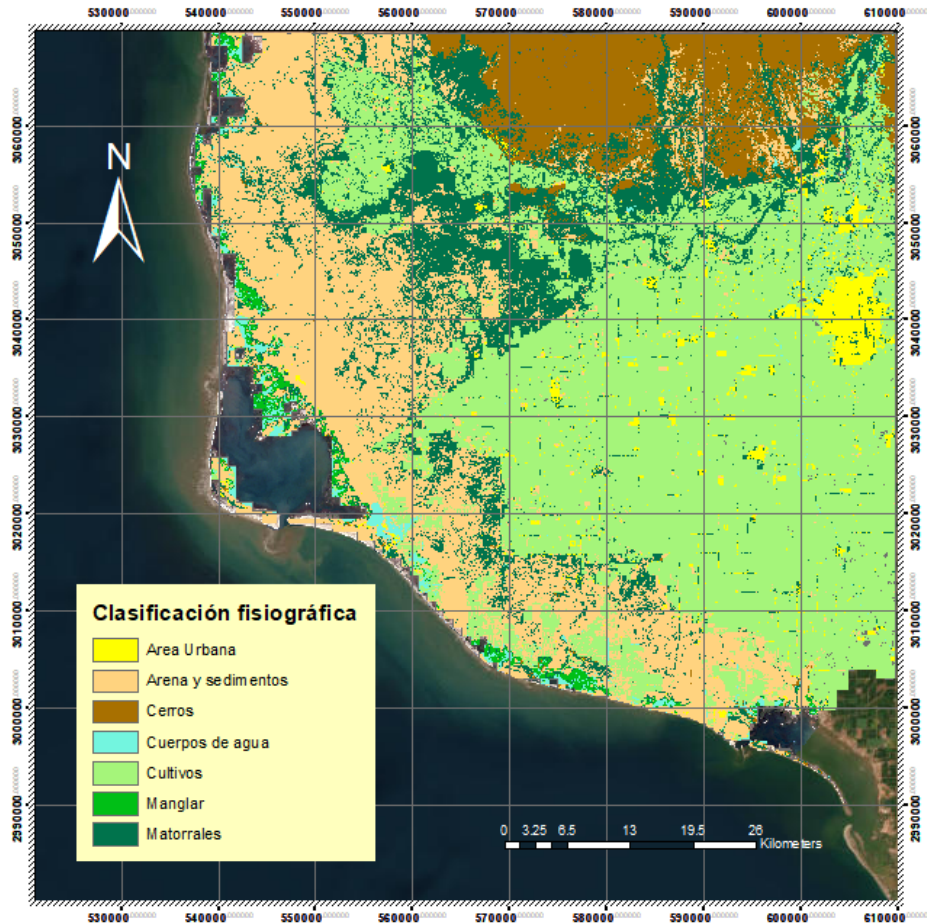
En el mapa de clasificación obtenido (*Figura (29 A)*) se puede observar que los campos de cultivo tienen la mayor parte del territorio con aproximadamente un 43.52% de cobertura, a su vez la arena y otros sedimentos tienen la segunda mayor cobertura con un 24.63% estando principalmente cerca de la costa junto con manglares y matorrales como vegetación natural cuya presencia es aproximadamente de un 18% en toda la región.

Cabe destacar que, aunque la presencia de área urbana sea de un 3.08%, la región tiene una mayor extensión territorial que el anterior sitio de los municipios de Guerrero, siendo Ciudad Obregón, al este de la región la mancha urbana más grande en el territorio analizado.



(Figura (29 A)

Mapa en color verdadero del sitio. Autoría propia



(Figura (29 B))

Mapa fisiográfico resultante de la clasificación de random forest para la región del Valle del Yaqui.
 Autoría propia

IV.3 Matriz de confusión resultante

El algoritmo de *random forest* utilizado dio como resultado una matriz de confusión donde se indica una muestra del árbol de decisión con más precisión de todo el bosque aleatorio. En esta, se muestra el número de clases y el número de predicciones correctas (falsos positivos/negativos y verdaderos positivos/negativos) que hubo en la compilación de variables y bandas de las imágenes satelitales.

Matriz de confusión resultante para la los municipios de Xochistlahuaca y Tlacoachistlahuaca

Para la clasificación fisiográfica de los municipios de Tlacoachistlahuaca y Xochistlahuaca se dibujaron 354 polígonos de entrenamiento, conformados por distintas clases que resaltaban diferentes niveles de reflectancia. Las clases con mayor importancia como vegetación y cultivo obtuvieron un mayor número de polígonos para obtener una mejor clasificación en

estas categorías. El número de polígonos por cada clase utilizada se señala en la siguiente tabla:

Tabla (9) Número de polígonos realizados por clase en los municipios de Xochistlahuaca y Tlacoachistlahuaca

No. Clase	Clases	No. De polígonos
1	Cultivo	128
18	Camino de asfalto	9
19	Camino de terracería	24
20	Área urbana	24
23	Vegetación natural	60
26	Cuerpo de agua	27
29	Cerro	51
30	Sombras de nubes	14
34	Nube	17

Autoría propia

La matriz de confusión resultante (*Figura (30)*) dio un índice de estimación de error (OBB) de 0.91%, y un índice kappa de 0.9266825, la cual cae en los límites de error aceptados, considerando el tamaño del sitio y el número de polígonos de entrenamiento creados.

```
Call:
  randomForest(x = x, y = y, mtry = min(param$mtry, ncol(x)))
  Type of random forest: classification
  Number of trees: 500
  No. of variables tried at each split: 2

  OOB estimate of error rate: 0.91%
Confusion matrix:
  1 18 19 20 23 26 29 30 34 class.error
1 1026 0 0 0 0 0 0 0 0 0.00000000
18 1 6 1 0 0 0 0 0 0 0.25000000
19 2 0 24 2 0 0 0 0 0 0.14285714
20 3 0 0 128 0 0 0 0 0 0.02290076
23 0 0 0 0 282 0 0 0 0 0.00000000
26 0 0 0 0 0 41 0 0 0 0.00000000
29 7 0 0 0 0 0 156 0 0 0.04294479
30 0 0 0 0 0 0 0 28 0 0.00000000
34 0 0 0 0 0 0 0 0 55 0.00000000
> print("Fin programado")
[1] "Fin programado"
```

Figura (30) Matriz de confusión de la clasificación resultante en los municipios de Xochistlahuaca y Tlacoachistlahuaca. Autoría propia

La muestra de datos que utilizó *random forest* para crear la matriz de confusión indica que hubo algunas clases con falsos negativos que señalan errores de clasificación. Estos errores se encuentran principalmente en las clases asignadas con el número 1 y 29, los cuales corresponden a cultivo y cerro respectivamente, así como también, la clase número 3 que

corresponde a áreas urbanas. Siendo así, que el número máximo de falsos negativos encontrados en la clase cultivo fue de 7, sin embargo, se estableció que el número máximo de falsos negativos/positivos que se podría encontrar en una matriz con menos de 2000 muestras sería 20, por lo que el número errores encontrados es aceptable considerando los parámetros establecidos.

Matriz de confusión resultante para la region del Valle del Yaqui

Por otro lado, para la clasificación fisiográfica del Valle del Yaqui se dibujaron 579 polígonos de entrenamiento, un número mayor a la región anterior debido a que tiene una mayor extensión territorial. Nuevamente la clase cultivo recibió un mayor número de polígonos dada su importancia en la clasificación. El número de polígonos por cada clase utilizada se señala en la siguiente tabla:

Tabla (10) Número de polígonos realizados por clase en el Valle del Yaqui

No. Clase	Clases	No. De polígonos
1	Arena y sedimentos	80
2	Matorrales	80
3	Cuerpo de agua	60
4	Área urbana	33
5	Cerro	48
6	Cultivo	252
7	Manglar	26

Autoría propia

La matriz de confusión resultante (*Figura (31)*) dio un índice de estimación de error (OBB) de 0.27%, y un índice kappa de 0.9953472, obteniendo así un mínimo de error aceptable.

```

call:
  randomForest(x = x, y = y, mtry = min(param$mtry, ncol(x)))
  Type of random forest: classification
  Number of trees: 500
  No. of variables tried at each split: 13

  OOB estimate of error rate: 0.27%
Confusion matrix:
  1  2  3  4  5  6  7  class.error
1 4130  4  0  0  1  0  0 0.0012091898
2  5 1740  2  1  2  2  1 0.0074158585
3  0  1 235  0  0  3  0 0.0167364017
4  1  3  0 430  0  3  0 0.0160183066
5  0  0  0  0 923  0  0 0.0000000000
6  0  3  0  0  0 4756  0 0.0006303845
7  0  1  0  0  0  0 95 0.0104166667
  
```

Figura (31) Matriz de confusión de la clasificación resultante en el Valle del Yaqui. Autoría propia

Como se puede observar, las clases que obtuvieron un mayor rango de error debido a sus características similares, fueron las número 1, 2 y 6, las cuales corresponden a las clases de arena, matorrales y cultivos respectivamente, a su vez la similitud de los sedimentos que conforman la arena y cultivos en descanso también crean confusión al algoritmo. Sin embargo, el número de falsos positivos nuevamente es bajo para la cantidad de muestras que contiene la matriz, por lo que la clasificación es satisfactoria dentro de los parámetros establecidos.

IV.4 Proceso de validación y delimitación de la frontera agrícola

Para la delimitación de la frontera agrícola de cada región fue necesario un proceso de validación y corrección de los rasters de clasificación resultantes, esto se llevó a cabo repitiendo el proceso de colocación o eliminación de polígonos basándose en el número de falsos positivos y negativos que fueron señalados en la matriz de confusión resultado de la clasificación con *random forest*.

En dicho algoritmo se observó que las clases que contenían mayor número de falsos positivos y negativos fueron las clases de cerro y cultivo en el caso de los municipios de Guerrero, y arena y cultivos en el caso del Valle del Yaqui zonas urbanas, sin embargo, esta última clase no se tomó en cuenta para la corrección debido a que la investigación se centra en la delimitación de la zona cultivable.

Para la región de los municipios de Guerrero se analizaron los polígonos de estas clases y se removieron o agregaron aproximadamente el 10% de los polígonos que no coincidían con la clase que indicaba la clasificación en el área, por otro lado, en la región del Valle del Yaqui solo fue necesario agregar o remover un 5% del total de polígonos.

Posteriormente, se separó la clase de cultivo del resto de clases y después de realizar técnicas de suavizado y eliminación de áreas excedentes (*Figuras (32 A) y (32 B)*), se realizó de nuevo el procedimiento de clasificación con *random forest*, obteniendo así una nueva matriz para cada región, esta vez con las nuevas modificaciones de los polígonos de entrenamiento.



Figura (32) Imagen en color verdadero

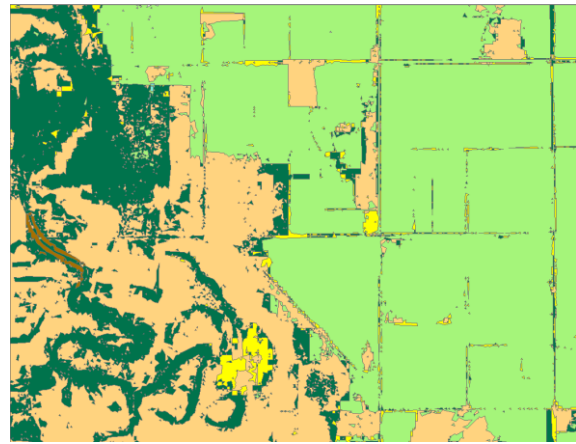


Figura (32 A)

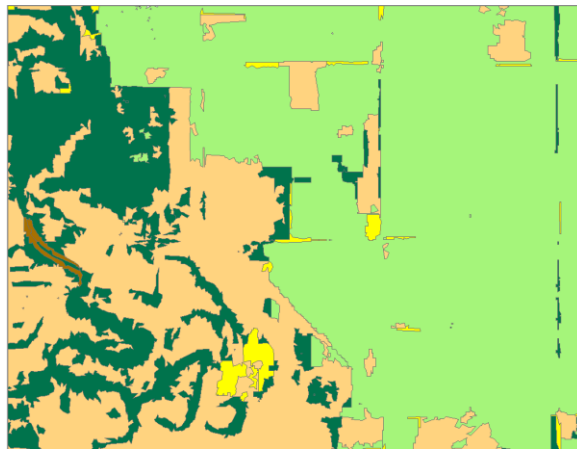


Figura (32 B)

Cercanía de una sección del mapa de clasificación en el Valle del Yaqui antes (Figura (32 A)) y después (Figura (32 B)) del proceso de suavizado y eliminación de polígonos excedentes. Autoría propia

Segunda matriz de confusión para el Valle del Yaqui.

La segunda matriz de confusión resultante (Figura (33)) para esta región dio un índice de estimación de error (OBB) de 0.3%, y un índice kappa de 0.9929845.

```

Call:
  randomForest(x = x, y = y, mtry = min(param$mtry, ncol(x)))
    Type of random forest: classification
    Number of trees: 500
No. of variables tried at each split: 13

      OOB estimate of  error rate: 0.3%
Confusion matrix:
      1  2  3  4  5  6  7  class.error
1 4490  5  0  0  1  1  0 0.0015565933
2  4 1768  3  5  1  1  0 0.0078563412
3  0  1 229  0  0  3  1 0.0213675214
4  2  1  0 491  0  3  0 0.0120724346
5  0  2  0  0 943  0  0 0.0021164021
6  0  3  0  0  0 4803 0 0.0006242197
7  0  0  1  0  0  0 99 0.0100000000
> print("Fin programado")
[1] "Fin programado"
> |

```

Figura (33) Segunda matriz de confusión de la clasificación resultante en el Valle del Yaqui. Autoría propia

Segunda matriz de confusión para los municipios de Xochistlahuaca y Tlacoachistlahuaca.

La segunda matriz de confusión resultante (Figura (34)) para esta región dio un índice de estimación de error (OBB) de 1.65%, y un índice kappa de 0.9370277.

```

Call:
  randomForest(x = x, y = y, mtry = min(param$mtry, ncol(x)))
    Type of random forest: classification
    Number of trees: 500
No. of variables tried at each split: 11

      OOB estimate of  error rate: 1.65%
Confusion matrix:
      1 18 19 20  23 26  29  class.error
1 1812  0  0  2  0  0  0 0.001102536
18  5 26  1  0  0  0  0 0.187500000
19  4  1 50  4  1  0  0 0.166666667
20  2  0  5 58  0  0  0 0.107692308
23  6  0  0  0 226  0  2 0.034188034
26  1  0  0  0  0  5  0 0.166666667
29  1  0  3  0  1  0 147 0.032894737
> print("Fin programado")
[1] "Fin programado"

```

Figura (34) Segunda matriz de confusión de la clasificación resultante en los municipios de Xochistlahuaca y Tlacoachistlahuaca. Autoría propia

Los nuevos mapas de clasificación (Ver figuras (35) y (36)) dieron como resultado un área de 5796.84 ha. (16% de la cobertura aprox.) para la región de los municipios de Tlacoachistlahuaca y Xochistlahuaca, y un área de 13463.24 ha. (7.4% de la cobertura aprox.)

para la región del Valle del Yaqui que no correspondía en realidad con áreas de cultivos (Ver figuras (37) y (38)) por lo que fueron removidas con las mismas técnicas de descarte que en los anteriores mapas de clasificación.

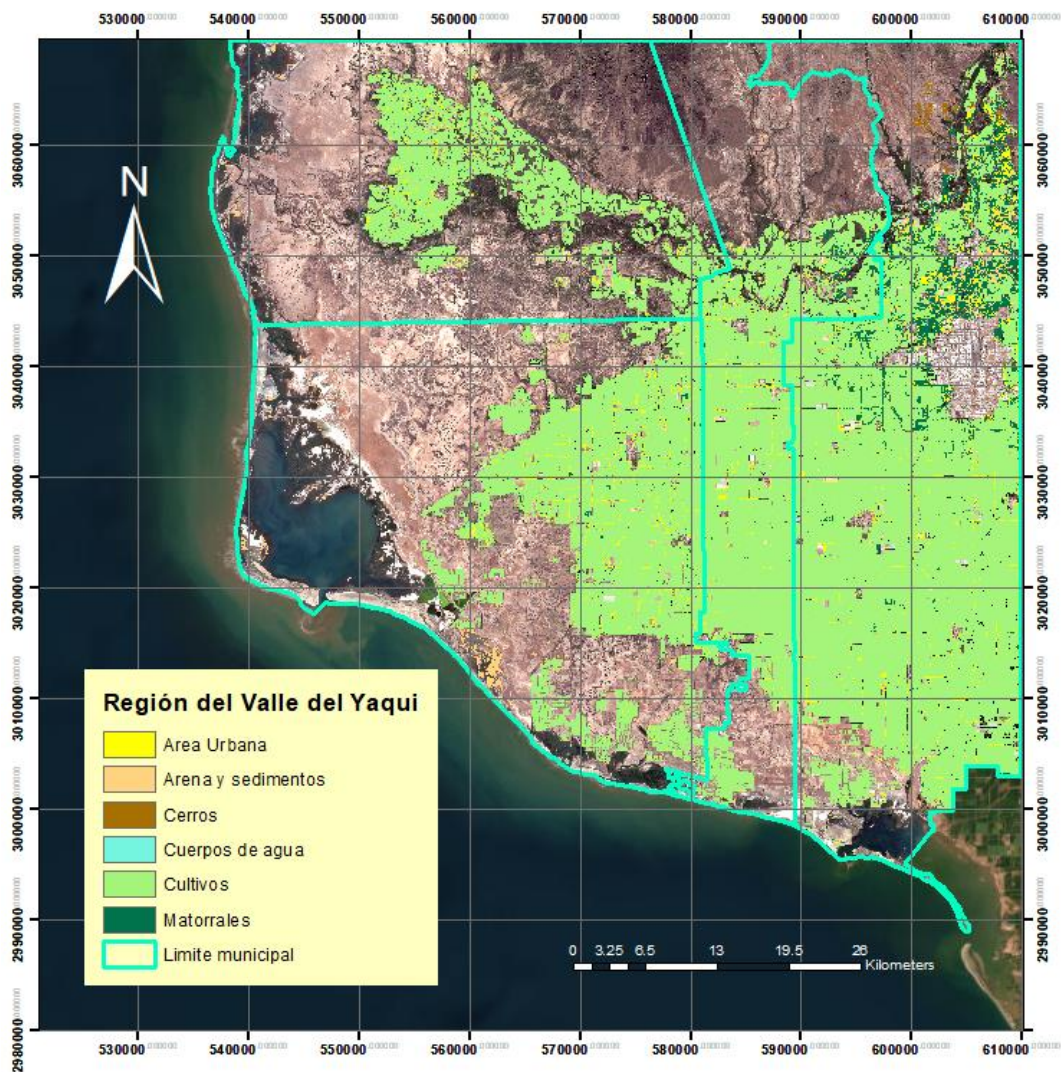


Figura (35) Segundo mapa fisiográfico resultante de la clasificación de random forest para la región del Valle del Yaqui después del proceso de validación. Autoría propia

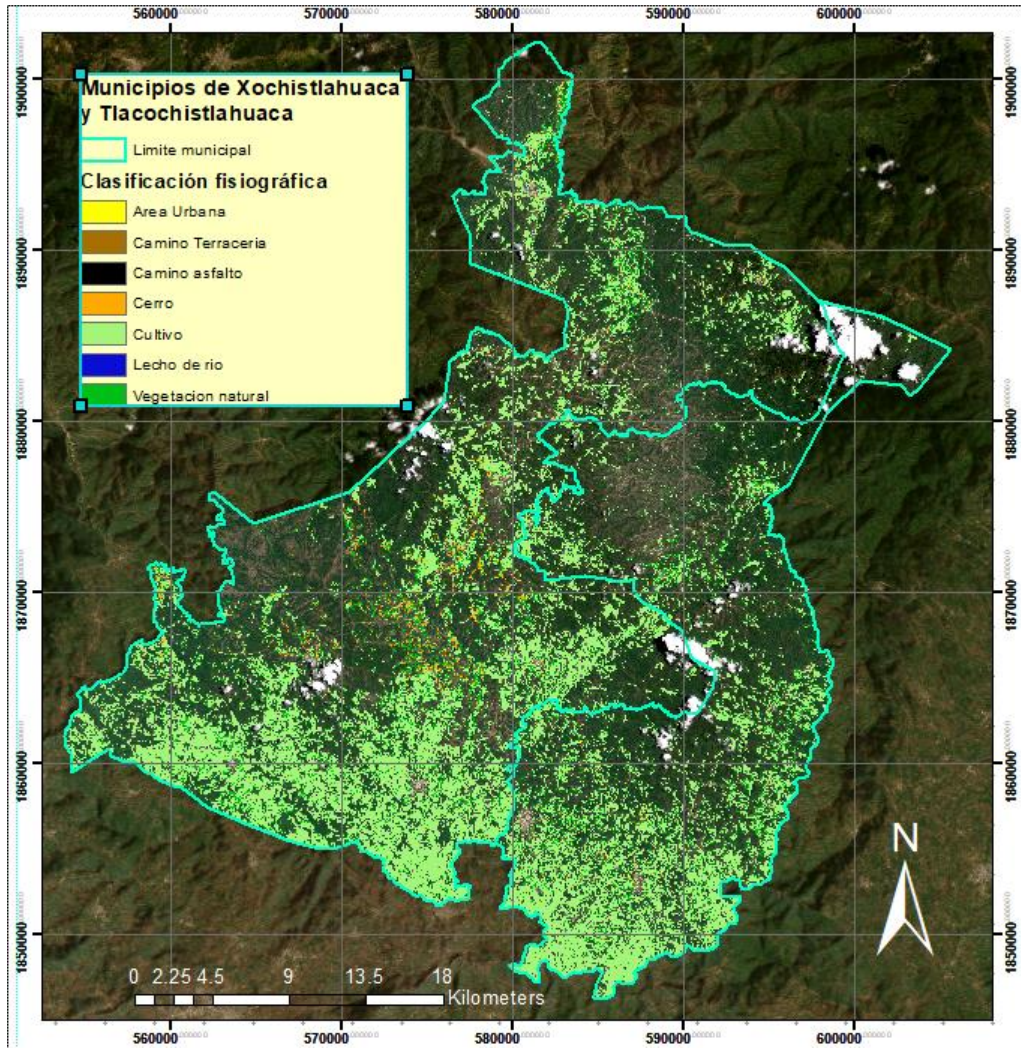


Figura (36) Segundo mapa fisiográfico resultante de la clasificación de random forest para los municipios de Xochistlahuaca y Tlacoachistlahuaca después del proceso de validación. Autoría propia

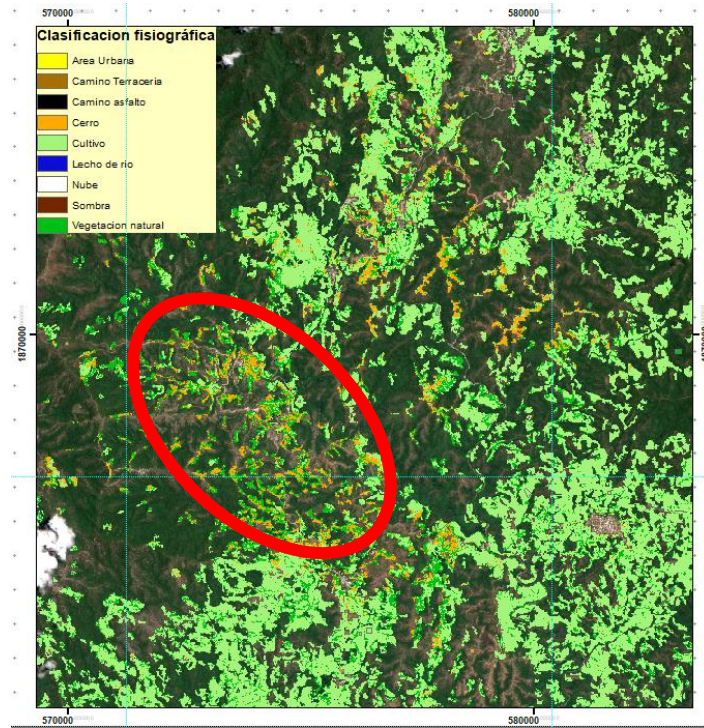


Figura (37) Municipio de Tlacoachistlahuaca, Guerrero. El círculo en rojo marca una gran zona de reclasificación del cultivo al realizar el procedimiento con un mejor análisis pasando a ser clasificados como cerros y vegetación natural. Autoría propia

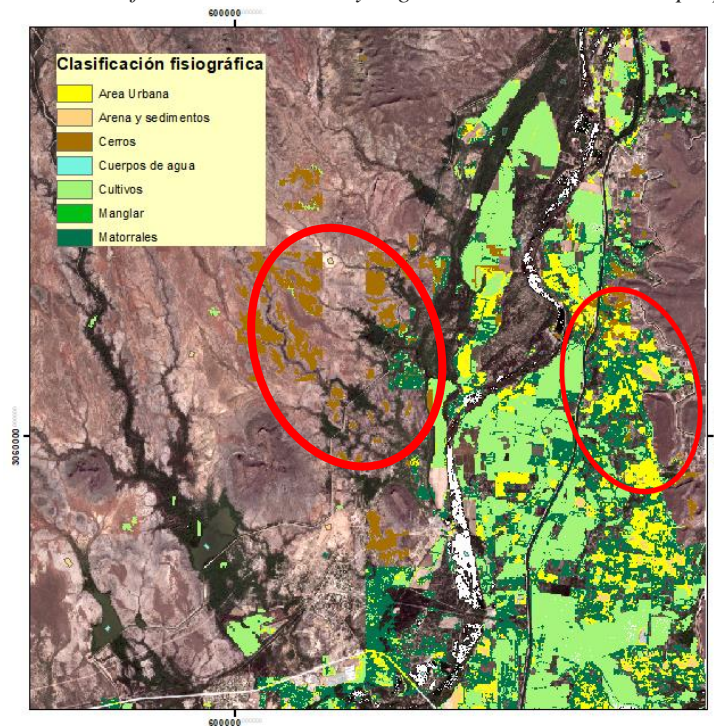


Figura (38) Municipio de Cajeme, Sonora. Los círculos en rojo muestran algunas zonas de reclasificación del cultivo al realizar el procedimiento con un mejor análisis, pasando a clasificarse como cerros y zonas urbanas. Autoría propia

Una vez señaladas las zonas que no pertenecen a la clase cultivo, se procedió a eliminarlas y realizar de nuevo el proceso de suavizado y eliminación de polígonos excedentes. Cabe destacar que se eliminaron polígonos que están entre 0.5 a 1.4 hectáreas ya que, los cultivos menores a estas medidas no se consideran importantes para la validación debido a que normalmente se consideran pequeñas huertas, terrenos o jardines.

Una vez realizadas todas las correcciones, en los municipios de Guerrero (*Ver figura (39)*), el resultado de **la frontera agrícola final mostró un 25.23% (31521.98 ha.)** de la cobertura total (124922.2276 ha.), esto contando la corrección final que removió un 4.4% del 29.63% de la clase cultivos original que no correspondía a dicha clase.

Por otro lado, en la región de Valle del Yaqui (*Ver figura (40)*), **la frontera agrícola final mostró un 38.68% (163674.93 ha.)** de la cobertura total (423085.3107 ha.) resultado de la corrección final que removió un 4.84% del 43.52% de la clase cultivo original.

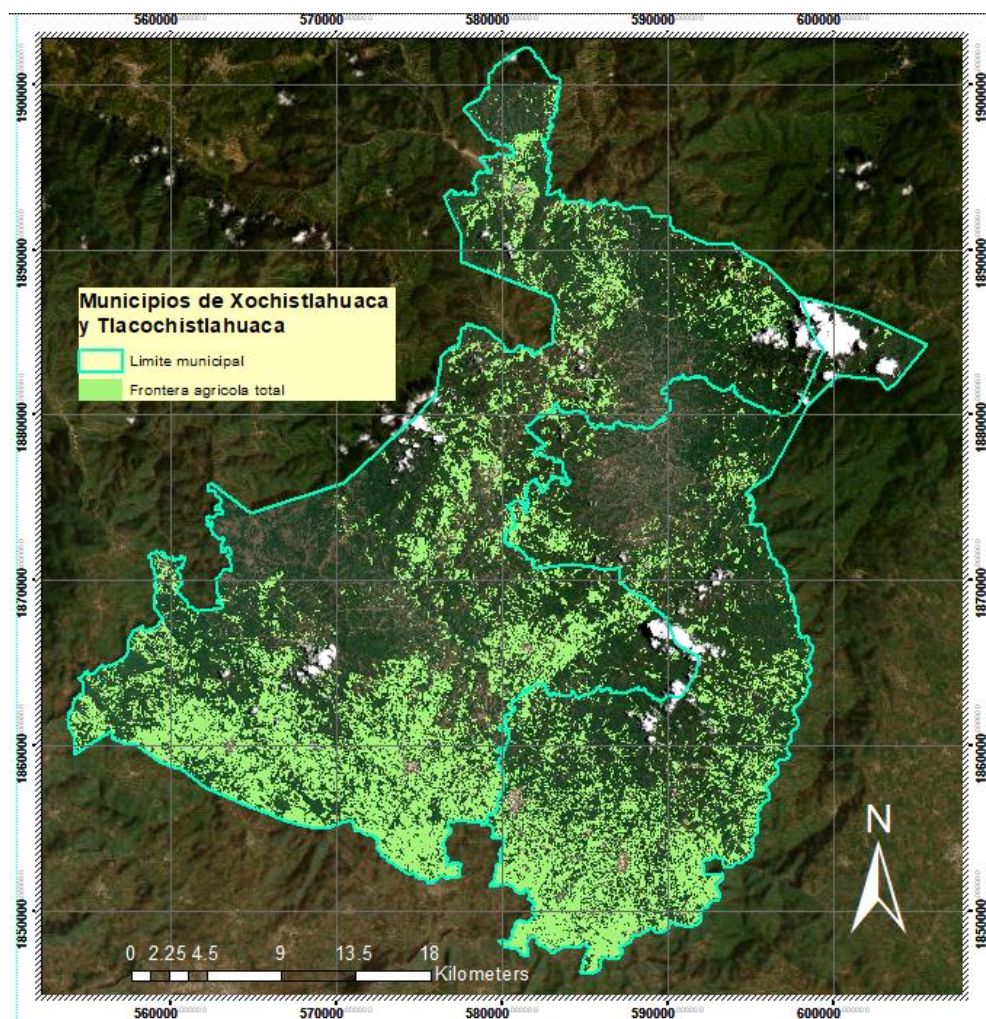


Figura (39) Frontera agrícola final (31521.986 ha.) de los municipios de Xochistlahuaca y Tlacoachistlahuaca. Autoría propia

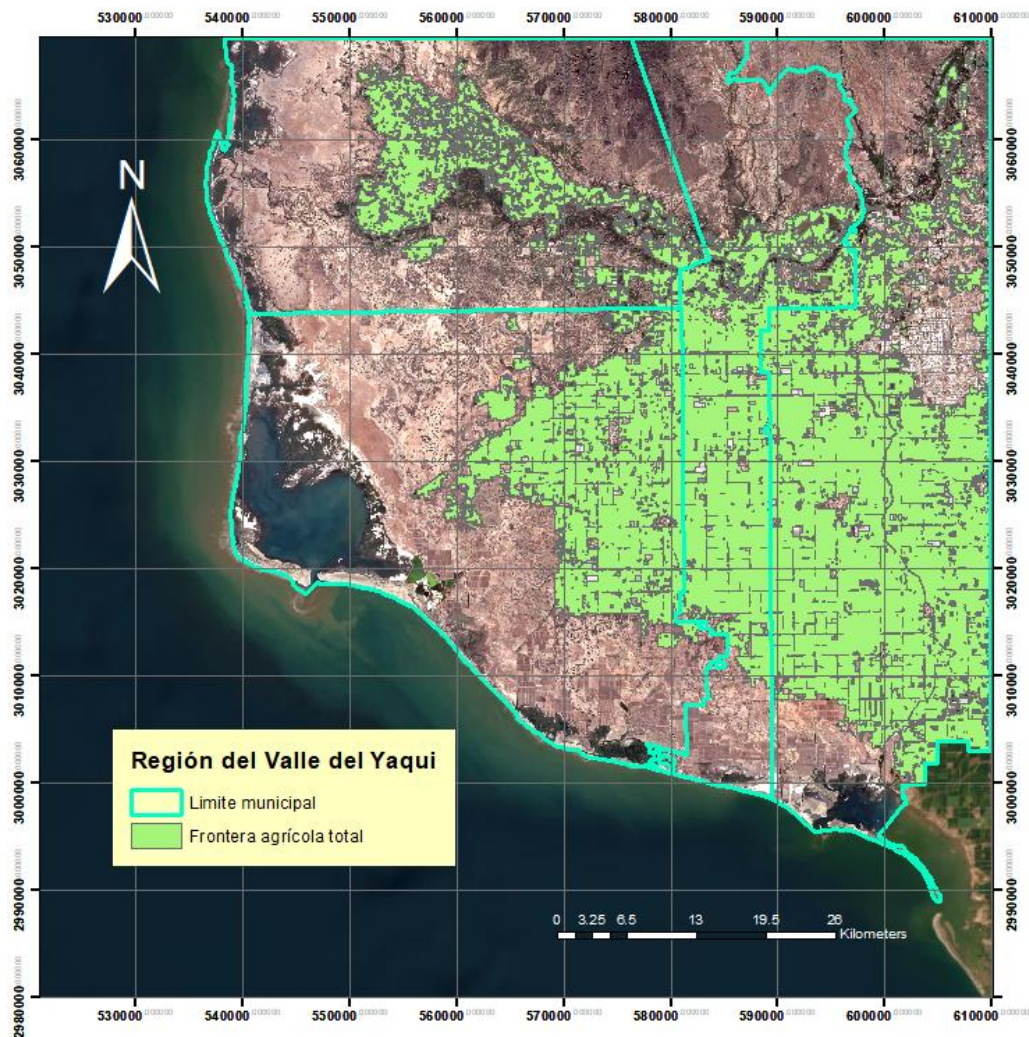


Figura (40) Frontera agrícola final (163674.93 ha.) del Valle del Yaqui. Autoría propia

Capítulo V. Discusión

Esta investigación tuvo como objetivo principal el establecer una metodología precisa sobre la delimitación de la frontera agrícola mexicana, pero a la vez, que no conllevara a un gasto excesivo de recursos y tiempo, haciéndola lo más práctica posible para un analista. Así mismo, los recursos utilizados (imágenes satelitales, datos estadísticos, entre otros) fueron seleccionados cuidadosamente para que la metodología contara con datos actualizados y precisos. El uso de drones, estudios de suelo y relieve *in situ* e imágenes satelitales con sensores de mejor resolución espacial, mejoraría las características revisadas en la investigación, sin embargo, se debe valorar el costo beneficio de dichos dispositivos además de la extensión territorial de los sitios de estudio, que en este caso cubrió más de un municipio en las regiones analizadas.

En relación con la metodología utilizada, se consideró que los índices de vegetación cumplieron un rol importante en la detección de cubierta vegetal, teniendo una mayor importancia para las regiones del sur y sureste del país, ya que cuentan con gran abundancia de vegetación natural y zonas cultivables, por lo que dichos índices ocuparon pesos más importantes para la evaluación multicriterio que la variable de pendiente. Esto se debe a que los habitantes de estas zonas suelen ocupar también terrenos inclinados para sembrar y, por lo tanto, tienden a exceder los límites de grados de pendiente establecidos para un correcto uso de suelo cultivable, por lo que en estos casos la pendiente no llega a ser una variable tan importante para los algoritmos de clasificación a diferencia de otras características del terreno. Aunque el cultivar en un terreno inclinado puede favorecer al agricultor para obtener un mayor aprovechamiento de área, estas prácticas conllevan riesgos tales como deslaves por lluvias y pérdida de nutrientes para los cultivos. Por ello mismo el uso de terrazas agrícolas por parte de los habitantes suele evitar estas situaciones.

Por otra parte, la distribución de los polígonos de entrenamiento tuvo una gran importancia en el éxito de las clasificaciones debido a que existía el problema de determinar límites e intervalos de los valores de los píxeles que no obtuvieron una clase definida, lo cual lleva a los algoritmos de clasificación a cometer errores.

Algunas de las clases que obtuvieron errores debido a que contenían reflectancias similares fueron cultivos en crecimiento con vegetación natural, por las condiciones fotosintéticas de sus hojas, por lo cual los polígonos de entrenamiento asignados a la vegetación natural tenían que colocarse en áreas con gran cubierta vegetal que creara una sombras visibles de las propias plantas, ya que esto le daría una característica más distintiva a diferencia de las plantas cultivadas, las cuales son más dispersas y forman patrones regulares creados por los propios agricultores.

Por otra parte, otras clases que obtuvieron algunos errores en la clasificación fueron las reflectancias de arena y sedimentos con las de cultivos en descanso, ya que usualmente los cultivos no utilizados se dejan abandonados durante largos periodos de tiempo, por lo cual la arena y polvo suelen cubrir estas áreas haciendo difíciles de diferenciar en una imagen satelital. Debido a esto, se optó por la creación de varios polígonos de entrenamiento en la

frontera entre estas 2 clases (Arena y cultivo) ya que el algoritmo de *random forest* utiliza modelos de regresión lineal y votación por clase, creando un efecto *nearest neighbour* (vecinos cercanos) evitando lo más posible que pixeles cercanos con valores similares se clasifiquen en una misma clase y se evite una confusión lo menos posible. A su vez se llegó a la conclusión que colocar polígonos de entrenamiento en lugares idóneos y bien definidos ayuda al algoritmo a tener una clasificación más heterogénea sobre todo en clases que tienen reflectancias similares.

Por último, durante la clasificación fisiográfica final, se notaron algunas inconsistencias ajenas al algoritmo *random forest*. La más importante se detectó en la parte sur de la región del Valle del Yaqui donde el algoritmo confundió granjas de camarones con campos de cultivo sin cosechar, esto debido a que, en la fecha de las imágenes satelitales utilizadas, estos campos estaban carentes de agua y por lo tanto tomaron las características de un campo de cultivo en descanso. Por casos como el anterior, es importante realizar un análisis del terreno previo y posterior a la creación de la clasificación para así detectar características del sitio que puedan afectar al proceso. A su vez, si se dispone de más tiempo para dicho proceso, se recomienda realizar un *timelapse* del sitio en cuestión para detectar cambios importantes en las características del terreno en distintos periodos agrícolas.

V.1 Conclusiones

En México aproximadamente el 16% del territorio está conformado por terreno agrícola, mantener el número aproximado de hectáreas que se ocupan para tal actividad es fundamental para entender y administrar el territorio nacional para diversos fines, es por ello que la frontera agrícola es un concepto que proporciona información geoespacial, medible y calculable sobre todos los campos de cultivo sembrados y utilizables en los últimos años.

Conocer el límite de esta frontera es útil para diversos campos de estudio, que se encargan de administrar y analizar el territorio nacional, principalmente en estudios ambientales, sociales y de desarrollo sostenible. Problemas como la tala de árboles para la creación de campos de cultivos y la invasión a áreas naturales protegidas pueden ser mejor analizados desde una perspectiva geográfica, por tal motivo es importante conocer en dónde concentrar los esfuerzos para controlar la invasión a estas áreas escasas en el país. Al mismo tiempo, en el ámbito social y económico se pueden saber las pérdidas y ganancias de las producciones agrícolas comparando los datos estadísticos de las áreas cultivadas anual o mensualmente, dando a conocer qué tipo de alimentos se producen, en qué cantidad y cuantas hectáreas ocupan cada uno. La demanda de vivienda urbana también suele ser un problema ya que se invaden o se compran terrenos agrícolas abandonados o en uso que reducen la producción en pueblos y zonas rurales empobrecidas, lo que conlleva a la invasión de más áreas naturales para la creación de nuevos campos de cultivo.

También, gracias a la observación y análisis geoespacial se pueden detectar diversos problemas que afectan a los cultivos, tales como; devastación por factores climáticos, plagas e incendios forestales, los cuales pueden analizarse desde gabinete sin necesidad de gastar

recursos y personal para cada rincón del país donde existan estos problemas en lugares y tiempos indeterminados.

Es por ello que, metodologías de monitoreo agrícola y forestal como la que se llevó a cabo en esta investigación buscan mejorar y actualizar los procesos que llevan a la creación de la delimitación de la frontera agrícola, ya que año con año esta misma siempre está en un constante cambio, por lo que métodos de análisis más simples, no siempre son capaces de calcular con exactitud las zonas con grandes extensiones, y a su vez, métodos más precisos, suelen conllevar más tiempo y no son capaces de abarcar grandes territorios.

La alimentación autosustentable para la población a través de los campos agrícolas es un reto que busca mejorar la calidad de vida y la economía de los productores, es por ello que mientras mejor se conozca y se analice el territorio cultivable y lo que lo rodea, mejor conoceremos los problemas a enfrentar y las posibles soluciones que beneficien a los campos de nuestro país.

Anexo 1

Script de *random forest* utilizado para esta investigación ejecutado en “R studio”

```
1 #Ejercicio randomforest
2
3 # Para ejecutar
4 # Rscript --vanilla Clas_RandomForest.r {dir:trabajo} {stak.TIF}
5 {Shape.shp} {nombre:archivo:salida.TIF}
6
7 args <- commandArgs(TRUE)
8
9 directorio_de_trabajo = "D:/Tesis/Rasters definitivos tesis" #
10 args[1]
11 pila_datos = "pila_D.tif" #args[2]
12 poligonos_entrenamiento = "AE_D.shp" #args[3]
13 archivo_de_salida = "RF_Tesis_1.tif" # args[4]
14
15 # directorio_de_trabajo = args[1]
16 #pila_datos = args[2]
17 # poligonos_entrenamiento = args[3]
18 # archivo_de_salida = args[4]
19
20 library(sp)
21 library(raster)
22 library(rgdal)
23 library(randomForest)
24 library(lattice)
25 library(ggplot2)
26 library(caret)
27 library(snow)
28 library(e1071)
29
30 setwd(directorio_de_trabajo)
31 getwd()
32
33 # readGDAL
34 # brick
35
36 img <- brick(pila_datos)
37
38 print("Se cargó la pila")
39
40 #plot(img, main = "img")
41
42 trainData <- shapefile(poligonos_entrenamiento)
43 print("AE_D.shp")
44
45 #plot(img, main = "img")
46
47 names(img) <- c(paste0("B", 1:(nlayers(img))), coll = "")
```

```

48 print("Se cambiaron nombres de bandas")
49
50 img2 <- img$B1 * 0
51 print("Se inicializo banda en ceros")
52
53 rp <- rasterize(trainData, img2, 'Class')
54 names(rp) <- "Class"
55 print("Se rasterizaron clases")
56
57 # plot(rp)
58 img3 <- addLayer(img, rp)
59 print("Se agregaron clases rasterizadas a la pila")
60
61 print("Extrayendo valores del raster, espere...")
62 dataSet3 <- extract(img3, trainData)
63 print("Valores almacenados")
64
65 dfAll = data.frame(matrix(vector(), 0, length(names(img3)) ))
66 print("Se inicializa data frame para recibir los valores del raster")
67
68 names(dfAll) <- names(img3)
69
70 print("Se renombran las columnas del dataframe")
71
72 df <- do.call("rbind", dataSet3)
73 print("Se ordenan en columnas los vsalores del raster, una columna x
74 banda del raster")
75
76 dfAll <- rbind(dfAll, df)
77 print("Guardando valores en data frame")
78
79 dfAll5 <- na.omit(dfAll)
80 # na.omit(dfAll5)
81 print("Eliminando nullos")
82
83 ejemplos <- ( (length(row(dfAll5))/length(names(img3)) * 50) %/% 100)
84
85 nsamples <- ejemplos
86 #nsamples <- 400
87 sdfAll <- subset(dfAll5[sample(1:nrow(dfAll5), nsamples), ])
88 print("Seleccionando datos para ajustar el modelo RF")
89 print("Inicia ajuste de modelo Random Forest")
90 date()
91 modFit_rf <- train(as.factor(Class) ~ ., method = "rf", data = sdfAll
92 )
93 date()
94 print("Inicia la clasificaci?n")
95 date()
96 beginCluster()
97 preds_rf <- clusterR(img, raster::predict, args = list(model =
98 modFit_rf))
99 endCluster()

```

```
100 date()
101 print("Termino la clasificaci?n ")
102 #plot(preds_rf, main = "Clasificada")
103
104 writeRaster(preds_rf, filename = archivo_de_salida, datatype
105 ='INT2S', format = "GTiff", overwrite = TRUE)
106 modFit_rf
107
    modFit_rf$finalModel
    print("Fin programado")
```

Bibliografías

- [1] Andrés Fernando Jiménez López et al., Análisis multiespectral de la vegetación para aplicaciones de sensado remoto, *Iteckne* vol.12 no.2, 2015. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-17982015000200007
- [2] Noé Aguilar Rivera, Percepción remota como herramienta de competitividad de la agricultura, *Revista Mexicana de Ciencias Agrícolas* Vol.6 Núm.2 [online]. 2015. <http://www.scielo.org.mx/pdf/remexca/v6n2/v6n2a14>
- [3] R. Allende Lastra, *Revista de estadística y geografía*, Volumen 1, No.2, secretaria de programación y presupuesto (SPP), 1980, Pág. 37.
- [4] Utilidad de frontera agrícola en la Delegación de la SAGARPA en el estado de San Luis Potosí, Servicio de Información Agroalimentaria y Pesquera (SIAP). https://www.gob.mx/cms/uploads/attachment/file/313900/Utilidad_de_la_Frontera_Agr_co_la_San_Luis_Potos_.pdf
- [5] Erika Danaé López Espinoza, M. en C., INAOE, “Segmentación de Coberturas de la Tierra Espectralmente Similares Empleando Campos Aleatorios de Markov y Características de Textura Estructural y Estocástica”, Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla, 2009
- [6] Juan Pablo Ardila Lopez et al. Validación de una Metodología de Clasificación de Imágenes Satelitales en un entorno Orientado a Objetos, 2005
- [7] E. Muñoz, A. Zozaya, and E. Lindquist, Satellite Remote Sensing of Forest Degradation, Using NDFI and the BFAST Algorithm, *IEEE Latin America Transactions*, VOL. 18, NO. 7, 2020. Pag: 1288-1295.
- [8] Nicolás del Toro Espín, et al., Comparación de métodos de clasificación de imágenes de satélite en la cuenca del río Argos, *Boletín de la Asociación de Geógrafos Españoles* N.º 67, 2015, Pág. 327-347.
- [9] Siamak Khorram et al. Remote Sensing, *SpringerBriefs in Space Development*, DOI: 10.1007/978-1-4614-3103-9, 2012, Pag 2.
- [10] Francisco Bautista et al. Técnicas de muestreo para manejadores de recursos naturales, Centro de investigaciones en Geografía ambiental, Universidad nacional autónoma de México 2004, Pag: 457 y 461.
- [11] Daniel J. Pérez, Introducción a los Sensores Remotos - Aplicaciones en Geología, UBA 2007, Pag. 5 y 14. http://aviris.gl.fcen.uba.ar/Curso_SR/Guia_Curso_sr_2007.pdf.
- [12] M. en G. Daniel M. López López, M. en G. Aura C. Torres Gómez, Introducción a los datos tipo raster, Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C, Pag: 17-20.
- [13] María Valentina Soria, María Angélica Matar de Saquis, Nociones sobre teledetección, Laboratorio de Procesamiento de Imágenes y SIG Instituto de Investigaciones Mineras

Facultad de Ingeniería, 2016, Pag 4.

<http://www.unsj.edu.ar/unsjVirtual/cartografiaaplicadaminas/wp-content/uploads/2016/10/Apuntes-de-c%C3%A1tedra-para-Cartograf%C3%ADa-Aplicada.pdf>

[14] Esperanza Sánchez Rodríguez, Sistemas de Información Geográfica: datos RASTER., Universidad de Murcia, 2020, pag 2.

[15] Emilio Chuvieco, Fundamentos de teledetección espacial, Segunda edición, Ediciones Rialp S.A. Madrid, 1995.

[16] Diego Hernán Beltrán Hernández, Aplicación de índices de vegetación para evaluar procesos de restauración ecológica en el parque forestal embalse del Neusa, , Universidad Militar Nueva Granada, facultad de ingeniería, Bogotá D.C., 2017, Pag 9. <https://repository.unimilitar.edu.co/handle/10654/17210>

[17] Benavides Mendoza Adalberto, Ramírez Homero, Respuestas de las plantas a la radiación electromagnética TY, Universidad Autónoma Agraria Antonio Narro (UAAAN), 2002.

[18] Carmelo Alonso et al., Determinación experimental de la firma espectral de la vegetación. Una sencilla práctica de introducción a la teledetección, Departamento de Teledetección, INDRA Espacio. C/ Mar Egeo, nº 4. 28850 San Fernando de Henares, Madrid, 1999.

[19] Pedro Muñoz Aguayo, Apuntes de teledetección: índices de vegetación, Centro de Información de Recursos Naturales, 2013. Pag 2.

[20] Servicio de Información Agroalimentaria y Pesquera (SIAP), Monitoreo de Cambios en Áreas con Actividad Agroalimentaria y Pesquera Utilizando Imágenes de Satélite de la Constelación SPOT, 2018. <https://www.gob.mx/siap/documentos/cambios-en-areas-con-actividad-agroalimentaria-y-pesquera-utilizando-imagenes-de-satelite-108352>

[21] Julio C Rodriguez, et al., Estimación de la evapotranspiración de un viñedo de uva de mesa (vitis vinífera) con imágenes satelitales sentinel, Universidad de Sonora, agosto 2021.

[22] Hernández, I. U., Ellis, E. A., Gallo, C. A. “Aplicación de teledetección y sistemas de información geográfica para el análisis de deforestación y deterioro de selvas tropicales en la región Uxpanapa, Veracruz”, 2013

[23] Anónimo, Enciclopedia de los Municipios y Delegaciones de México, Instituto Nacional para el Federalismo y el Desarrollo Municipal, 2017 <https://www.derechomunicipal.org.mx/single-post/2018/07/24/enciclopedia-de-los-municipios-y-delegaciones-de-m%C3%A9xico>

[24] Precipitación máxima anual de las acumuladas mensuales (1902-2011), Repositorio Institucional, Instituto de Ciencias de la Atmósfera y Cambio Climático, UNAM, 2011. <https://ri.atmosfera.unam.mx/acdm/srv/spa/catalog.search#/metadata/019967a7-00d0-497c-beaa-ebad866b9209>

- [25] Temperatura media mensual (1902-2015), Repositorio Institucional, Instituto de Ciencias de la Atmósfera y Cambio Climático, UNAM, 2015. <https://ri.atmosfera.unam.mx/acdm/srv/spa/catalog.search#/metadata/a587ea57-e07d-43a2-9ad4-b560483bae5f>
- [26] Continuo de Elevaciones Mexicano (CEM), INEGI, (s.f.), <https://www.inegi.org.mx/app/geo2/elevacionesmex/>
- [27] Geografía y Medio Ambiente, Climatología, INEGI, (s.f.), <https://www.inegi.org.mx/temas/climatologia/#Descargas>
- [28] Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), Suelos, Suelos de México (s.f.), 2012 Pag 5. https://apps1.semarnat.gob.mx:8443/dgeia/informe_12/pdf/Cap3_suelos.pdf
- [29] Arturo García Navarro, Edafología, Universidad de Extremadura, Departamento de Biología y Producción de los Vegetales, Área de Edafología y Química Agrícola, 2005. <https://www.eweb.unex.es/edafo/FAO>
- [30] Uso de suelo y vegetación, Geografía y Medio Ambiente, INEGI. <https://www.inegi.org.mx/temas/ususuelo/>
- [31] Dr. Amin Zarur Menes et al., Enciclopedia de los Municipios de México, Estado de Guerrero, Xochistlahuaca, Centro Nacional de Desarrollo Municipal, Gobierno del Estado de Guerrero, 2021 <https://www.mexicantextiles.com/library/amusgo/xochistlahuacainfo.pdf>
- [32] Tomás Bustamante Arroyo, Agricultura y perspectivas de desarrollo, SIPIG-UNAM, 2009, Pag 4. https://www.nacionmulticultural.unam.mx/edespig/diagnostico_y_perspectivas/RECUADROS/CAPITULO%206/1%20AGRICULTURA%20Y%20PERSPECTIVAS.pdf
- [33] Valle del Yaqui: el granero de México, MUNSA, 2015. <https://www.munsa.com.mx/blog/valle-del-yaqui-el-granero-de-mexico#:~:text=M%C3%A1s%20de%20450%2C000%20hect%C3%A1reas%20ba%C3%Bladas,en%20el%20granero%20de%20M%C3%A9xico.>
- [34] Secretaría de Agricultura y Desarrollo Rural, (SADER), 2020. <https://www.agricultura.gob.mx/sites/default/files/sagarpa/document/2020/03/19/1896/19032020-guerrero-pcef-informe-eval-2015-2017-08112018.pdf> página 8
- [35] Anuario Estadístico de la Producción Agrícola, Servicio de Información Agroalimentaria y Pesquera (SIAP). <https://nube.siap.gob.mx/cierreagricola/>
- [36] Simon Fonteyne y Nele Verhulst, Rotación de cultivos: beneficios directos a la economía de los productores, CIMMYT, 2019. <https://idp.cimmyt.org/rotacion-de-cultivos-beneficios-directos-a-la-economia-de-los-productores/>
- [37] Alma Trasviña Barriga et al., Rehabilitación de un suelo salino con yeso agrícola en un cultivo de nogal en el Valle del Yaqui, Terra Latinoamericana vol.36 no.1 Chapingo, 2018,

Pág. 1. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57792018000100085

[38] Richardson, AJ y CL Wiegand, "Distinguir la vegetación de la información de fondo del suelo", *Photogrametric Engineering and Remote Sensing*, 1977, Pág. 1541–1552. https://www.asprs.org/wp-content/uploads/pers/1977journal/dec/1977_dec_1541-1552.pdf

[39] Taghreed A H Naji, Study of vegetation cover distribution using DVI, PVI, WdVI indices with 2D-space plot, 2018. <https://iopscience.iop.org/article/10.1088/1742-6596/1003/1/012083/pdf>

[40] Emanuele Alcaras, et al., Normalized Burn Ratio Plus (NBR+): A New Index for Sentinel-2 Imagery, 2022, Pág. 2.

[41] Florina Cristina Roşca, Climatic dysfunctionalities observed with the aid of ndmi and savi indices in the leu-rotunda and dăbuleni plains, 2013, Pág. 505 y 500 <https://www.proquest.com/openview/afb6d7d408ff41ec23bdd8bd3d0f5b00/1?pq-origsite=gscholar&cbl=2026591>

[42] Castellanos, G. y Moreno, M. y F. D. Maldonado, Obtención de imágenes índice de vegetación SAVI a partir de imágenes milirreflectancia TOA, CEREGEO. FCyT. Universidad Autónoma de Entre Ríos (UADER), 2016.

[43] NDMI: Índice De Humedad De Diferencia Normalizada, EOS Data Analytics. <https://eos.com/es/make-an-analysis/ndmi/>

[44] B. Leo, Random forests, Kluwer Academic Publishers, vol. 45, 2001, Pág. 5-32.

[45] R. Prytz, S. Nowaczyk, T. Rögnvaldsson, and S. Byttner, Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data, *Engineering Applications of Artificial Intelligence*, vol. 41, 2015, Pág. 139-150.

[46] G. Biau and E. Scornet, A random forest guided tour, *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, vol. 25, 2016, Pág. 197–227.

[47] Javier Redondo Antón, et al., Comparativa de modelos de random forest y redes neuronales aplicados al mantenimiento predictivo con valores ausentes y datos desbalanceados, Universidad Complutense, Madrid, 2021, Pág. 29. https://eprints.ucm.es/id/eprint/68457/1/Javier_Redondo_TFM.pdf

[48] V. Mathew, T. Toby, V. Singh, B. M. Rao, and M. G. Kumar, Prediction of remaining useful lifetime (rul) of turbofan engine using machine learning, *International conference on circuits and systems (ICCS)*, 2017, Pág. 306–311.

[49] Yébenes Gómez, Á., & Giner Sotos, Á., *Clasificación Digital De Imágenes Por Satélite*, 2010.

- [50] Understand Random Forest Algorithms With Examples (Updated 2023), Analytics Vidhya, 2023. <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- [51] Rogelio Hernández Montes, Cynthia Alejandra Martínez Pinto, Silvana Guadalupe Navarro Jiménez, Redes neuronales y árboles de decisión para la clasificación de objetos astronómicos, Tecnológico Nacional de México, ITCG, Jalisco, México, 2019. https://www.rcs.cic.ipn.mx/rcs/2019_148_7/Redes%20neuronales%20y%20arboles%20de%20decision%20para%20la%20clasificacion%20de%20objetos%20astronomicos.pdf
- [52] Bense, T., Introducción a la percepción remota. VI Jornadas de Educación En Percepción Remota En El Ámbito Del Mercosur y I Uruguayas de Educación En Percepción Remota, 2007. <http://www.teledet.com.uy/tutorial-imagenes-satelitales/satelites-resolucion-espacial.htm>
- [53] Arango, M., Branch, J., & Botero, Clasificación no supervisada de coberturas vegetales sobre imágenes digitales de sensores remotos: “LANDSAT– ETM+.” Facultad Nacional de Agricultura Medellín, 2005, Pág. 2611–2634.
- [54] Anónimo, El tamaño de la celda y el remuestreo en análisis, Esri, 2021 <https://desktop.arcgis.com/es/arcmap/latest/extensions/spatial-analyst/performing-analysis/cell-size-and-resampling-in-analysis.htm>
- [55] David Revelo Luna, et al., Análisis de los índices de vegetación NDVI, GNDVI y NDRE para la caracterización del cultivo de café (Coffea arabica), 2020. <https://www.redalyc.org/journal/852/85269294002/html/#:~:text=Resumen%3A%20Los%20C3%ADndices%20de%20vegetaci%C3%B3n,estado%20nutricional%20o%20estado%20h%C3%ADdrico.>
- [56] Henry A. Adornado, Masao Yoshida, Crop Suitability and Soil Fertility Mapping using Geographic Information System (GIS), 2008, Pag. 5.
- [57] Balzter, H., Cole, B., Thiel, C., & Schmullius, C, Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. Remote Sensing, 2015. <https://doi.org/10.3390/rs71114876>
- [58] ESRI, Métodos de clasificación de datos, Clasificación de roturas naturales de Jenks en ArcMap. https://wiki.gis.com/wiki/index.php/Jenks_Natural_Breaks_Classification
- [59] Rafael Zambrano, Qué es R y por qué utilizarlo, Open Webinars, 2019. <https://openwebinars.net/blog/que-es-r-y-por-que-utilizarlo>
- [60] ESRI, Train Random Trees Classifier (Image Analyst). <https://pro.arcgis.com/en/pro-app/3.1/tool-reference/image-analyst/train-random-trees-classifier.htm>
- [61] Juan Ignacio Barrios Arce, La matriz de confusión y sus métricas, Inteligencia Artificial en Salud, Health Big Data, 2019. <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>

[62] NV5 GEOSPATIAL SOFTWARE, Calculate Confusion Matrices, 2024.
<https://www.nv5geospatialsoftware.com/docs/CalculatingConfusionMatrices.html>

[63] James, Gareth, Witten, Daniela, Hastie, Trevor; Tibshirani, Robert, Una introducción al aprendizaje estadístico, 2013, Págs. 316–321.

64 Janitza, Silke, Hornung, Romano, “Sobre la sobreestimación del error fuera de bolsa de Random Forest” Instituto de Procesamiento de Información Médica, Biometría y Epidemiología, Universidad de Munich, Munich, Alemania, 2018.

Bibliografías de figuras

[A] Mapeo satelital de las redes hidrográficas anexadas a las redes viales, zona de la Oroya, Ministerio de transportes y comunicaciones, Perú, 2016.

[B] Electromagnetic spectrum-es, Wikipedia, s.f.)
https://es.wikipedia.org/wiki/Espectro_visible#/media/Archivo:Electromagnetic_spectrum-es.svg

[C] Geoffrey J. Meaden et al. Canterbury, Los sistemas de información geográfica y la telepercepción en la pesca continental y la acuicultura, Cap. 4 La telepercepcion como fuente de datos, Kent CT2 7 JD, Reino Unido, 1992.
<https://www.fao.org/3/t0446s/t0446s04.htm>

[D] Calentamiento global y efecto invernadero (lo natural, la intensificación artificial), TRINI aprende, s.f. https://trinidadarroyo.org/aprende/?page_id=161

[E] Mariela Lorena Miño, Detección de basurales ilegales, rellenos sanitarios, ex basurales, tosqueras y chatarrerías en el gran buenos aires, mediante teledetección y sistemas de información geográfica, Universidad Nacional de Córdoba, 2012.
[.https://www.famaf.unc.edu.ar/documents/1317/9-Gulich-Mi%C3%B1o.pdf](https://www.famaf.unc.edu.ar/documents/1317/9-Gulich-Mi%C3%B1o.pdf)

[F] ENIIT, Sensores para la agricultura, s.f. <https://eniit.es/sensores-para-la-agricultura/>

[G] 61 Julio C Rodriguez, et al., Estimación de la evapotranspiración de un viñedo de uva de mesa (vitis vinífera) con imágenes satelitales sentinel, Universidad de Sonora, agosto 2021.

[H] Hernández, I. U., Ellis, E. A., Gallo, C. A. “Aplicación de teledetección y sistemas de información geográfica para el análisis de deforestación y deterioro de selvas tropicales en la región Uxpanapa, Veracruz”, 2013

[I] Monografías, ¡Así es la Costa Chica, Gro., México! (Estado de Guerrero,) s.f.
<https://www.monografias.com/trabajos96/a-asi-es-costa-chica-gro-mexico/a-asi-es-costa-chica-gro-mexico>

- [J] Cd. Obregon en Sonora, Fierro por la 200!!!, VALLE DEL YAQUI, s.f.
<https://obson.wordpress.com/2008/11/25/valle-del-yaqui/>
- [K] Gladys Toribio, Cursos teledetección, NDVI, ¿qué es y para qué sirve?, s.f.
<https://www.cursosteledeteccion.com/ndvi-que-es-y-para-que-sirve/>
- [L] Analytcs Vidhya, Understand Random Forest Algorithms With Examples (Updated 2024), s.f. <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- [M] Analytcs Vidhya, Understand Random Forest Algorithms With Examples (Updated 2024), s.f. <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- [N] Arc GIS Desktop, El tamaño de la celda y el remuestreo en análisis, s.f.
<https://desktop.arcgis.com/es/arcmap/latest/extensions/spatial-analyst/performing-analysis/cell-size-and-resampling-in-analysis.htm>
- [Ñ] Geo innova, Composiciones RGB de imágenes satélite, 2019. <https://geoinnova.org/blog-territorio/composiciones-rgb-de-imagenes-satelite/>
- [O] Prawo-budowlane, Infracción del límite de la parcela por parte de un vecino, 2006
https://www.prawo-budowlane.info/wznowienie-granic-z-naruszeniem-wlasnosci,458,material_prawo_budowlane.html
- [P] 123RF, Inmage Lab Pte Ltd, s.f. https://es.123rf.com/photo_79591552_vista-a%C3%A9rea-del-cultivo-de-ma%C3%ADz-campo-de-cultivo-de-ma%C3%ADz-a-partir-de-punto-de-vista-drone.html
- [Q] Pedro Muñoz Aguayo, Apuntes de teledetección: índices de vegetación, Centro de Información de Recursos Naturales, 2013.
- [R] Ali Essahlaoui, Groundwater potential of Middle Atlas plateaus, Morocco, using fuzzy logic approach, GIS and remote sensing, GIS classification 'Natural Breaks (Jenks)' of GWP, 2016. https://www.researchgate.net/figure/GIS-classification-Natural-Breaks-Jenks-of-GWP_fig5_302918488
- [S] Prácticas de Teledetección, Universidad de Murcia s.f.
<https://www.um.es/geograf/sig/teledet/lorca.html>
- [T] Arc GIS Desktop, Generalización de imágenes de ráster clasificadas, s.f.
<https://desktop.arcgis.com/es/arcmap/latest/tools/spatial-analyst-toolbox/generalization-of-classified-raster-imagery.htm>
- [U] Juan Ignacio Barrios Arce, La matriz de confusión y sus métricas, Inteligencia Artificial en Salud, Health Big Data, 2019. <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- [V] Microsoft Build, Resultados de los modelos de aprendizaje automático, 2023.
<https://learn.microsoft.com/es-es/dynamics365/finance/finance-insights/confusion-matrix>