



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE CIENCIAS DE LA INGENIERÍA

CARRERA INGENIERÍA AMBIENTAL

Proyecto de Investigación previo
a la obtención del título de
Ingeniero Ambiental

Título del Proyecto de Investigación:

**“ESTUDIO COMPARATIVO DE DOS MÉTODOS DE DETECCIÓN DE
CULTIVOS DE CACAO, BANANO Y PALMA EN LA ZONA OESTE DE LA
PROVINCIA DE LOS RÍOS, A PARTIR DE LAS IMÁGENES SATELITALES
LANDSAT-8 Y SENTINEL 2”**

Autor:

Elían Briohildo Intriago Giler

Directora de Proyecto de Investigación:

Geóg. Lidia Vlassova, PhD

Quevedo - Los Ríos – Ecuador.

2022



DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **Elían Briohildo Intriago Giler**, declaro que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

f. _____

Elían Briohildo Intriago Giler

C.C. # 1208177954



CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN

La suscrita, **Geóg. Lidia Vlassova, PhD**, Docente de la Universidad Técnica Estatal de Quevedo, certifica que el estudiante **Elían Briohildo Intriago Giler**, realizó el Proyecto de Investigación de grado titulado **“Estudio comparativo de dos métodos de detección de cultivos de cacao, banano y palma en la zona oeste de la provincia de Los Ríos, a partir de las imágenes satelitales Landsat-8 y Sentinel 2”**, previo a la obtención del título de Ingeniero Ambiental, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.



Firmado electrónicamente por:
LIDIA VLASSOVA

.....
Geóg. Lidia Vlassova, PhD

DIRECTORA DEL PROYECTO DE INVESTIGACIÓN



CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO

La suscrita Geóg. Lidia Vlassova PhD, mediante el presente cumpla en presentar a usted, el informe del Proyecto de Investigación titulado **“ESTUDIO COMPARATIVO DE DOS MÉTODOS DE DETECCIÓN DE CULTIVOS DE CACAO, BANANO Y PALMA EN LA ZONA OESTE DE LA PROVINCIA DE LOS RÍOS, A PARTIR DE LAS IMÁGENES SATELITALES LANDSAT-8 Y SENTINEL 2”**, presentado por el estudiante Sr. Elian Briohildo Intriago Giler, estudiante de la Carrera de Ingeniería Ambiental, que fue revisado bajo mi dirección según resolución Centésima Séptima Resolución del consejo académico de la Facultad de Ciencias de la Ingeniería, de fecha 31 de enero del 2022, desarrollado de acuerdo al Reglamento de la Unidad de Integración Curricular de la Universidad Técnica Estatal de Quevedo y que cumple con el requerimiento de análisis de URKUND, el cual mostró 0% de similitud.



Document Information

Analyzed document	Intriago_proyecto_inv.docx (D146176848)
Submitted	2022-10-12 04:31:00
Submitted by	
Submitter email	lvlassova@uteq.edu.ec
Similarity	0%
Analysis address	lvlassova.uteq@analysis.orkund.com

Sources included in the report

W URL: <https://www.sciencedirect.com/science/article/pii/S2666017221000055>
Fetched: 2021-03-05 16:51:07



1



Firmado electrónicamente por:
LIDIA VLASSOVA

Geóg. Lidia Vlassova PhD.

DIRECTORA DEL PROYECTO DE INVESTIGACIÓN



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA
CARRERA DE INGENIERÍA AMBIENTAL

PROYECTO DE INVESTIGACIÓN

Título:

“Estudio comparativo de dos métodos de detección de cultivos de cacao, banano y palma en la zona oeste de la provincia de Los Ríos, a partir de las imágenes satelitales Landsat-8 y Sentinel 2”

Presentado a la Consejo Directivo como requisito previo a la obtención del título de Ingeniero Ambiental.

Aprobado por:

OSCAR
OSWALDO
PRIETO
BENAVIDES

Firmado digitalmente por
OSCAR OSWALDO
PRIETO
BENAVIDES
Fecha: 2022.10.09
15:26:04 -05'00'

PRESIDENTE DEL TRIBUNAL

Ing. Oscar Prieto Benavides PhD.

**YARELYS
FERRER
SANCHEZ**

Firmado digitalmente por YARELYS
FERRER SANCHEZ
Nombre de reconocimiento (DN):
cn=YARELYS FERRER SANCHEZ,
serialNumber=080722171544,
ou=ENTIDAD DE CERTIFICACION
DE INFORMACION, o=SECURITY
DATA S.A. 2, c=EC
Fecha: 2022.11.09 20:22:58 -05'00'



Firmado electrónicamente por:
**CARLOS ALBERTO
NIETO CAÑARTE**

INTEGRANTE DEL TRIBUNAL
Blga. Yarelys Ferrer Sánchez PhD.

INTEGRANTE DEL TRIBUNAL
Ing. Carlos Nieto Cañarte MSc.

QUEVEDO – LOS RÍOS – ECUADOR

2022

AGRADECIMIENTO

Agradezco a Otto y Silvia, mis padres, quiénes con su apoyo incondicional durante toda mi carrera universitaria han permitido este logro tan importante en mi vida. Sin duda alguna, son los principales artífices.

De la misma forma, agradezco a todos los docentes de la carrera de Ingeniería Ambiental de la Universidad Técnica Estatal de Quevedo. De forma especial, agradezco a la geógrafa Lidia Vlassova, quien, como mi tutora de tesis, fue la principal guía para el desarrollo del presente proyecto de investigación.

Finalmente, agradezco a Gabriela y Karla, mis compañeras y amigas incondicionales en cada semestre de la carrera, con quiénes compartí momentos de frustración, tristeza, satisfacción y felicidad.

DEDICATORIA

*A mi abuelo Gregorio, quién está en mi corazón
hasta el día que nos encontremos otra vez.*

A mis hermanos Damián y Alan.

RESUMEN EJECUTIVO Y PALABRAS CLAVES

La teledetección es muy importante en el ordenamiento territorial, por lo tanto, los nuevos métodos requieren ser comprobados en situaciones reales. Esta investigación comparó la precisión del método tradicional en teledetección, Maximum Likelihood (ML), frente a un método basado en inteligencia artificial denominado Random Forest (RF), en la detección de cultivos de banano, cacao y palma (y otras coberturas). Se utilizaron imágenes satelitales de Landsat-8 y Sentinel-2, las cuales fueron descargas desde el sitio web de la USGS; a las bandas de la imagen de Sentinel-2 se les aplicó corrección atmosférica. Para las dos imágenes se calcularon los índices espectrales NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), GCI (Green Coverage Index) y MSI (Moisture Stress Index) y se crearon las regiones de interés (ROIs) sobre el stack de índices calculados de cada imagen (86 en Landsat-8 y 85 en Sentinel-2); para crear los ROIs (tanto para entrenamiento y validación de resultados) se tomaron coordenadas *in situ* y desde Google Earth Pro versión 7.3.4.8642. La clasificación con ML se realizó en el software ENVI 5.3 y la clasificación con RF se la realizó en QGIS 3.22.7 con las herramientas de Dzetsaka. Para la validación de resultados se utilizó la matriz de confusión tradicional y el coeficiente de Kappa; además, se empleó la prueba estadística no paramétrica Wilcoxon-Mann-Whitney para verificar si existieron diferencias estadísticas entre las clasificaciones. Los resultados demostraron que, ML detectó con mayor precisión los cultivos de interés, tanto en la imagen de Landsat-8 (sin diferencias significativas) como en la de Sentinel-2; no obstante, la clasificación general más precisa (incluyendo todas las coberturas) fue la de Sentinel-2 utilizando RF. Por lo tanto, se sugiere continuar estudiando el potencial de RF con condiciones más favorables (Imagen en temporada seca y mayor cantidad de muestras).

Palabras clave: Maximum Likelihood, Random Forest, índices espectrales, coberturas de suelo.

ABSTRACT AND KEYWORDS

Remote sensing is very important in land management; therefore, new remote sensing methods need to be tested in real situations. This research compared the accuracy of the traditional remote sensing method, Maximum Likelihood (ML), versus an artificial intelligence-based method called Random Forest (RF), in the detection of banana, cocoa, and palm crops (and other land covers). Landsat-8 and Sentinel-2 satellite images were used, which were downloaded from the USGS website; the atmospheric correction was applied to the Sentinel-2 bands. The spectral indices NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), GCI (Green Coverage Index), and MSI (Moisture Stress Index) were calculated for the two images, and regions of interest (ROIs) were created on the stack of indices calculated from each image (86 in Landsat-8 and 85 in Sentinel-2); to make the ROIs (both for training and validation), coordinates were taken *in situ* and from Google Earth Pro V 7. 3.4.8642. ML classification was run in ENVI 5.3 software and RF classification was run in QGIS 3.22.7 with the Dzsaka plugin. The traditional confusion matrix and the Kappa coefficient were used to validate the results; in addition, the Wilcoxon-Mann-Whitney nonparametric statistical test was used to verify statistical differences between the classifications. The results showed that ML detected with better accuracy the crops of interest in Landsat-8 (without significant differences) and Sentinel-2 images; however, the most accurate overall classification (including all coverages) was with Sentinel-2 image using RF. Therefore, this study suggests continuing to study the RF potential with more favorable conditions (dry season image and more samples).

Keywords: Maximum Likelihood, Random Forest, spectral index, land cover.

TABLA DE CONTENIDO

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	ii
CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN.....	iii
CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO.....	iv
CERTIFICADO DE APROBACIÓN POR TRIBUNAL DE SUSTENTACIÓN	v
AGRADECIMIENTO	vi
DEDICATORIA.....	vii
RESUMEN EJECUTIVO Y PALABRAS CLAVES	viii
ABSTRACT AND KEYWORDS	ix
TABLA DE CONTENIDO	x
ÍNDICE DE TABLAS.....	xiii
ÍNDICE DE FIGURAS	xiv
ÍNDICE DE ECUACIONES	xv
ÍNDICE DE ANEXOS	xvi
CÓDIGO DUBLIN.....	xvii
Introducción.....	1
CAPÍTULO I.....	3
CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN	3
1.1 Problema de investigación	4
1.1.1 Planteamiento del problema	4
1.1.2 Diagnóstico.....	5
1.1.3 Pronóstico	5
1.1.4 Formulación del problema.....	6
1.1.5 Sistematización del problema.....	6
1.2 Objetivos.....	6
1.2.1 Objetivo general	6

1.2.2	Objetivos específicos	6
1.3	Justificación	7
CAPÍTULO II.....		9
FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN		9
2.1	Marco conceptual.....	10
2.1.1	Firmas espectrales.....	10
2.1.2	Ordenamiento territorial	10
2.1.3	Sistemas de producción agrícola sostenibles	10
2.1.4	Teledetección.....	11
2.1.5	Imágenes multiespectrales	11
2.1.6	Bandas espectrales.....	12
2.1.7	Sentinel-2.....	12
2.1.8	Landsat-8	14
2.1.9	Clasificación de imágenes multiespectrales	15
2.1.10	Índices espectrales	16
2.2	Marco referencial	17
2.2.1	Mapeo de la cobertura vegetal en la subcuenca Quillcay (Ancash, Perú) con el clasificador de Árbol de decisiones	18
2.2.2	Una evaluación de los datos de Landsat, Sentinel-2, Sentinel-1 y MODIS para el mapeo de tipos de cultivos.....	18
2.2.3	Comparación de algoritmos de clasificación para detectar recuperación costera típica en la provincia Guangdong con imágenes de Landsat-8 y Sentinel-2 ..	19
2.2.4	Deep learning para la clasificación de usos de suelo agrícola con Sentinel-2	20
CAPÍTULO III		22
METODOLOGÍA DE LA INVESTIGACIÓN		22
3.1	Localización.....	23
3.2	Tipo de investigación.....	25

3.3	Métodos de investigación	25
3.4	Fuentes de recopilación de información	26
3.5	Diseño de la investigación	26
3.5.1	Objetivo 1: Preprocesar las imágenes multiespectrales de Landsat-8 y Sentinel-2.....	27
3.5.2	Objetivo 2: Calcular los índices multiespectrales NDVI, GNDVI, GCI y MSI a partir de las imágenes multiespectrales preprocesadas.	31
3.5.3	Objetivo 3: Ejecutar la clasificación supervisada de imágenes con los métodos Maximum Likelihood y Random Forest.	33
3.5.4	Objetivo 4: Comprobar el nivel de precisión de los métodos utilizados en la clasificación supervisada	38
3.5.5	Objetivo 5: Analizar los resultados obtenidos de la clasificación supervisada de imágenes multiespectrales de Landsat-8 y Sentinel-2	40
3.6	Recursos humanos y materiales	40
3.6.1	Materiales tecnológicos	40
3.6.2	Materiales de oficina	40
3.6.3	Programas informáticos.....	41
3.6.4	Sitios web	41
CAPÍTULO IV		42
RESULTADOS Y DISCUSIÓN		42
4.1	Preprocesamiento de imágenes multiespectrales	43
4.1.1	Selección de imagen Landsat-8	43
4.1.2	Selección de imagen Sentinel-2.....	44
4.2	Recorte de imágenes para el área de estudio y unión de bandas espectrales necesarias.....	46
4.3	Cálculo de índices espectrales	50
4.3.1	Cálculo individual de los índices NDVI, GNDVI, MSI y GCI.....	50
4.3.2	Stack de índices	53

4.4	Clasificación supervisada.....	56
4.4.1	Creación de áreas de entrenamiento	56
4.4.2	Creación de los ROIs	58
4.4.3	Cálculo de separabilidad espectral de ROIs	59
4.4.4	Clasificación supervisada	63
4.5	Validación de resultados	71
4.5.1	Áreas de entrenamiento al azar.....	71
4.5.2	Matrices de confusión.....	72
4.5.3	Coeficientes de Kappa	75
4.6	Análisis estadístico.....	75
4.7	Discusión.....	77
CAPÍTULO V.....		79
CONCLUSIONES Y RECOMENDACIONES		79
5.1	Conclusiones.....	80
5.2	Recomendaciones	81
CAPÍTULO VI		82
BIBLIOGRAFÍA		82
6.1	Bibliografía.....	83
CAPÍTULO VII.....		94
ANEXOS		94

ÍNDICE DE TABLAS

Tabla 1.	Bandas de Sentinel-2.	12
Tabla 2.	Bandas de Landsat-8.....	14
Tabla 3.	Modelo de registro de coordenadas para definición de áreas de entrenamiento ..	34
Tabla 4.	Modelo matriz de confusión con adición del resultado del coeficiente de Kappa	39
Tabla 5.	Características de imagen de Landsat-8	43
Tabla 6.	Características de imagen de Sentinel-2.....	45

Tabla 7. Áreas de entrenamiento	56
Tabla 8. ROIs para Stacks de índices	58
Tabla 9. Separabilidad espectral Jeffries-Matusita de los ROIs del stack de índices calculados de Landsat-8.....	60
Tabla 10. Separabilidad espectral Jeffries-Matusita de los ROIs del stack de índices calculados de Sentinel-2	62
Tabla 11. Extensión de coberturas identificadas con el clasificador Maximum Likelihood para la imagen de Landsat-8.....	63
Tabla 12. Extensión de coberturas identificadas con el clasificador Maximum Likelihood para la imagen de Sentinel-2	65
Tabla 13. Extensión de coberturas identificadas con el clasificador Random Forest para la imagen de Landsat-8.....	67
Tabla 14. Extensión de coberturas identificadas con el clasificador Random Forest para la imagen de Sentinel-2	69
Tabla 15. Áreas de entrenamiento al azar para creación de nuevos ROIs y generación de matrices de confusión y cálculo de coeficientes de Kappa	71
Tabla 16. Coeficientes de Kappa para las clasificaciones realizadas con Maximum Likelihood y Random Forest.	75
Tabla 17. Estadísticos de prueba de Suma de Rangos de Wilcoxon para las comparaciones estadísticas entre métodos clasificadores. a: variable de agrupación ML1RF2. b:variable de agrupación ML1RF2.....	76

ÍNDICE DE FIGURAS

Figura 1. Longitudes de onda	11
Figura 2. Mapa de ubicación del área de estudio: Cantones Mocache, Palenque, Vinces y Baba.....	24
Figura 3. Flujograma de la investigación	27
Figura 4. Interfaz de sitio web de la USGS	28
Figura 5. Corrección atmosférica de bandas espectrales de Sentinel-2 con Semi Automatic Classification Plugin del año 2021.	29
Figura 6. Polígono definitivo del sitio de estudio en función de las coberturas reales de las imágenes de Landsat-8 y Sentinel-2.....	30

Figura 7. Configuración de unión de bandas espectrales necesarias en función del polígono de la zona de estudio	31
Figura 8. Ejemplo de cálculo de índice espectral para la imagen Sentinel-2.....	33
Figura 9. Configuración para el algoritmo clasificador Random Forest con las herramientas de “Dzetsaka”.....	36
Figura 10. Scripts de “Water in Wetlands Index” y “Natural Color” para la comprobación de zonas inundadas en imágenes de Sentinel 2 a partir de QGIS y Sentinel Hub.....	37
Figura 11. Combinación de bandas del índice “Water in Wetlands Index” del 17 de abril del 2018 en el área de estudio.....	38
Figura 12. Combinación RGB de imagen multiespectral adquirida de Landsat-8 con fecha 27/11/2016 más el área de estudio sobrepuesta.....	44
Figura 13. Combinación RGB de imagen satelital adquirida de Sentinel-2 con fecha 17/04/2018 más el área de estudio sobrepuesta.....	46
Figura 14. Recorte de área de estudio para imagen de Landsat-8.....	48
Figura 15. Recorte de área de estudio para imagen de Sentinel-2	49
Figura 16. Representación de índices espectrales de Landsat-8.....	51
Figura 17. Representación de índices espectrales de Sentinel-2	52
Figura 18. Stack de índices calculados de Landsat-8.....	54
Figura 19. Stack de índices calculados de Sentinel-2.....	55
Figura 20. Clasificación supervisada con ML de la imagen de Landsat-8	64
Figura 21. Clasificación supervisada con ML de la imagen de Sentinel-2	66
Figura 22. Clasificación supervisada con RF de la imagen de Landsat-8	68
Figura 23. Clasificación supervisada con RF de imagen de Sentinel-2	70
Figura 24. Matrices de confusión de clasificaciones realizadas.....	73

ÍNDICE DE ECUACIONES

Ecuación 1. NDVI	32
Ecuación 2. GNDVI.....	32
Ecuación 3. GCI.....	32
Ecuación 4. MSI	32

ÍNDICE DE ANEXOS

Anexo 1. Metadatos completos de la imagen adquirida de Landsat-8.....	95
Anexo 2. Metadatos completos de la imagen adquirida de Sentinel-2	96
Anexo 3. Unión de bandas necesarias para la imagen de Sentinel-2 con la opción “Layer Stacking”	98
Anexo 4. Informe completo de separabilidad espectral de ROIs de Landsat-8	99
Anexo 5. Informe completo de separabilidad espectral de ROIs de Sentinel-2.....	100
Anexo 6. Toma de coordenadas para ROIs de zonas con coberturas de cacao.....	101
Anexo 7. Toma de coordenadas para ROIs (de validación) de zonas con coberturas de cacao y zonas con otras coberturas.....	101
Anexo 8. Generación de puntos al azar para análisis de diferencias estadísticas con la herramienta “Create Random Points”	102
Anexo 9. Extracción de puntos al azar para análisis de diferencias estadísticas con la herramienta “Extract Multi Values to Point”	102
Anexo 10. Gráfico Q-Q para análisis de normalidad de los datos de ML-L8.....	103
Anexo 11. Gráfico Q-Q para análisis de normalidad de los datos de RF-L8.....	103
Anexo 12. Gráfico Q-Q para análisis de normalidad de los datos de ML-S2.....	104
Anexo 13. Gráfico Q-Q para análisis de normalidad de los datos de RF-S2.....	104

CÓDIGO DUBLIN

Título:	Estudio comparativo de dos métodos de detección de cultivos de cacao, banano y palma en la zona oeste de la provincia de Los Ríos, a partir de las imágenes satelitales Landsat-8 y Sentinel 2				
Autor:	<u>Intriago Giler, Elian Briohildo</u>				
Palabras clave:	Máximum Likelihooh	Random Forest	Índices espectrales	Coberturas de suelo	
Fecha de publicación:	Diciembre 2022				
Editorial:	Universidad Técnica Estatal de Quevedo				
Resumen: (hasta 300 palabras)	<p>Resumen.- Se comparó la precisión de Maximum Likelihood (ML) y Random Forest (RF), en la detección de cultivos de banano, cacao y palma (y otras coberturas). Se utilizaron imágenes de Landsat-8 y Sentinel-2. Se calcularon los índices espectrales NDVI, GNDVI, GCI y MSI y se crearon las regiones de interés (ROIs); para crear los ROIs (de entrenamiento y validación) se tomaron coordenadas <i>in situ</i> y desde Google Earth. Para la validación se utilizó una matriz de confusión y coeficientes de Kappa; además, se empleó el test Wilcoxon-Mann-Whitney para verificar diferencias estadísticas. ML detectó con mayor precisión los cultivos de interés, tanto en Landsat-8 (sin diferencias significativas) como en Sentinel-2; no obstante, la clasificación general más precisa (incluyendo todas las coberturas) fue la de Sentinel-2 utilizando RF. Por lo tanto, se sugiere continuar estudiando el potencial de RF con condiciones más favorables (Imagen en temporada seca y mayor cantidad de muestras).</p> <p>Abstract.- The accuracy of Maximum Likelihood (ML) and Random Forest (RF) was compared in the detection of banana, cocoa, and palm crops (and other covers). Landsat-8 and Sentinel-2 images were used. NDVI, GNDVI, GCI and MSI spectral indices were calculated, and regions of interest (ROIs) were created; to make the ROIs (training and validation) coordinates were taken <i>in situ</i> and from Google Earth. For validation, a confusion matrix and the Kappa coefficients were obtained; in addition, the Wilcoxon-Mann-Whitney test was used to verify statistical differences. ML detected the crops of interest with greater precision, both in Landsat-8 (without significant differences) and in Sentinel-2; however, however, the most accurate overall classification (including all coverages) was with Sentinel-2 image using RF. Therefore, this study suggests continuing to study the RF potential with more favorable conditions (dry season image and more samples).</p>				
Descripción:	122 hojas : dimensiones, 29 x 21 cm + CD-ROM 6162				
URI:					

Introducción.

Uno de los problemas actuales más graves que afecta al medio ambiente natural del mundo es la expansión e intensificación de la agricultura desorganizada e insostenible. Además, el aumento de la demanda de alimentos, piensos y energía conduce a una agricultura y una transformación del paisaje más intensivas [1]. Dicha expansión agrícola insostenible genera degradación del suelo, sobre todo en áreas fértiles [2], en donde el control de las actividades humanas por parte de entidades estatales es limitado.

En Sudamérica la expansión agrícola insostenible también representa un problema. En las últimas décadas ocurrieron grandes transformaciones del paisaje, sobre todo por el incremento de cultivos con fines de exportación. Como ejemplo claro de este problema, en cinco países del continente desde el ciclo 2000/2001 hasta el ciclo 2010/2011 se observó un incremento del 43% de la superficie cultivada [3]. Por lo tanto, resulta indispensable una planificación territorial integral que permita la evolución planificada de los sistemas territoriales en la región.

A nivel nacional la problemática no es distinta, mucho más siendo Ecuador un país exportador de productos agrícolas, entre los cuales figuran el banano, el maíz amarillo, la caña de azúcar, soja, cacao, entre otros [4]. El Ministerio de Agricultura, Ganadería, Acuacultura y Pesca del Ecuador (MAGAP) afirma que la degradación de los suelos en el país está relacionada a prácticas agrícolas sin considerar un ordenamiento territorial basado en la vocación agrícola de los suelos. Esta problemática también se da en la provincia Los Ríos, la cual es la segunda zona agro productiva más apta a nivel nacional [5]. En dicha provincia, los cultivos predominantes son el cacao, el café, el arroz, el banano y la palma africana.

El insumo principal para el ordenamiento territorial es información confiable sobre las coberturas de suelo. Precisamente, una medida que permite cubrir la carencia de información objetiva y actualizada sobre la distribución de cultivos es la teledetección a través del procesamiento y clasificación de imágenes multiespectrales. Para la clasificación es común que se utilicen métodos tradicionales; no obstante, en los últimos años se han propuesto estrategias alternativas basadas en técnicas de inteligencia artificial y

aprendizaje automático [6]. Entre dichas técnicas avanzadas figuran métodos de redes neuronales artificiales, máquinas de vectores de soporte, teoría difusa y árboles de decisión [7]. El presente proyecto comparó los resultados obtenidos entre el método tradicional de Máxima Probabilidad o Maximum Likelihood (ML) en inglés, versus un método alternativo denominado Random Forest (RF), para detección de cacao, banano y palma a partir de imágenes satelitales de Landsat-8 y Sentinel-2.

CAPÍTULO I

CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

1.1 Problema de investigación.

1.1.1 Planteamiento del problema.

Existe una necesidad urgente de aumentar la producción agrícola para satisfacer la creciente demanda de alimentos impulsada por el crecimiento de la población y las preferencias alimentarias cambiantes [8]. Como consecuencia de esta necesidad de aumento en la producción agrícola, y de los problemas ambientales que incurren en la expansión de cultivos y el uso progresivo de la tierra para fines agrícolas sin un ordenamiento territorial integral, se incumplen varios objetivos planteados en la Agenda 2030 para lograr el Desarrollo Sostenible [9].

En la zona oeste de la provincia Los Ríos, la problemática sigue la misma línea; el cultivo de cacao a pesar de ser una fuente de seguridad económica conlleva un coste medioambiental elevado [10]. Entre los costes ambientales figuran principalmente la contaminación del suelo y la contaminación de los recursos hídricos superficiales y subterráneos. Los problemas mencionados son producto de la expansión insostenible de este tipo de cultivos y los residuos tóxicos que se generan a lo largo del proceso agrícola.

Por otra parte, el cultivo de banano también causa impactos ambientales; el uso de agua aumenta durante las temporadas secas, presumiendo que es una consecuencia de las afectaciones del cambio climático; además, rompe la tendencia de otros procesos productivos, generando mayor huella de carbono en la etapa agrícola que en la fase de transporte [11]. Por lo expuesto, es evidente que, a pesar de ser un pilar fundamental en el Producto Interno Bruto (PIB) de la zona, la expansión agrícola desordenada del presente cultivo genera una huella de carbono considerable.

En relación a la expansión agrícola de la palma, las transformaciones resultantes del uso de la tierra para la producción de este cultivo, están asociadas tanto con la degradación ambiental, como con las emisiones sustanciales de dióxido de carbono [12]. Además, la tendencia siempre ha sido culpar del problema a los pequeños agricultores [13]; sin embargo, la promulgación de políticas públicas que catalizan la expansión de palma africana [14], benefician principalmente a grandes productores, los mismos que proliferan

la distribución desordenada de este tipo de monocultivos.

En relación de la problemática que antecede, es evidente que existe la necesidad de aplicar acciones y/o medidas que permitan responder integralmente ante los impactos ambientales generados por la expansión insostenible de zonas agrícolas. Una medida es la planificación territorial eficaz con información geográfica objetiva y actualizada. El tratamiento de imágenes satelitales de acceso libre permite gestionar acciones oportunas en relación con la expansión agrícola; no obstante, hay metodologías relativamente nuevas, que requieren ser comprobadas en casos reales de estudio.

1.1.2 Diagnóstico.

Según el Plan de Desarrollo y Ordenamiento Territorial (PDyOT) de Baba, el uso de suelo está dominado por arboricultura tropical, cultivos de arroz, banano, cultivos de ciclo corto, pasto cultivado, cuerpos de agua, vías y zonas urbanas [15]. Por otro lado, según el PDyOT de Mocache, el área registra 15 tipos de usos de suelo, entre los cuales figuran: arborícola tropical, arroz, banano, bosque natural, cultivos de ciclo corto, pasto cultivado, pasto natural, vegetación arbustiva y cuerpos de agua [16]. Por su parte, en Palenque predominan la actividad comercial, cultivos de maíz, cacao, plátano, café y guanábana, y cuerpos de agua como ríos, esteros, pozas y albarradas [17]. Finalmente, el uso de suelo en Vinces se subdivide en cultivos de cacao, banano, arroz, soya, sandía, mango y maíz, cuerpos de agua y zonas pobladas [18].

En relación con las condiciones climáticas del área de estudio, la temperatura promedio durante el año es de 25°C con fluctuaciones entre 24°C y 27°C; los meses de lluvia son de diciembre a mayo, con precipitación máxima de 429 mm en febrero. Cabe agregar que, en el área de estudio la mayor parte del suelo es de textura fina, media y moderadamente gruesa [19].

1.1.3 Pronóstico.

Con el uso de la teledetección, aplicando métodos precisos para la detección de cultivos y otras coberturas de suelo, se puede identificar características relevantes de la vegetación

tales como: estado saludable, actividad fotosintética, estrés hídrico; además de, posibles usos insostenibles del suelo, lo que ayudará a los tomadores de decisiones de carácter estatal o privado a responder oportuna e integralmente ante cualquier problemática identificada.

1.1.4 Formulación del problema.

¿Los métodos que utilizan algoritmos de aprendizaje automático como Random Forest, tienen mayor, menor o similar precisión que el método Maximum Likelihood para la detección de cultivos de cacao, banano y palma?

1.1.5 Sistematización del problema.

¿Cuán óptimas son las imágenes multiespectrales adquiridas para el cálculo de los índices espectrales?

¿Es posible representar la distribución y características de los cultivos de cacao, banano y palma del área de estudio con los métodos Maximum Likelihood y Random Forest?

¿Cuál es el nivel de precisión de los métodos utilizados para la clasificación de las imágenes multiespectrales derivadas de Landsat-8 y Sentinel-2?

1.2 Objetivos.

1.2.1 Objetivo general.

Comparar la precisión de los métodos Maximum Likelihood y Random Forest en la detección de cultivos de cacao, banano y palma de la zona oeste de la provincia de Los Ríos, a partir de imágenes multiespectrales de los satélites Landsat-8 y Sentinel-2.

1.2.2 Objetivos específicos.

- Preprocesar las imágenes multiespectrales de Landsat-8 y Sentinel-2.
- Calcular los índices espectrales NDVI, GNDVI, GCI y MSI a partir de las imágenes multiespectrales preprocesadas.

- Ejecutar la clasificación supervisada de imágenes con los métodos Maximum Likelihood y Random Forest.
- Comprobar el nivel de precisión de los métodos utilizados en la clasificación supervisada.
- Analizar los resultados de la clasificación supervisada de imágenes multiespectrales de Landsat-8 y Sentinel-2.

1.3 Justificación.

En el área de estudio los cultivos de cacao, banano y palma ocupan una gran extensión de la superficie de la zona; debido a esto, representan una parte importante del producto interno bruto de la provincia. Por lo tanto, la detección y representación de sus características a partir de los datos de sensores remotos es transcendental para realizar un análisis integral de su distribución en la zona, y de esta forma mejorar el nivel de eficiencia de la planificación territorial.

El análisis integral de la distribución de los cultivos y sus características usando las herramientas de los sistemas de información geográfica, permitiría ordenamiento territorial [20] basado en las interrelaciones del medio biótico y abiótico y formulación de las estrategias más eficientes. Con dicha gestión integral, es posible direccionar las acciones de diversos procesos productivos hacia una menor huella de carbono, con la finalidad de contribuir al objetivo 15 del Desarrollo Sostenible sobre la importancia de la vida de diversos ecosistemas terrestres.

Pese a lo nuevas que parezcan las técnicas de teledetección, el procesamiento y la clasificación de imágenes multiespectrales para el análisis integral de la distribución agrícola, la determinación de su frontera y la documentación del problema, han estado ligadas a la utilización de SIG desde hace décadas [21]. No obstante, han surgido nuevos métodos a partir de los avances en el campo de la programación informática; muchos de estos métodos utilizan algoritmos de aprendizaje automático, entre los cuales figura el empleado en este proyecto de investigación (Random Forest) [22].

En Ecuador y en la provincia Los Ríos aún no es común utilizar algoritmos de clasificación de coberturas de suelo; no obstante, con el empleo de este tipo de técnicas se pueden obtener resultados más precisos, con menos tiempo y recursos. Por tal motivo, es importante seguir evaluando la aplicabilidad de este tipo de herramientas. En el presente estudio, se comparó la precisión del método Random Forest y el método tradicional Maximum Likelihood, para la detección de cultivos de cacao, banano y palma en el oeste de la provincia de Los Ríos, a partir de imágenes satelitales de Landsat-8 y Sentinel-2; calculando los índices espectrales NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Vegetation Index), GCI (Green Coverage Index), MSI (Moisture Stress Index).

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN

2.1 Marco conceptual.

2.1.1 Firmas espectrales.

Una firma espectral es una curva que representa la variación de la reflectancia de un objeto en función de la longitud de onda [23]. De este modo, las distintas superficies responden de manera diferente a la radiación electromagnética, lo que significa que se puede obtener una firma espectral específica para cada tipo de superficie u objeto. Por lo tanto, las distintas superficies se pueden identificar en base a sus firmas espectrales, pero será necesario que el espectro sea suficientemente detallado en términos de intervalos de longitud de onda y que cubra un rango espectral ancho [23].

2.1.2 Ordenamiento territorial.

El ordenamiento territorial es una ciencia que integra a diferentes sectores y que tiene por objeto una adecuada organización del territorio, con la finalidad de darle un mejor uso, aprovechamiento y protección. Por lo tanto, el ordenamiento territorial se puede definir como una política pública integral que trata de solucionar problemas a partir de oportunidades territoriales, de acuerdo con los objetivos pretendidos por los tomadores de decisiones en función de las necesidades de la población y del medio en el que desarrolla sus actividades [24].

2.1.3 Sistemas de producción agrícola sostenibles.

La sostenibilidad de los sistemas de producción agrícola se refiere a la capacidad de dichos sistemas para mantener su productividad a pesar de las perturbaciones económicas y naturales, externas o internas. De este modo, los sistemas de producción agrícolas sostenibles son aquellos capaces de contrarrestar las presiones negativas (resilientes) del entorno, gracias a las fortalezas naturales del área y a las intervenciones positivas de índole social, técnico y económico [25].

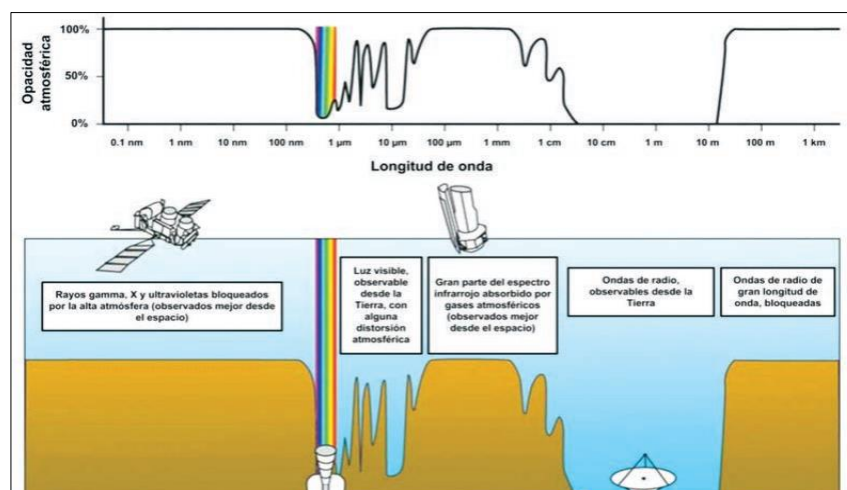
2.1.4 Teledetección.

La teledetección es una disciplina que emplea diversas técnicas y herramientas para captar información real de un fenómeno sobre la superficie terrestre. En una imagen de teledetección, se incluye información de contenido, información espacial e información temporal en metadato [26]. La teledetección y la ciencia de la información geográfica (GIScience) se han utilizado desde hace décadas para analizar los recursos naturales y el medio ambiente, proporcionando datos importantes para su gestión; no obstante, con el avance de la tecnología, han surgido nuevas técnicas y herramientas más sofisticadas y precisas que han aumentado la importancia de esta disciplina [27].

2.1.5 Imágenes multispectrales.

Las imágenes multispectrales son una matriz 3D [28] que usualmente tienen de 3 a 10 bandas espectrales y escáneres que funcionan en regiones particulares del espectro electromagnético. Entre los satélites de observación de la Tierra que generan imágenes multispectrales están Landsat, SPOT, RapidEye y Worldview-2 [29]. Por lo general, las imágenes multispectrales contienen información de longitudes de onda que el ojo humano no puede ver (Figura 1), y a partir de allí se explica su utilidad.

Figura 1
Longitudes de onda.



FUENTE: Ordoñez (2012). *Espectro electromagnético y espectro radioeléctrico.*

2.1.6 Bandas espectrales.

Las bandas espectrales son un intervalo de longitudes de onda del espectro electromagnético. Por ejemplo, la banda roja, la banda verde y la banda azul tienen longitudes de onda entre 0.4 y 0.7 μm , y la unión de estas tres bandas espectrales en una imagen constituye el espectro visible (RGB). Por lo tanto, se llama banda a la información adquirida por los detectores de un satélite en una banda espectral, entendiéndose como “bandas de una imagen” de teledetección [30]. Las imágenes multiespectrales contienen diversas bandas espectrales con varias longitudes de ondas útiles para el análisis de distintos fenómenos.

2.1.7 Sentinel-2.

El proyecto Sentinel-2 es una misión de satélites europeos diseñada para entregar datos de teledetección terrestre; la misma que consta de satélites gemelos, los cuales son Sentinel-2A y Sentinel-2B. Los satélites mencionados proporcionan observaciones de alta resolución espacial (10 m, 20 m y 60 m dependiendo de la banda) [31]. Con Sentinel-2 se obtienen imágenes de la Tierra desde una altitud aproximada de 786 km y permiten el análisis remoto de distintos fenómenos [32].

2.1.7.1 Bandas espectrales de Sentinel-2.

Las imágenes satelitales de Sentinel-2 están constituidas por 13 bandas espectrales [33], las cuales se detallan en la tabla 1.

Tabla 1
Bandas de Sentinel-2

Número de banda	Nombre de banda	Resolución espacial (m)
1	Aerosol costero	60
2	Azul	10
3	Verde	10
4	Rojo	10

5	Borde rojo de vegetación 1	20
6	Borde rojo de vegetación 2	20
7	Borde rojo de vegetación 3	20
8	Infrarroja cercana (NIR)	10
8A	Infrarroja cercana (NIR) estrecha	20
9	Vapor de agua	60
10	SWIR - Cirrus	60
11	SWIR 1	20
12	SWIR 2	20

FUENTE: European Space Agency [33].

2.1.7.2 Tipos de productos de Sentinel-2.

El programa Sentinel ofrece productos con distintos tipos de datos; los tipos de productos que dispone Sentinel-2 son los de nivel 1C y de nivel 2A. Los productos de nivel 1C contienen reflectancia en la parte superior de la atmósfera en geometría cartográfica, con un volumen de datos de 600MB cada 100 km por 100 km²; por otro lado, los productos de nivel 2A tienen reflectancia del fondo de la atmósfera en geometría cartográfica con un volumen de datos de 600MB cada 100 km por 100 km² [33].

2.1.7.3 Corrección atmosférica de productos de Sentinel-2.

La corrección atmosférica a imágenes de Sentinel-2 se la realiza con la finalidad de obtener resultados más precisos en las clasificaciones digitales. Por lo tanto, el proceso de corrección atmosférica es un preprocesamiento necesario cuando las imágenes disponibles para los usuarios tienen valores de reflectancia “Top Of Atmosphere” en geometría cartográfica (producto de nivel 1C). La corrección atmosférica permite pasar a imágenes de reflectancia Bottom Of Atmosphere (producto de nivel 2A) [34], entendiéndose como un nivel más sofisticado de procesamiento de la imagen satelital.

2.1.8 Landsat-8.

Landsat-8 es un satélite (el octavo) del programa satelital Landsat, el cual se lanzó en febrero de 2013 [35]. Con más de 40 años, los satélites Landsat proporcionan el registro temporal más largo de observaciones de la superficie terrestre basadas en el espacio, y el exitoso lanzamiento de Landsat-8 en 2013 continúa con este legado. En Landsat-8, se han mejorado diversas características (como la calibración y la resolución radiométrica) con respecto a Landsat-7 Enhanced Thematic Mapper (ETM+) [36]. Este satélite proporciona imágenes con resolución espacial de 15 m, 30 m y 100 m dependiendo de la banda [37].

2.1.8.1 Bandas espectrales de Landsat-8.

Las imágenes satelitales de Landsat-8 están constituidas por 11 bandas espectrales [37], las cuales se detallan en la tabla 2.

Tabla 2
Bandas de Landsat-8

Número de banda	Nombre de banda	Resolución espacial (m)
1	Aerosol costero	30
2	Azul	30
3	Verde	30
4	Rojo	30
5	Infrarroja cercano (NIR)	30
6	Infrarroja de Onda Corta 1 (SWIR 1)	30
7	Infrarroja de Onda Corta 2 (SWIR 2)	30
8	Pancromático	15
9	Cirrus	30
10	Infrarroja térmico 1 (TIRS 1)	100
11	Infrarroja térmico 2 (TIRS 2)	100

FUENTE: Earth Observing System [37].

2.1.9 Clasificación de imágenes multiespectrales.

La clasificación de imágenes multiespectrales es un proceso en el cual se extraen clases de información de una imagen ráster multibanda. El ráster resultante de la clasificación de imagen se puede utilizar para crear mapas temáticos [38]. En la clasificación de imágenes se utilizan métodos de clasificación supervisada y no supervisada [39].

2.1.9.1 Clasificación supervisada y no supervisada.

La clasificación supervisada utiliza firmas espectrales obtenidas de las regiones de interés creadas para clasificar una imagen; dichas regiones de interés son creadas en función de las coberturas que el usuario desea clasificar y representar en un mapa temático. Por otro lado, la clasificación no supervisada busca clases espectrales (o clústeres) en una imagen multibanda sin la intervención del analista, es decir, sin creación de regiones de interés por el usuario [38]. En el presente trabajo se implementaron dos clasificadores supervisados, los cuales fueron Maximum Likelihood y Random Forest, como ya se especificó en apartados anteriores.

2.1.9.2 Método de Máxima Probabilidad (ML).

El método de máxima probabilidad o Maximum Likelihood en inglés, es uno de los tradicionales utilizados en el área de la teledetección. En este método las matrices de covarianza y vector de media de clase se utilizan para minimizar la distancia de clase y maximizar la probabilidad de que una característica pertenezca a la clase seleccionada mediante el uso de funciones de discriminación cuadráticas o lineales. Este método, sin embargo, puede sobre ajustarse a las probabilidades posteriores y dar como resultado una clasificación inexacta cuando se violan los supuestos [40].

2.1.9.3 Árboles de decisión (AD).

Los AD se representan mediante un gráfico con estructura arbórea que ha sido inducida desde un conjunto de datos de entrenamiento [41]. Todo AD comienza con un nodo al que pertenecen todos los casos de la muestra que se quiere clasificar, a este nodo se le

denomina nodo raíz (además hay nodos internos y terminales). Los nodos internos se corresponden con una condición elegida en el conjunto de atributos y cada ramificación de un nodo interno representa un valor del dominio de la condición del atributo. Los nodos terminales u hojas contienen una única etiqueta, que corresponde, en este caso, a la cubierta del suelo. Desde la raíz del árbol hasta una determinada hoja, el camino conforma una regla de clasificación [41].

2.1.10 Índices espectrales.

Un índice espectral es una técnica de realce de imágenes, que a través de fórmulas matemáticas, en donde las variables son los valores de una o más bandas espectrales en cada píxel, permite representar (de forma llamativa) una condición de interés en el terreno [42]. Por lo tanto, se utilizan índices espectrales para representar e identificar de forma particular las condiciones u objetos sobre la superficie terrestre que los analistas intentar estudiar [42].

2.1.10.1 Índice NDVI (Normalized Difference Vegetation Index).

El NDVI proporciona una métrica objetiva para detectar los niveles de vegetación saludable, ya que su ecuación permite identificar el cociente resultante entre la diferencia de las intensidades de la luz reflejadas por la banda roja e infrarroja y la suma de las mismas [43]. Los datos del NDVI se recopilan utilizando datos de teledetección que están disponibles públicamente y permiten comparar estudios en diferentes regiones geográficas [44]. Además, gracias a la características de este índice, es posible reflejar directamente el efecto de interacción de múltiples factores, incluido el uso de la tierra y las características climáticas [45], entre otros.

2.1.10.2 Índice GNDVI (Green Normalized Vegetation Index).

El índice de vegetación de diferencia normalizada verde (GNDVI, por sus siglas en inglés) es un índice de vegetación que utiliza las bandas verde e infrarroja cercana, y que se emplea para realizar estimaciones de la actividad fotosintética, determinando el consumo de agua y nitrógeno de la cubierta vegetal [46]. La principal diferencia con el NDVI es

que, el índice GNDVI es más sensible a la variación de la clorofila en el cultivo que el índice NDVI y presenta un punto de saturación más alto; además, GNDVI puede ser utilizado en cultivos en etapas avanzadas de desarrollo, mientras que el índice NDVI es más idóneo para cultivos en etapas iniciales de crecimiento [47].

2.1.10.3 Índice GCI (Green Coverage Index).

El índice de clorofila GCI (Green Coverage Index) es un nuevo indicador de vegetación, que, junto a otros afines, permite evaluar el estado sanitario de la vegetación y realzar únicamente, aquellos elementos formados por estructura vegetal. El GCI se puede calcular en entornos como ArcGIS, QGIS, gvSIG, SNAP, ENVI o LEOworks y las bandas necesarias para el cálculo son la infrarroja cercana y la banda verde [48].

2.1.10.4 Índice MSI (Moisture Stress Index).

El índice de estrés por humedad (MSI: Moisture Stress Index en inglés) es una técnica de teledetección que sirve para estimar la humedad del suelo a diferentes profundidades en sitios distantes o no monitoreados de forma física por el ser humano [49]. Además, puede usarse para determinar rápidamente la humedad del suelo para regiones grandes y fuera de alcance. Por lo tanto, el presente índice espectral es de suma importancia en los estudios de teledetección a nivel mundial; las bandas que participan en el cálculo son la infrarroja media y la infrarroja cercana [49].

2.2 Marco referencial.

Cada vez son más frecuentes los estudios de clasificación supervisada que utilizan distintos algoritmos e imágenes de diversos satélites. Estas investigaciones tratan de definir la mejor vía para clasificar distintos objetos de estudio de acuerdo con los intereses de cada región. En el presente marco referencial se evidencian algunas investigaciones que han medido el nivel de precisión de los algoritmos utilizados en este proyecto, a partir de distintas imágenes satelitales (incluidas Landsat-8 y Sentinel-2) y con intenciones similares de clasificación.

2.2.1 Mapeo de la cobertura vegetal en la subcuenca Quillcay (Ancash, Perú) con el clasificador de Árbol de decisiones.

Santiago, Mallqui y Ríos (2021) publicaron el artículo titulado *Mapeo de la cobertura vegetal en la subcuenca Quillcay (Ancash, Perú) con el clasificador de árbol de decisiones* [50]. Como su nombre lo indica, los investigadores delimitaron la cobertura vegetal de la subcuenca en estudio, empleando el método de clasificación supervisada de árbol de decisiones. El proceso de delimitación constó de varias etapas, que van desde la revisión bibliográfica hasta la validación de resultados [50]

Inicialmente, a través de revisión de fuentes secundarias se identificaron los tipos de coberturas vegetales que existen en el área de la subcuenca, para después obtener la imagen satelital de Landsat-8 del sitio web USGS [50]. Una vez obtenida la imagen, se le realizaron correcciones radiométricas y atmosféricas, para pasar a definir las variables geográficas (altitud y pendientes). Finalmente, se calculó el índice de Humedad de Diferencia Normalizada (NDMI), se obtuvieron las firmas espectrales y se realizó la clasificación por el método Spectral Angle Mapper (con toma de puntos en campo), concluyendo con la delimitación mediante árboles de decisión y la respectiva validación de resultados [50].

En la subcuenca objeto de estudio se identificaron las coberturas predominantes de: bofedal (190.71 ha), matorral (2295 ha), pajonal (4405.86 ha) y bosque (409.32 ha), las mismas que se desarrollan en altitudes entre 3200 – 5600 msnm, en pendientes mayores a 8.5% [50]. El árbol de decisiones fue creado con 14 nodos binarios, los cuales incluyeron características geográficas y espectrales, discriminando píxeles de áreas con escasa o sin vegetación; obteniendo una validación de resultados exitosa, con una precisión general del 90% en la matriz de confusión y 0.8998 en el coeficiente de Kappa [50].

2.2.2 Una evaluación de los datos de Landsat, Sentinel-2, Sentinel-1 y MODIS para el mapeo de tipos de cultivos.

Con respecto a estudios que comparan utilidad de imágenes de distintos satélites, Peng Song, Huang, Hansen y Potapov (2021) realizaron el artículo titulado *An evaluation of*

Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping [51], en el cual se evaluó la utilidad y la precisión potencial de Landsat-8, Landsat-7, Sentinel-2, Sentinel-1 y MODIS para la representación cartográfica de los cultivos de maíz y soya en Estados Unidos. Para el diseño de la investigación, se ejecutó una serie de experimentos de clasificación, utilizando las imágenes satelitales adquiridas y datos del Departamento de Agricultura de los Estados Unidos [51].

El área de estudio se dividió en bloques iguales de 20 km², y se clasificaron de acuerdo con la intensidad promedio de cultivos. Una vez ejecutado el preprocesamiento, se diseñaron experimentos (cuatro categorías) para evaluar la precisión de clasificación, así como la importancia relativa de cada sensor, banda espectral y fecha de generación de las imágenes [51]. Se utilizó el clasificador de aprendizaje automático “árboles de decisión”, el cual contribuyó con dos resultados: estimación de precisión de la clasificación suponiendo áreas de entrenamiento suficientes y correctas, y la selección y clasificación interpretando la estructura del árbol resultante [51].

Entre los resultados relevantes de la investigación, se destaca que, la precisión general de todos los sensores satelitales fue del 90% tanto para la representación de soja y de maíz, mientras que solo con Landsat se obtuvo una precisión de 96.8% para maíz y para soja, y con Sentinel-2 una precisión de 96.6% [51]. No obstante, el equipo investigador infiere que los datos actuales productos de teledetección con resolución moderada, pueden lograr una precisión nominal de hasta 95% para cultivos de expansión en escala nacional, que ocupen grandes extensiones de área [51].

2.2.3 Comparación de algoritmos de clasificación para detectar recuperación costera típica en la provincia Guangdong con imágenes de Landsat-8 y Sentinel-2.

Como ya se indicó en el inicio del presente apartado, la comparación de algoritmos de clasificación es bastante frecuente y necesaria. Por ejemplo, Bin Ai, Ke Huang, Shaojie Sun, Zhuokai Jian y Xiaoding Liu (2022) publicaron el artículo titulado *Comparison of Classification Algorithms for Detecting Typical Coastal Reclamation in Guangdong Province with Landsat-8 and Sentinel 2 Images* [52]; se plantearon como objetivo

encontrar el método apropiado para representar a gran escala la recuperación costera típica en la provincia China de Guangdong, para lo cual se utilizaron métodos de clasificación, entre los cuales figura el Árbol de decisiones [52].

El proceso inició adquiriendo los datos satelitales de Landsat-8 y Sentinel-2 sin presencia de nubes sobre el área, para después realizar la corrección radiométrica, la corrección atmosférica para la imagen de Landsat y correcciones geométricas y mosaicos continuos para todas las imágenes [52]. Los algoritmos de clasificación fueron bosque aleatorio, máquina de vectores de soporte, árbol de decisión y métodos orientados a objetos basados en reglas. El algoritmo óptimo para representar geográficamente la recuperación de la zona costera objeto de estudio se determinó mediante un análisis estadístico (validación de resultados con matriz de confusión y McNemar's Test) [52].

Entre los resultados destacados está que la precisión general varió de 83.05 % a 88.57 cuando los algoritmos se aplicaron a las imágenes de Sentinel-2 para la clasificación del oeste de la zona costera; por otro lado, la clasificación de la zona costera del este varió de 86.47% a 88.50% [52]. El mejor clasificador para las imágenes de Sentinel-2 fue el orientado a objetos con un coeficiente de Kappa de 0.827. Con respecto a las imágenes de Landsat-8, se obtuvieron valores similares, siendo el mejor clasificador el orientado a objetos en la clasificación del oeste (86.39%); mientras que, la clasificación con el método de Árbol de decisiones fue de 82.62% (oeste) y 81.36% (Delta del río y este de la zona costera) [52].

2.2.4 Deep learning para la clasificación de usos de suelo agrícola con Sentinel-2.

En otro estudio comparativo de algoritmos de clasificación, Campos –Taberner, García-Haro, Martínez y Gilabert (2020) publicaron un artículo titulado *Deep learning para la clasificación de usos de suelo agrícola con Sentinel-2* [53], en el cual se comparó el rendimiento de un método de inteligencia artificial denominado 2-BiLSTM versus árboles de decisión, los k-vecinos más cercanos, redes neuronales, máquinas de soporte vectorial y bosques aleatorios, para la clasificación de usos de suelo agrícola a partir de imágenes del satélite Sentinel-2 [53].

El estudio se llevó a cabo en la Comunitat Valenciana, España. Se descargaron dos escenas de Sentinel-2, las mismas que abarcan conjuntamente $220 \times 220 \text{ km}^2$. Se seleccionaron 16 usos de suelo no urbano para realizar la comparación entre los métodos clasificadores previamente mencionados. Para la validación de los algoritmos de clasificación se utilizó la matriz de confusión y un índice de precisión derivado de la misma [53].

Entre los resultados se destaca que, el algoritmo más preciso fue el de 2-BiLSTM con 98.6%, seguido de los algoritmos 1-BiLSTM, 3-BiLSTM, 4-BiLSTM y bosques aleatorios con 97.3%, 95.1%, 94.7% y 94.3%, respectivamente. En un escalón más abajo, se encuentran los algoritmos de máquinas de soporte vectorial, redes neuronales, k-vecinos más cercanos y árboles de decisión con 92.3%, 92.2%, 88.1% y 85.3%, respectivamente [53]. La mayor precisión con el mejor algoritmo (el 2-BiLSTM) se obtuvo sobre la clase tierra con un 99.9% de identificación precisa, mientras que, el error más grande fue para la clase frutales con precisiones de usuario y productor del 91.1% y 92.6%, respectivamente. Una vez seleccionado el mejor método clasificador, se lo utilizó para la generación de un mapa de usos del suelo y un mapa de confianza de clasificación [53].

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

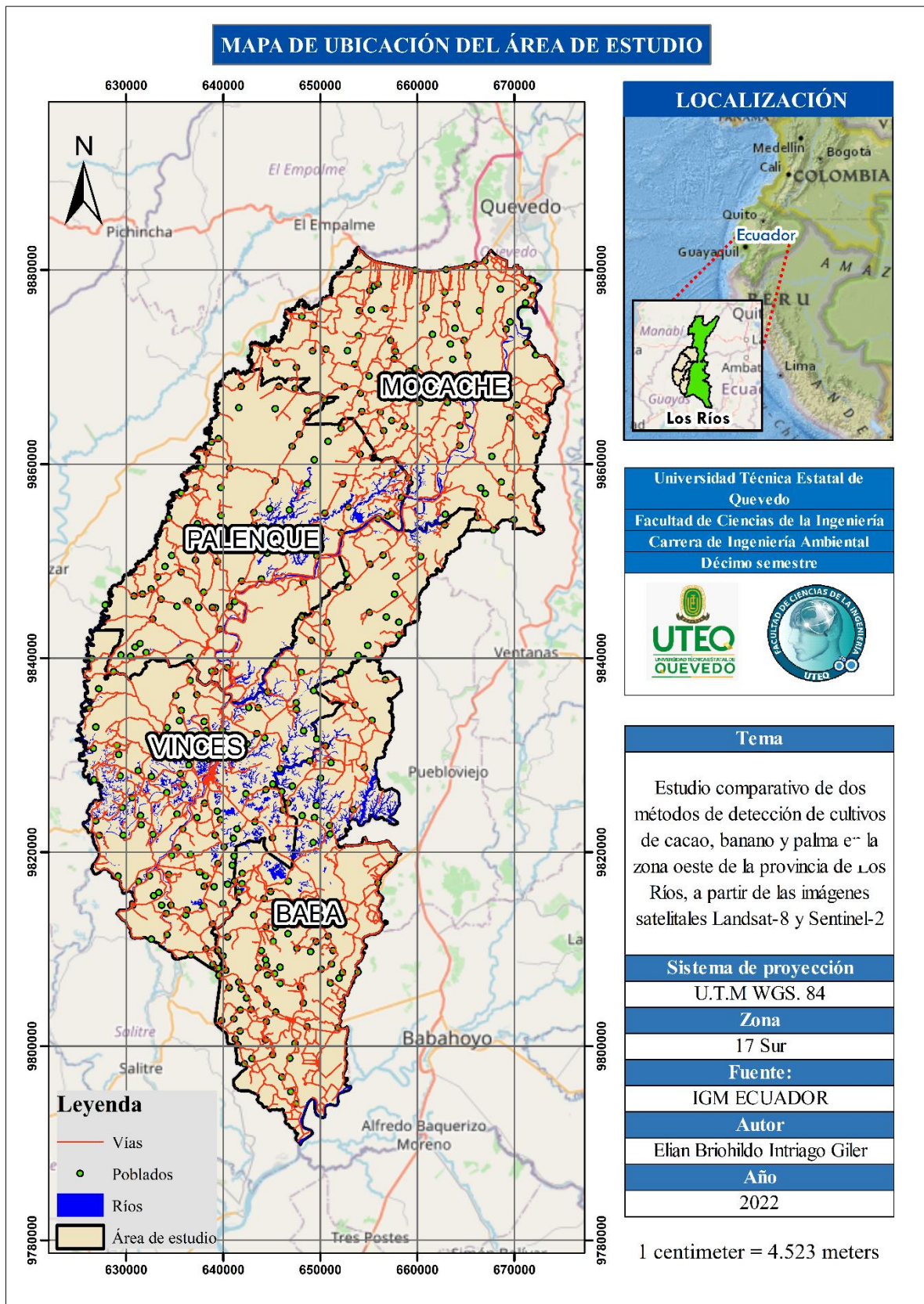
3.1 Localización.

El estudio se ejecutó en cuatro cantones de la zona oeste de la provincia Los Ríos, Ecuador. Los cantones son Baba, Vinces, Mocache y Palenque. El área de estudio limita al norte con los cantones Quevedo y El Empalme, al sur con Babahoyo y Salitre, al este con Ventanas y Pueblo Viejo, y al oeste con los cantones Balzar, Colimes y Palestina (Figura 2).

Los Ríos es una de las zonas más agro-productivas del país, donde la producción agrícola se ha convertido en un pilar fundamental del PIB. La investigación se centró en la detección de cultivos de banana, palma y cacao, comparando la efectividad de los métodos clasificadores, así como el nivel de utilidad de las imágenes satelitales de Landsat-8 y Sentinel-2.

Figura 2

Mapa de ubicación del área de estudio: Cantones Mocache, Palenque, Vinces y Baba.



ELABORADO: AUTOR

3.2 Tipo de investigación.

La investigación desarrollada es de tipo exploratoria, cuantitativa y transversal.

- **Exploratoria:** El proyecto de investigación es de tipo exploratorio, debido a que aún no se había estudiado la precisión del método de inteligencia artificial Random Forest a partir de imágenes satelitales de Landsat-8 y Sentinel-2 en el área de estudio, ni mucho menos ejecutada una comparación de métodos para la clasificación supervisada de imágenes multiespectrales.
- **Cuantitativa:** El estudio es de tipo cuantitativo, ya que en la validación de resultados se realizó una matriz de confusión y se obtuvo el Coeficiente de Kappa para cada clasificación supervisada ejecutada. Se seleccionó el mejor método clasificador que permitió obtener la información objetiva sobre la extensión de los cultivos de cacao, banano y palma en el área de estudio.
- **Transversal:** La investigación es de tipo transversal, ya que comparó la precisión del método de inteligencia artificial Random Forest y el método tradicional Maximum Likelihood, en la detección de cultivos de cacao, banano y palma en el área de estudio a partir de las imágenes satelitales de Landsat-8 y Sentinel-2

3.3 Métodos de investigación.

El proyecto de investigación utilizó el siguiente método:

- **Método inductivo:** Este método se utilizó en la investigación, dado que, para determinar el mejor clasificador para las imágenes multiespectrales, se ejecutó un proceso desde lo particular (procesamiento de las imágenes, cálculo de índices espectrales y validación de resultados) a lo general (determinación del mejor método clasificador a partir de una imagen multiespectral).

3.4 Fuentes de recopilación de información.

La información para el proyecto de investigación se obtuvo de las siguientes fuentes:

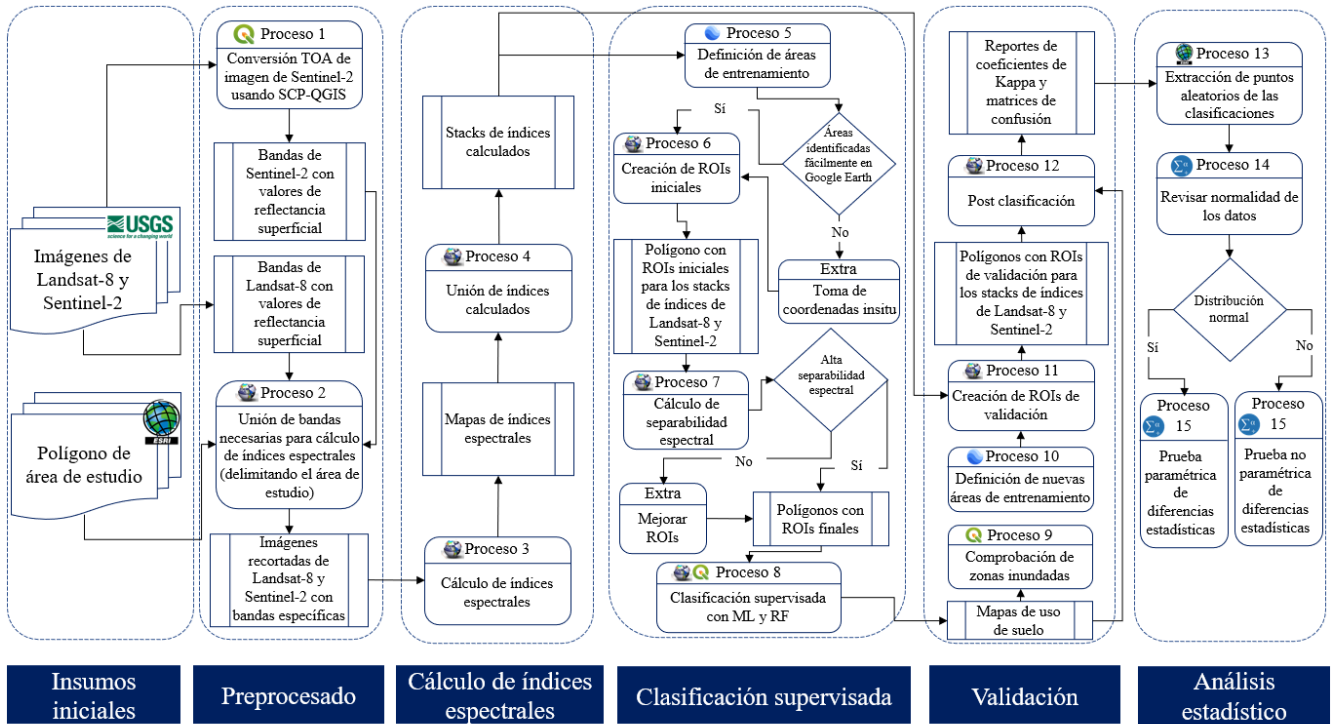
- **Fuentes primarias:** Para la definición de áreas de entrenamiento, previo a la ejecución de la clasificación supervisada con los métodos ya mencionados, se tomaron coordenadas *in situ* (puntos referenciales) para el cultivo de cacao; de la misma forma, se tomaron coordenadas desde Google Earth Pro para las demás coberturas de interés. En el mismo sentido, se tomaron coordenadas *in situ* y desde Google Earth Pro para la validación de resultados con las matrices de confusión y coeficientes de kappa.
- **Fuentes secundarias:** La obtención de información secundaria constituye la base de la investigación. Se descargaron las imágenes multiespectrales de Landsat-8 y Sentinel-2 desde el sitio web de la USGS, las mismas que sirvieron para el cálculo de los índices espectrales, la clasificación supervisada con los métodos previamente descritos y su cálculo de precisión (validación de resultados).

3.5 Diseño de la investigación.

En la figura 3 se observa todo el proceso metodológico para la detección de los cultivos objetivo a partir de imágenes de Landsat-8 y Sentinel-2.

Figura 3

Flujograma de la investigación.



ELABORADO: AUTOR

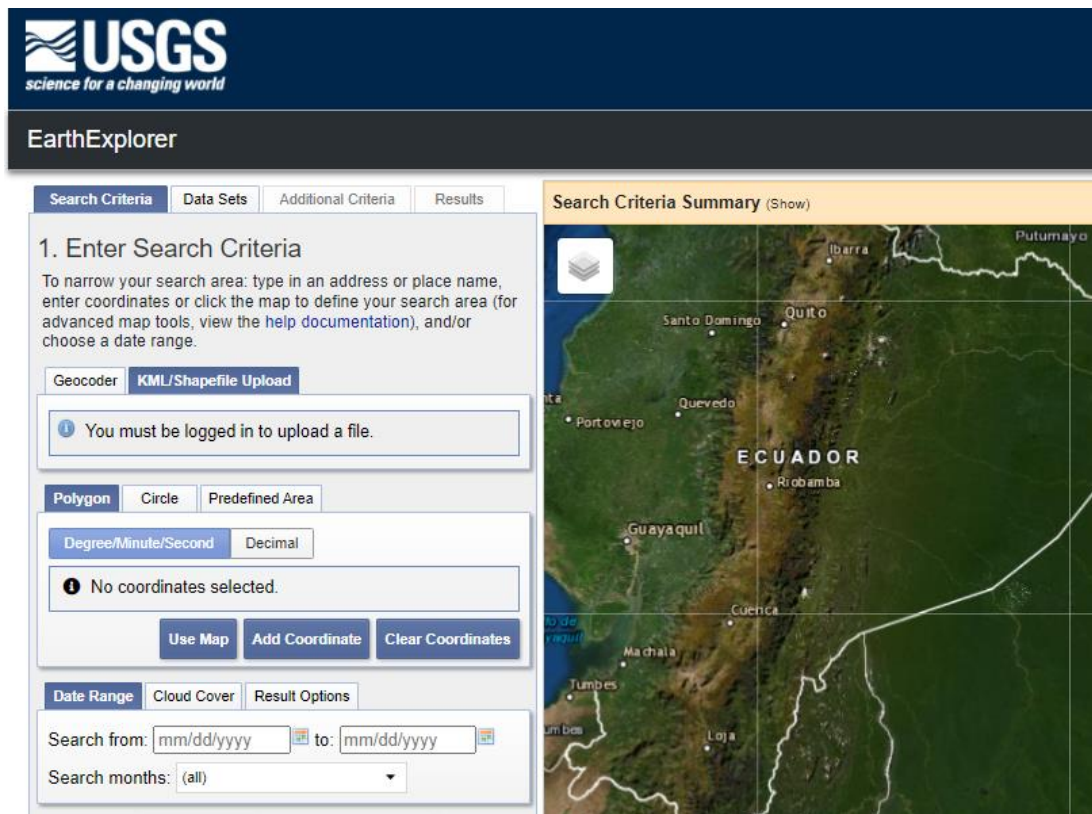
3.5.1 Objetivo 1: Preprocesar las imágenes multispectrales de Landsat-8 y Sentinel-2.

3.5.1.1 Búsqueda de imágenes.

Para la búsqueda de imágenes satelitales de Landsat-8 y Sentinel-2 se utilizó “USGS Earth Explorer”. Este portal permite adquirir datos geospaciales de múltiples satélites que orbitan el planeta Tierra. La interfaz del sitio web se muestra en la figura 4.

Figura 4

Interfaz de sitio web de la USGS.



FUENTE: USGS Science for a changing world [54].

Para el proceso de búsqueda se utilizó la opción “KML/Shapefile” dentro la ventana “Search Criteria”. En el apartado de “KML/Shapefile” se agregaron las coordenadas y el intervalo temporal de búsqueda. Posteriormente, se seleccionó la opción “Data sets” y se especificaron los satélites proveedores de las imágenes multiespectrales (Landsat-8 y Sentinel-2). Para finalizar, se buscó y se descargó una imagen multiespectral despejada (sin presencia de nubes sobre el área de estudio) en formato multibanda por cada satélite.

3.5.1.2 Corrección atmosférica de las bandas espectrales de la imagen de Sentinel-2.

A la imagen satelital de Sentinel-2 se le realizó un procedimiento denominado corrección atmosférica. La corrección atmosférica se realizó debido a que la imagen disponible para el área de estudio se encuentra solo en nivel de procesado “1C”, y no en “2A” (imágenes disponibles corregidas atmosféricamente); este proceso permite mejorar la precisión de los

clasificadores [34]. Este procedimiento no fue necesario para la imagen de Landsat-8, ya que las imágenes de la Colección 1 nivel 2 (On – Demand) del programa satelital, se pueden descargar (bajo pedido por correo electrónico) con valores de reflectancia superficial (corregidas atmosféricamente).

La corrección atmosférica para la imagen de Sentinel-2 se realizó en el software QGIS 3.22.7 con el complemento Semi Automatic Classification Plugin (SCP) [55] y utilizado por otros autores [56]–[59]. Posterior a la instalación del complemento, se siguió esta ruta: “SCP” > “Preprocesamiento” > “Sentinel-2” (Figura 5). Toda vez en el apartado, se seleccionó la carpeta en donde se encontraban todas las bandas espectrales, y el archivo “MTD_MSI”, el cual contuvo la información de todas las bandas de la imagen satelital. Finalmente, se marcó con un visto la opción que indicaba “Aplicar la corrección atmosférica”.

Figura 5

Corrección atmosférica de bandas espectrales de Sentinel-2 con Semi Automatic Classification Plugin del año 2021.

Conversion de Sentinel-2

Directorio conteniendo bandas Sentinel-2: C:/Users/Personal/Desktop/L1C_T17MPU_A005814_20180417T154048/S2B_MSIL1C_20180417T153619_N0206_R068_T17MPU_20180417T2022

Select metadata file (MTD_MSI): C:/Users/Personal/Desktop/L1C_T17MPU_A005814_20180417T154048/S2B_MSIL1C_20180417T153619_N0206_R068_T17MPU_20180417T2022

☒ Aplicar la corrección atmosférica DOS1 ☒ Use value as NoData 0

☒ Preprocess bands 1, 9, 10

☒ Crear Juego de Bandas y utilizar sus herramientas ☒ Add bands in a new Band set

Metadatos

Satélite: Sentinel-2B Date (YYYY-MM-DD): 2018-04-17 Product: S2MSI1C

	Banda	Valor de Cuantificación	Irradiancia solar
1	T17MPU_20180417T153619_B01.jp2	10000	1874.3
2	T17MPU_20180417T153619_B02.jp2	10000	1959.77
3	T17MPU_20180417T153619_B03.jp2	10000	1824.93
4	T17MPU_20180417T153619_B04.jp2	10000	1512.79
5	T17MPU_20180417T153619_B05.jp2	10000	1425.78
6	T17MPU_20180417T153619_B06.jp2	10000	1291.13
7	T17MPU_20180417T153619_B07.jp2	10000	1175.57
8	T17MPU_20180417T153619_B08.jp2	10000	1041.28
9	T17MPU_20180417T153619_B09.jp2	10000	817.58
10	T17MPU_20180417T153619_B10.jp2	10000	365.41
11	T17MPU_20180417T153619_B11.jp2	10000	247.08
12	T17MPU_20180417T153619_B12.jp2	10000	87.75
13	T17MPU_20180417T153619_B8A.jp2	10000	953.93

Editar metadatos

ELABORADO: AUTOR

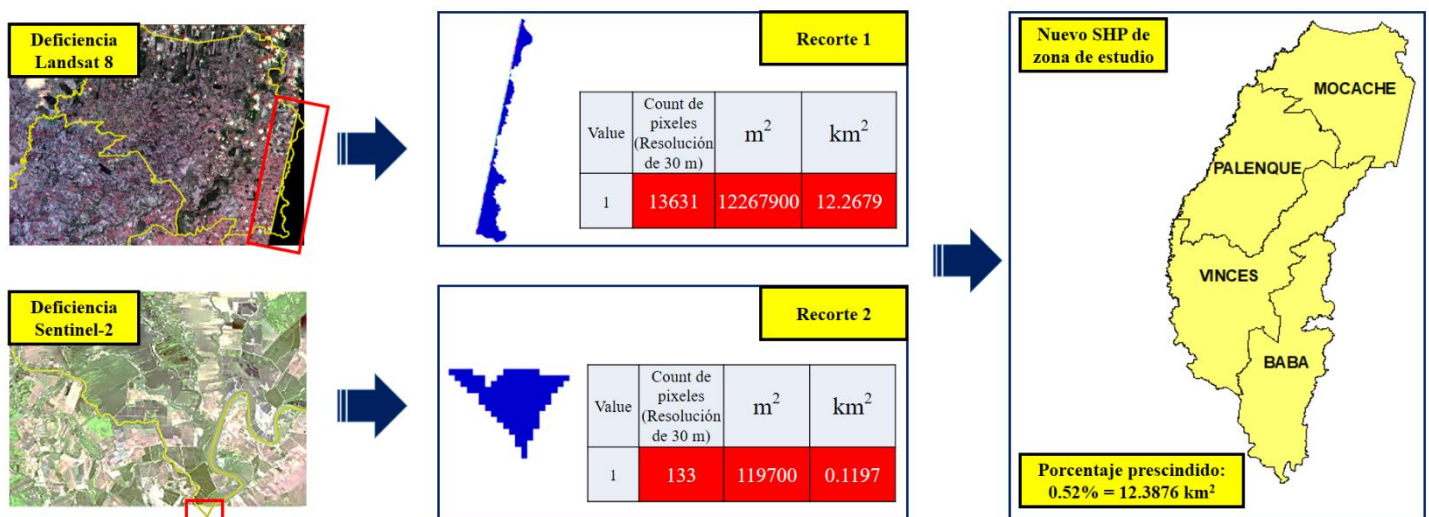
3.5.1.3 Extracción de polígono real del área de estudio y recorte adicional.

Inicialmente se extrajo un polígono (archivo SHP) que incorpore los cantones Baba, Palenque, Mocache y Vines (área de estudio). La extracción se realizó en el software

ArcMap, mediante la selección de dichos cantones en la tabla de atributos de la capa shape de cantones del Ecuador y exportando la selección como una nueva capa shape, utilizando la herramienta “Export Data” en el apartado de “Data”. No obstante, las imágenes de Landsat-8 y Sentinel-2 descargadas, no cubrieron toda la zona de estudio, por lo tanto, se modificó el polígono del área en función de la cobertura real de las imágenes satelitales. Se recortó de forma manual en el software ArcMap, con le herramienta “Cut Polygons”, prescindiendo de un total de 12.3874 km², es decir, menos del 1% del área de los cantones analizados (Figura 6). Esto se realizó con la finalidad de evitar errores en la eventual clasificación supervisada.

Figura 6

Polígono definitivo del sitio de estudio en función de las coberturas reales de las imágenes de Landsat-8 y Sentinel-2.



ELABORADO: AUTOR

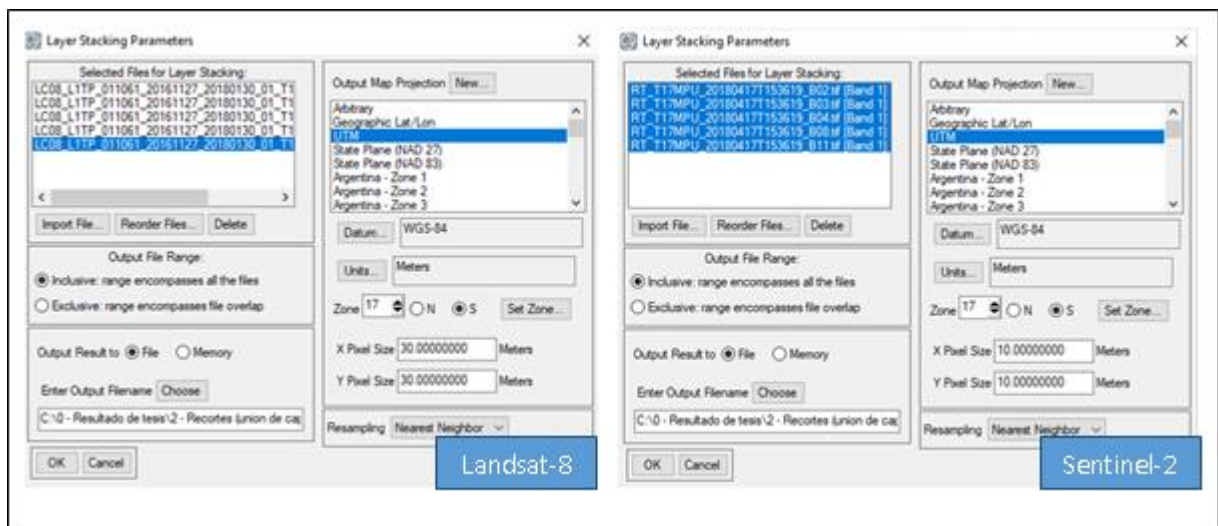
3.5.1.4 Unión de bandas espectrales.

Una vez comprobadas las imágenes en ArcMap tanto para Landsat-8 y Sentinel-2, y realizada la corrección atmosférica para las bandas espectrales de la imagen satelital de Sentinel-2, se unieron las bandas necesarias por cada satélite. Las bandas que se agruparon en una sola imagen multiespectral (por cada satélite) fueron aquellas necesarias para el cálculo de los índices espectrales y para presentar una combinación RGB (espectro visible). Para la unión se utilizó la herramienta “Layer Stacking” la cual se encuentra

dentro de la caja de herramientas del software ENVI 5.3. Dentro de la opción mencionada, se escogieron las bandas necesarias y se realizó la unión solo para la extensión del polígono del sitio de estudio (delimitando la extensión de la unión dentro de la misma herramienta) (Figura 7).

Figura 7

Configuración de unión de bandas espectrales necesarias en función del polígono de la zona de estudio.



ELABORADO: AUTOR

3.5.2 Objetivo 2: Calcular los índices multispectrales NDVI, GNDVI, GCI y MSI a partir de las imágenes multispectrales preprocesadas.

3.5.2.1 Cálculo del índice NDVI.

Para el cálculo del NDVI se utilizó la ecuación 1, como se expresa en la investigación “Unsupervised Learning-based Vegetation change detection with Landsat-8 data” [43]. En la ecuación NIR representa información de reflexión en la banda del infrarrojo cercano y RED representa información de reflexión sobre la banda roja de las imágenes satelitales [43].

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad \text{Ecuación 1. } NDVI$$

3.5.2.2 Cálculo del índice GNDVI.

Con la finalidad de calcular el índice espectral GNDVI se utilizó la ecuación 2, como se expresa en la investigación “FusionNDVI: A Computational Fusion Approach for High-Resolution Normalized Difference Vegetation Index” [60]. En la ecuación NIR representa la banda infrarrojo cercano y G representa la banda verde de las imágenes satelitales [60].

$$GNDVI = \frac{NIR-G}{NIR+G} \quad \text{Ecuación 2. } GNDVI$$

3.5.2.3 Cálculo del índice GCI.

En el cálculo del Green Coverage Index (GCI) se utilizó la Ecuación 3 [61]. El índice de clorofila verde se utilizó para estimar el contenido de clorofila de las hojas en varias especies de plantas. El contenido de clorofila refleja el estado fisiológico de la vegetación; disminuye en plantas estresadas y, por tanto, puede utilizarse como medida de la salud de las plantas [62]. Las variables inmersas en la ecuación representan lo descrito en la ecuación 3.

$$GCI = \frac{NIR}{G} - 1 \quad \text{Ecuación 3. } GCI$$

3.5.2.4 Cálculo del índice MSI.

Para el cálculo del índice MSI se utilizó la ecuación 4, como se expresa en la investigación “Estimation of Soil Moisture Percentage Using LANDSAT-based Moisture Stress Index” [49]. En la ecuación del MSI, MidIR representa la banda del infrarrojo medio y NIR la banda del infrarrojo cercano de las imágenes satelitales adquiridas [49].

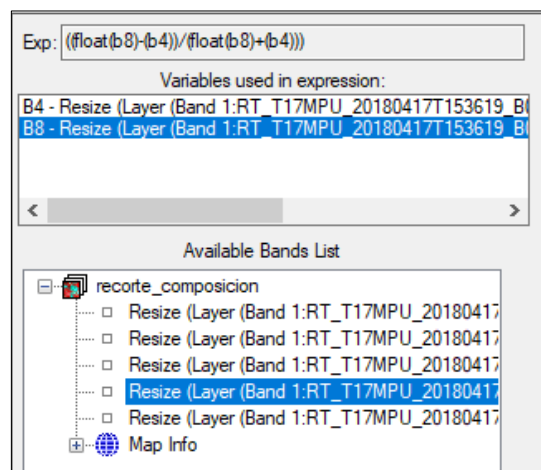
$$MSI = \frac{MidIR}{NIR} \quad \text{Ecuación 4. } MSI$$

3.5.2.5 Proceso de cálculo en software.

Para el cálculo de los índices multispectrales se utilizó el software ENVI, empleando la herramienta Band Math. En la ventana emergente de la herramienta mencionada, se escribieron las ecuaciones 1, 2, 3 y 4, y se le asignó la banda espectral necesaria que solicita la ecuación pertinente en función la imagen satelital. Un ejemplo del proceso se muestra en la figura 8.

Figura 8

Ejemplo de cálculo de índice espectral para la imagen Sentinel-2.



ELABORADO: AUTOR

3.5.3 Objetivo 3: Ejecutar la clasificación supervisada de imágenes con los métodos Maximum Likelihood y Random Forest.

3.5.3.1 Creación de áreas de entrenamiento.

La clasificación del conjunto de la imagen es el proceso por el que a cada elemento contenido en la imagen se le asigna una categoría, basándose en los atributos contenidos en las áreas de entrenamiento [63]. Como paso previo a la clasificación supervisada con los métodos Maximum Likelihood y Random Forest, se definieron las áreas de entrenamiento de las coberturas de interés a clasificar.

Para la definición de las áreas de entrenamiento se tomaron coordenadas UTM (Universal Transversal de Mercator) mediante el software Google Earth Pro, en función de las coberturas de interés a clasificar. No obstante, para las categorías más complejas de identificar mediante Google Earth Pro se realizó la toma de coordenadas en campo (*in situ*), con la finalidad de mejorar la separabilidad espectral de los ROIs (Regiones de interés por su siglas en inglés) a crear en función de las áreas de entrenamiento. El formato de la tabla y las categorías de interés se muestran en la tabla 3.

Tabla 3

Modelo de registro de coordenadas para definición de áreas de entrenamiento

Categorías	Registro de coordenadas UTM		Procedencia	
	X	Y	Google Earth	Toma en Campo
Zona con asentamientos humanos (ZAH)				
Zonas con cultivos de banano (ZCB)				
Zonas con cultivos de cacao (ZCC)				
Zonas con cultivos de palma (ZCP)				
Zonas con cultivos de cuerpos de agua (ZCA)				
Zonas sin vegetación aparente (ZVA)				
Zonas con otras coberturas (ZOC)				
Nubes (N)				

ELABORADO: AUTOR

3.5.3.2 Regiones de interés (ROIs).

Posterior a la definición de las áreas de entrenamiento, se procedió a utilizar dichas coordenadas para definir las regiones de interés para cada stack de índices calculados. La definición se realizó en el software ENVI 5.3 y permitió crear las categorías predefinidas. Para el stack de Landsat-8, se definieron 65 ROIs de entrenamiento, repartidos en 22 para ZCC, 11 para ZOC, 10 para ZCP, 8 para ZCB, 4 para ZCA y N y 3 para ZVA y ZAH, representando un total de 4822 píxeles de 900 m² (4.4 km²). Por otro lado, para el stack de Sentinel-2, se definieron 59 ROIs de entrenamiento, repartidos en 22 para ZOC, 10 para ZCP, 7 para ZCC, 6 para ZCB, 5 para ZAH y ZVA y 4 para ZCA, representando un total de 18350 píxeles de 100 m² (1.84 km²).

Las regiones de interés de todas las coberturas fueron diferenciadas por colores. Finalmente, se calculó la separabilidad espectral de Jeffries-Matusita, Transformed Divergence [64]; esto permitió verificar si existió o no una relación/similitud espectral entre categorías [65] y mejorar la calidad de los ROIs creados en caso de ser necesario. Finalmente, se exportaron dos archivos shape, que contuvieron todos los ROIs para cada stack de índices calculados.

Este mismo proceso fue aplicado para la creación de ROIs de validación de resultados (ver 3.5.4). Para el stack de Landsat-8 se crearon 21 ROIs de validación, repartidos en 7 para ZCO, 3 para N y ZVA, 2 para ZCP, ZCC y ZAH y 1 para ZCB y ZCA, representando un total de 3256 píxeles de 900 m² (2.9 km²). Para el stack de Sentinel-2 se crearon 26 ROIs de validación, repartidos en 11 para ZCO, 4 para ZCC, 3 para ZCP y ZCA, 2 para ZVA y ZAH y 1 para ZCB, representando un total de 14960 píxeles de 100 m² (1.5 km²).

3.5.3.3 Aplicación del clasificador Maximum Likelihood (ML).

Al finalizar la creación de los ROIs, se aplicó el clasificador tradicional de Máxima Probabilidad (ML) para los dos stacks de índices calculados, en el software ENVI 5.3, utilizando funciones discriminantes [66]. Este método asume que las estadísticas para cada clase (coberturas de interés a clasificar) en cada banda se distribuyen de forma normal y calcula la probabilidad de que un píxel dado pertenezca a una clase específica. De esta manera, cada píxel se asigna a la clase que tiene la probabilidad más alta; no obstante, si la probabilidad más alta es menor que un umbral especificado, el píxel permanece sin clasificar [66]. Se empleó este método para clasificar el stack de índices calculados a partir de la imagen de Landsat-8 (ML-L8) y para clasificar el de Sentinel-2 (ML-S2).

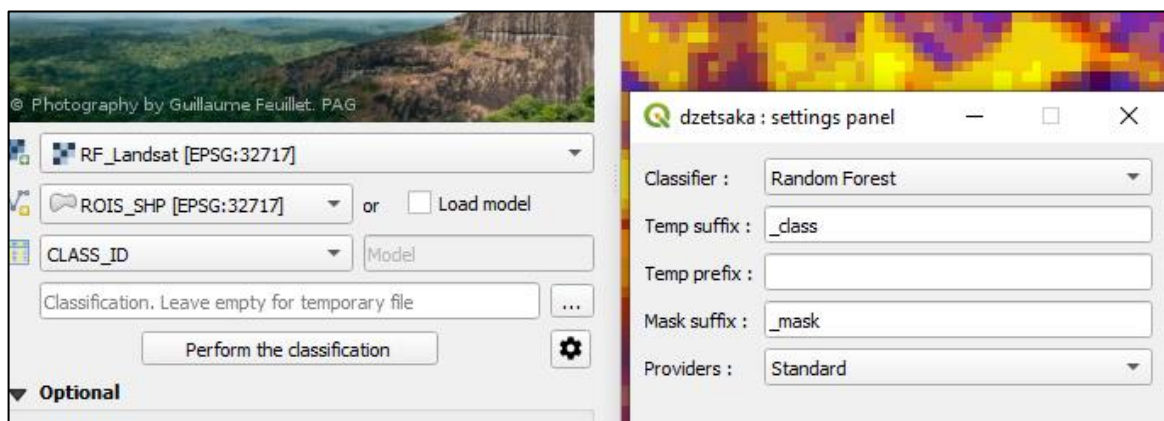
Para la ejecución de este método clasificador, se importó al área de trabajo del software ENVI, el stack de índices calculados de Landsat-8 y el stack de índices calculados de Sentinel-2. Posteriormente, se utilizó la herramienta “Maximum Likelihood Classification” dentro del apartado de “Classification” en el Toolbox del software. Se seleccionó el stack correspondiente a clasificar y las categorías creadas en las regiones de interés. No se estableció ningún umbral de probabilidad, de modo que los píxeles se asignaron a las clases de mayor probabilidad [40].

3.5.3.4 Aplicación del clasificador Random Forest (RF).

Se aplicó este clasificador no paramétrico de inteligencia artificial para el stack de índices calculados a partir de la imagen de Landsat-8 (RF-L8) y para clasificar el de Sentinel-2 (RF-S2), en el software QGIS. Para la ejecución de este algoritmo clasificador, se importaron los ROIs a la bandeja de capas del software, así como los stacks de índices calculados, tanto de Sentinel-2 como de Landsat-8. Posteriormente, se instaló el complemento “Dzetsaka” [67] citado en las metodologías de varios estudios [68]–[71]. Abierto el complemento, se aseguró que el stack de índices calculados y los ROIs, tengan la misma proyección cartográfica, además de, seleccionar el clasificador RF y la columna “class” para ejecutar la clasificación (Figura 9).

Figura 9.

Configuración para el algoritmo clasificador Random Forest con las herramientas de “Dzetsaka”.



ELABORADO: AUTOR

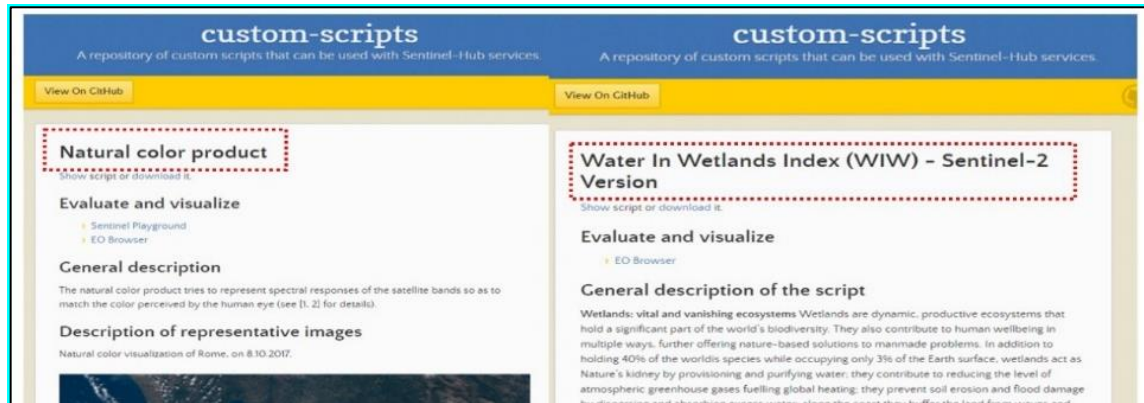
3.5.3.5 Comprobación de zonas inundadas.

Posterior a la clasificación supervisada, se comprobó la existencia de zonas inundadas en el área de estudio en la imagen de Sentinel-2. Para el proceso mencionado se utilizó el sitio web Sentinel Hub y el software QGIS. Inicialmente se creó una nueva configuración en Sentinel Hub, en la cual se agregaron dos capas de bandas combinadas de Sentinel-2 (Water in Wetlands Index y Natural Color). Para agregar las bandas combinadas se descargaron sus “scripts” desde un repositorio de scripts personalizados (ver Figura 10).

Toda vez obtenidos los scripts de las bandas combinadas, se agregaron como nuevas capas dentro de la configuración creada en Sentinel Hub.

Figura 10.

Scripts de “Water in Wetlands Index” y “Natural Color” para la comprobación de zonas inundadas en imágenes de Sentinel 2 a partir de QGIS y Sentinel Hub.

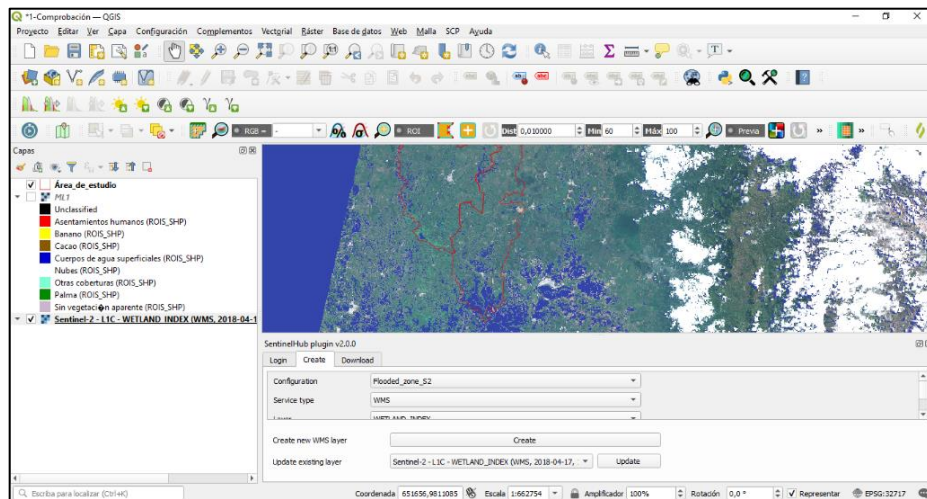


FUENTE: Sentinel Hub [72]

En el software QGIS se instaló el plugin “SentinelHub”; a partir de este plugin se enlazó la configuración creada en el sitio web de Sentinel Hub con QGIS. Toda vez enlazada la configuración, se procedió a visualizar las combinaciones de bandas de “Water in Wetlands Index” y “Natural Color” en la fecha de la imagen de Sentinel 2 (Tabla 6). A través de la visualización de las combinaciones de bandas y su respectiva comparación con la clasificación supervisada para la imagen de Sentinel-2, se logró comprobar la existencia de zonas inundadas en el área. En la figura 11 se observa la combinación de bandas del índice “Water in Wetlands Index” en la fecha de la imagen de Sentinel-2.

Figura 11

Combinación de bandas del índice “Water in Wetlands Index” del 17 de abril del 2018 en el área de estudio.



ELABORADO: AUTOR

3.5.4 Objetivo 4: Comprobar el nivel de precisión de los métodos utilizados en la clasificación supervisada.

Después de la clasificación supervisada, es importante evaluar el grado de precisión que obtuvieron los clasificadores. La evaluación del grado de precisión se realizó con la finalidad de validar las clasificaciones realizadas. Para este proceso se utilizaron los ROIs de validación creados y se obtuvieron coeficientes de Kappa y matrices de confusión, metodologías que se detallan a continuación.

3.5.4.1 Matriz de confusión.

Se utilizó la matriz de confusión, descrita en el estudio [73] y utilizada varios investigaciones de teledetección [63], [74]–[76]. La matriz de confusión, también llamada tabla de contingencias consta de una tabla de doble entrada, es decir que cada categoría (coberturas de interés a clasificar) representa una fila y una columna dentro de la tabla. El grado de precisión de la clasificación realizada se observa de acuerdo con la diagonal de la matriz (muestra la cantidad de píxeles que coinciden por categoría). Por otro lado, los valores que se encuentran en sentido vertical representan los píxeles que se coinciden con otras categorías (error de clasificación) [63]. Las matrices de confusión se realizaron en el

software ENVI 5.3 para cada clasificador en función de las imágenes satelitales. El modelo de la matriz de confusión se muestra en la tabla 4.

Tabla 4

Modelo matriz de confusión con adición del resultado del coeficiente de Kappa

Clasificación	Categoría 1	Categoría 2	Categoría 3	Total
Categoría 1				
Categoría 2				
Categoría 3				
Total				
Acuerdo/exactitud				
Error de omisión				
Error de comisión				
Precisión general				

FUENTE: Vera (2020). *Análisis de uso de suelo a partir de imágenes satelitales Sentinel-2 en el cantón Buena Fe, Provincia de Los Ríos* [65].

3.5.4.2 Índice de Kappa.

Como medida de precisión, se utilizó el coeficiente de Kappa propuesto por Jacob Cohen en 1960 [77] y utilizada en varias investigaciones [63], [75], [76]. El proceso de cálculo se realizó en el software ENVI, para las cuatro clasificaciones realizadas. Entre 0.01 y 0.20 representa una concordancia leve, entre 0.21 y 0.40 representa una concordancia aceptable, entre 0.41 y 0.60 una concordancia moderada, entre 0.61 y 0.80 considerable, y entre 0.81 y 1 una concordancia casi perfecta [77], [78].

3.5.5 Objetivo 5: Analizar los resultados obtenidos de la clasificación supervisada de imágenes multiespectrales de Landsat-8 y Sentinel-2.

Se ejecutó un análisis estadístico para comprobar si existieron diferencias significativas entre las clasificaciones de las imágenes de Landsat-8 y Sentinel-2 [79]. Inicialmente se extrajeron puntos al azar de las cuatro clasificaciones. El número de puntos extraídos fue proporcional al área de estudio (2348 km²) [80]. Por lo tanto, se tomaron 2348 puntos al azar por cada clasificación (1 punto por km²). Para la extracción de puntos al azar se utilizaron las herramientas “Create Random Points” y “Extract Multi Values to Point”.

Para el análisis estadístico mencionado se utilizó el software SPSS. Inicialmente se comprobó el tipo de distribución de los datos, la cual no cumplió con los parámetros de normalidad. Por lo tanto, se utilizó la prueba no paramétrica de Wilcoxon-Mann-Whitney para muestras independientes, también denominada “Suma de Rangos de Wilcoxon” [81].

En la prueba de Wilcoxon-Mann-Whitney se combinan los datos de dos muestras en una sola. No obstante, no calcula la diferencia entre las medianas y el cálculo de diferencias de las medianas no es intrínseco en esta prueba. Sin embargo, fue útil para la comparación estadística ya que la diferencia de dispersión de datos de un grupo con respecto al otro es la esencia de esta prueba; la misma que se establece a partir de un subconjunto de datos calculados para los dos grupos que se comparan [82].

3.6 Recursos humanos y materiales.

3.6.1 Materiales tecnológicos.

- GPS

3.6.2 Materiales de oficina.

- Lápiz
- Hojas de papel

3.6.3 Programas informáticos.

- ArcMap10.4.1
- ENVI 5.3
- QGIS 3.22.7
- Google Earth Pro 7.3.4.8248
- IBM SPSS Statistics 26

3.6.4 Sitios web.

- EarthExplorer
- Sentinel Hub

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 Preprocesamiento de imágenes multiespectrales.

4.1.1 Selección de imagen Landsat-8.

Para la selección de la imagen multiespectral de Landsat-8 de acuerdo con el área de estudio, se utilizó la colección 1 nivel 2 (On demand) del programa satelital. Se utilizó una imagen del 27 de noviembre del 2016. Varios datos relevantes sobre la imagen multiespectral se muestran en la tabla 5.

Tabla 5
Características de imagen de Landsat-8

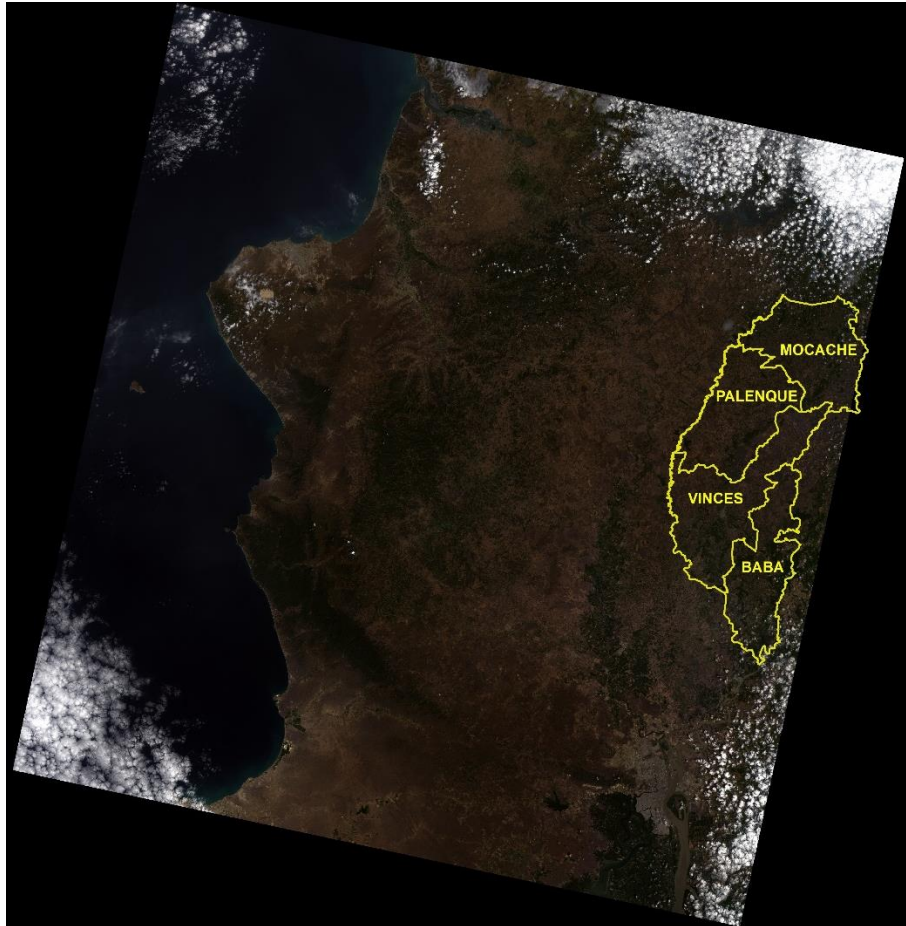
Imagen Landsat-8	
Resolución espacial (m)	15, 30 y 100
Número de bandas espectrales	11
Fecha	27-11-2016
Porcentaje de nubosidad	8.99%
Datos de colección y nivel	Colección 1 Nivel 2
DATUM	WGS84
Proyección de mapa	Universal Transversal de Mercator
Zona UTM	17 sur

FUENTE: USGS Science for a changing world [54]

Como se observa en la Tabla 5, la imagen de Landsat-8 adquirida cuenta con tres resoluciones espaciales; las bandas 1,2,3,4,5 y 6 tienen una resolución espacial es de 30 m, la banda 8 tiene una resolución de 15 m, la banda 9 una resolución de 30 m y las bandas 10 y 11 una resolución de 100 m. Por otro lado, el porcentaje de nubosidad de la imagen multiespectral es de 8.99%. La imagen adquirida con el área de estudio sobrepuesta se muestra en la figura 12.

Figura 12

Combinación RGB de imagen multiespectral adquirida de Landsat-8 con fecha 27/11/2016 más el área de estudio superpuesta.



FUENTE: USGS Science for a changing world [54]

A pesar de que la nubosidad de toda la imagen es de 8.99%, en el área ocupada por Baba, Vincés, Palenque y Mocache dicha cantidad es mucho menor. Es decir, el porcentaje real de nubes sobre el área de estudio es mucho menor, lo que aumenta las probabilidades de obtener una clasificación más precisa.

4.1.2 Selección de imagen Sentinel-2.

Con respecto a la imagen multiespectral de Sentinel-2, se seleccionó una imagen tomada por Sentinel-2B (satélite idéntico a Sentinel-2A) que forma parte del programa satelital. Al igual que con la imagen de Landsat-8, se consideró como prioridad la nula y/o escasa

presencia de nubes sobre el área de estudio. Los datos relevantes de la imagen multiespectral se muestran en la tabla 6 y en la figura 13.

Tabla 6
Características de imagen de Sentinel-2

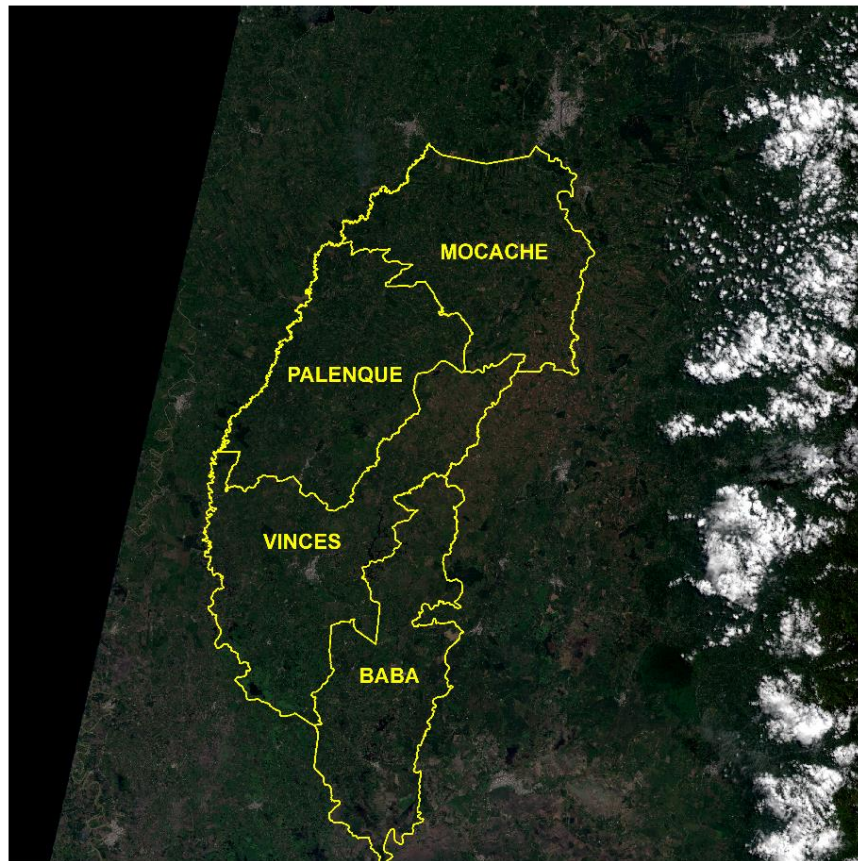
Imagen Sentinel-2	
Resolución espacial (m)	10, 20 y 60
Número de bandas espectrales	13
Fecha	17-04-2018
Porcentaje de nubosidad	4.32%
Nivel de procesado	1C
DATUM	WGS84
Proyección de mapa	Universal Transversal de Mercator
Zona	17 sur

FUENTE: USGS Science for a changing world [54]

De acuerdo a la Tabla 6, la imagen adquirida de Sentinel-2 tiene una mayor resolución espacial en sus bandas espectrales en comparación con la imagen de Landsat-8. La resolución espacial de las bandas 1, 9 y 10 es de 60 m, la resolución de las bandas 5, 6, 7, 8A, 11 y 12 es de 20 m, mientras que las de las bandas 2, 3, 4 y 8 son de 10 m. Por otro lado, esta imagen tuvo menor porcentaje de nubosidad (4.32%) que la imagen de Landsat-8 (8.99%). Además, la característica más relevante de todas es su nivel de procesado (1C), lo que evidencia la necesidad de aplicar una corrección atmosférica a la imagen adquirida. La imagen multiespectral se muestra en la figura 13.

Figura 13

Combinación RGB de imagen satelital adquirida de Sentinel-2 con fecha 17/04/2018 más el área de estudio sobrepuesta.



FUENTE: USGS Science for a changing world [54].

Como se observa en la imagen satelital de Sentinel-2 (Figura 13), la presencia de nubes se da hacia el este del área de estudio, por lo tanto, la imagen adquirida es de utilidad para realizar la clasificación supervisada.

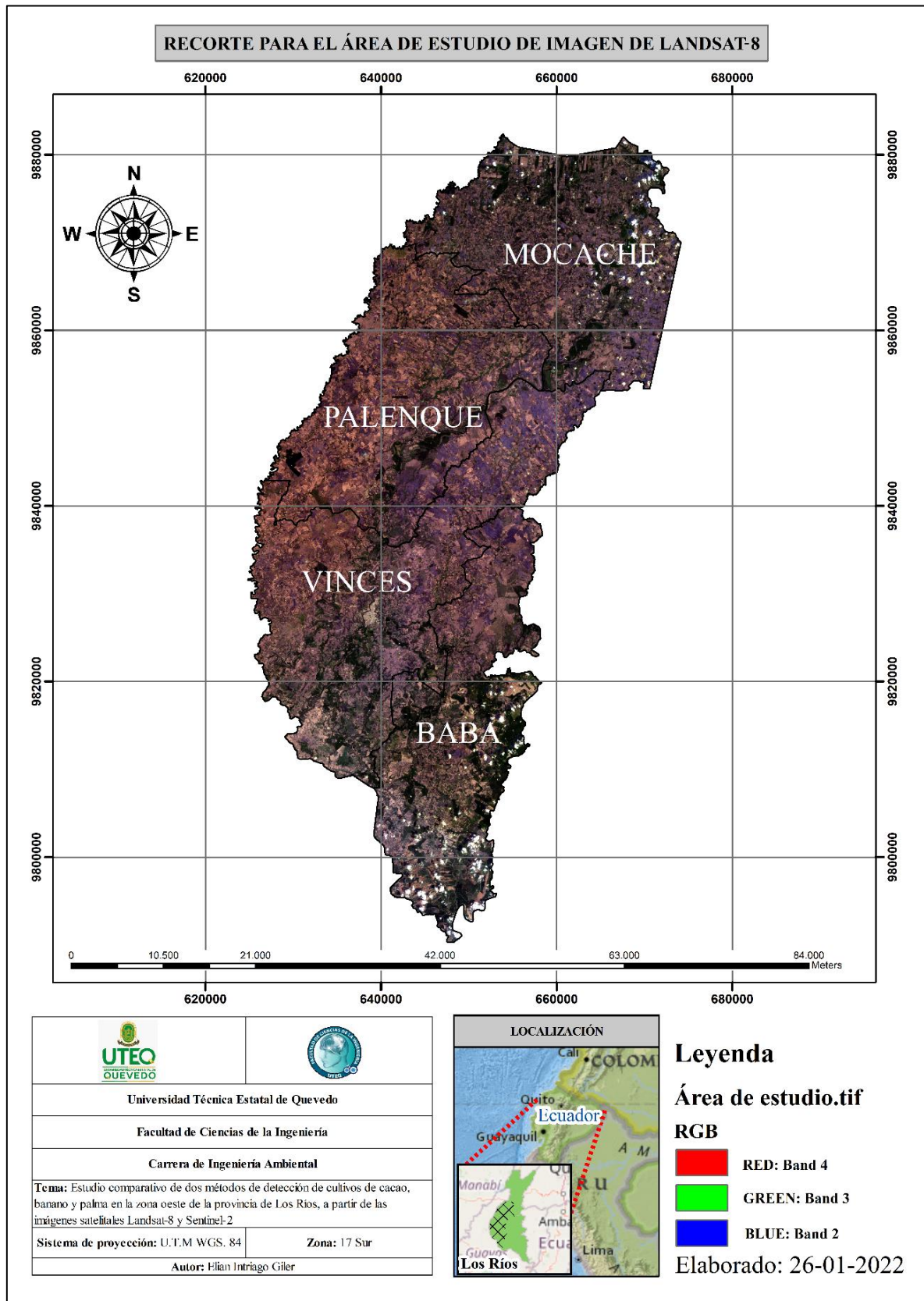
4.2 Recorte de imágenes para el área de estudio y unión de bandas espectrales necesarias.

Después de realizar la corrección atmosférica para todas las bandas espectrales de la imagen de Sentinel-2, se unieron las bandas espectrales necesarias para: el cálculo de índices espectrales, mostrar la combinación RGB y para ejecutar la clasificación supervisada, tanto en Landsat-8 como en Sentinel-2. Para la imagen de Landsat-8 se unió en una sola imagen multiespectral las bandas 2 (Azul), 3 (Verde), 4 (Roja), 5 (Infrarroja

cercana) y 6 (Infrarrojo de onda corta); mientras que para la imagen de Sentinel-2, se unieron las bandas 2 (Azul), 3 (Verde), 4 (Roja), 8 (Infrarrojo cercano) y 11 (Infrarrojo de onda corta). Posterior a la unión de bandas, se realizó el recorte para el área de estudio tanto para la imagen de Sentinel-2 como para Landsat-8. El recorte de las imágenes se lo hizo con la finalidad de calcular los índices espectrales y ejecutar la clasificación supervisada solo para el área del recorte. Los resultados del recorte se muestran en las figuras 14 y 15.

Figura 14

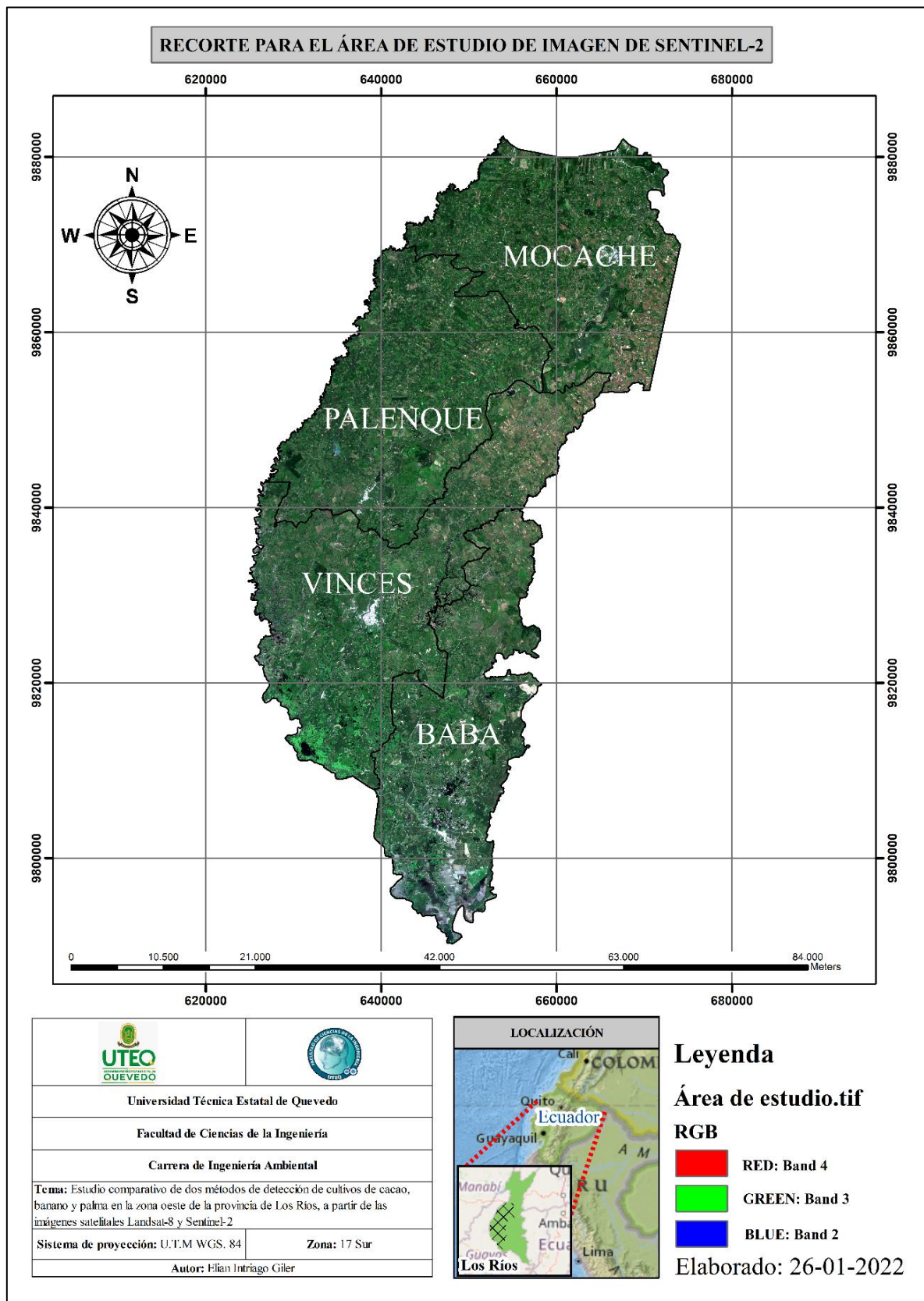
Recorte de área de estudio para imagen de Landsat-8.



ELABORADO: AUTOR

Figura 15

Recorte de área de estudio para imagen de Sentinel-2.



ELABORADO: AUTOR

4.3 Cálculo de índices espectrales.

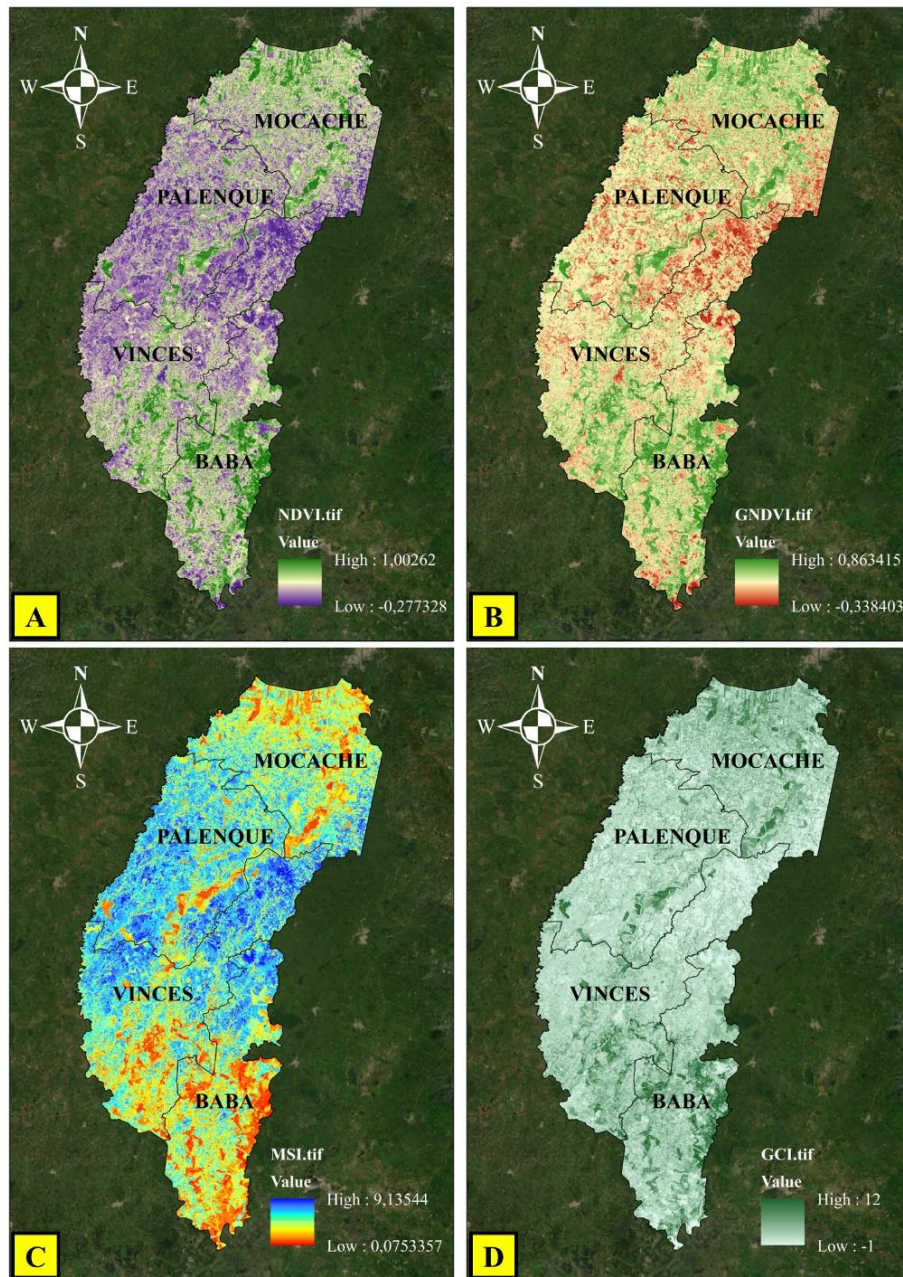
4.3.1 Cálculo individual de los índices NDVI, GNDVI, MSI y GCI.

Los índices espectrales se calcularon en el software ENVI y se representaron en el software ArcMap (Figura 16 y Figura 17). El cálculo de los índices espectrales y sus stacks (unión de índices calculados) (Figura 18 y Figura 19) fueron el eje primordial para la clasificación supervisada en el proyecto de investigación con los dos métodos determinados. Las líneas de código utilizadas para calcular los índices espectrales en el software ENVI a través de la herramienta “Band Math” en función de las ecuaciones 1, 2, 3 y 4 (presentadas en la metodología) fueron las siguientes:

- Línea de código ecuación 1 (NDVI-Landsat-8): $((\text{float}(\text{b5}) - (\text{b4})) / (\text{float}(\text{b5}) + (\text{b4})))$
- Línea de código ecuación 2 (GNDVI-Landsat-8): $((\text{float}(\text{b5}) - (\text{b3})) / (\text{float}(\text{b5}) + (\text{b3})))$
- Línea de código ecuación 3 (GCI-Landsat-8): $((\text{float}(\text{float}((\text{b5}) / (\text{b3})))) - (1)))$
- Línea de código ecuación 4 (MSI-Landsat-8): $((\text{float}((\text{float}(\text{b6})) / (\text{float}(\text{b5}))))))$
- Línea de código ecuación 6 (NDVI-Sentinel-2): $((\text{float}(\text{b8}) - (\text{b4})) / (\text{float}(\text{b8}) + (\text{b4})))$
- Línea de código ecuación 7 (GNDVI-Sentinel-2): $((\text{float}(\text{b8}) - (\text{b3})) / (\text{float}(\text{b8}) + (\text{b3})))$
- Línea de código ecuación 8 (GCI-Sentinel-2): $((\text{float}(\text{float}((\text{b8}) / (\text{b3})))) - (1)))$
- Línea de código ecuación 9 (MSI-Sentinel-2): $((\text{float}((\text{float}(\text{b11})) / (\text{float}(\text{b8}))))))$

Figura 16

Representación de índices espectrales de Landsat-8: A(NDVI) – B(GNDVI) – C(MSI) – D(GCI).



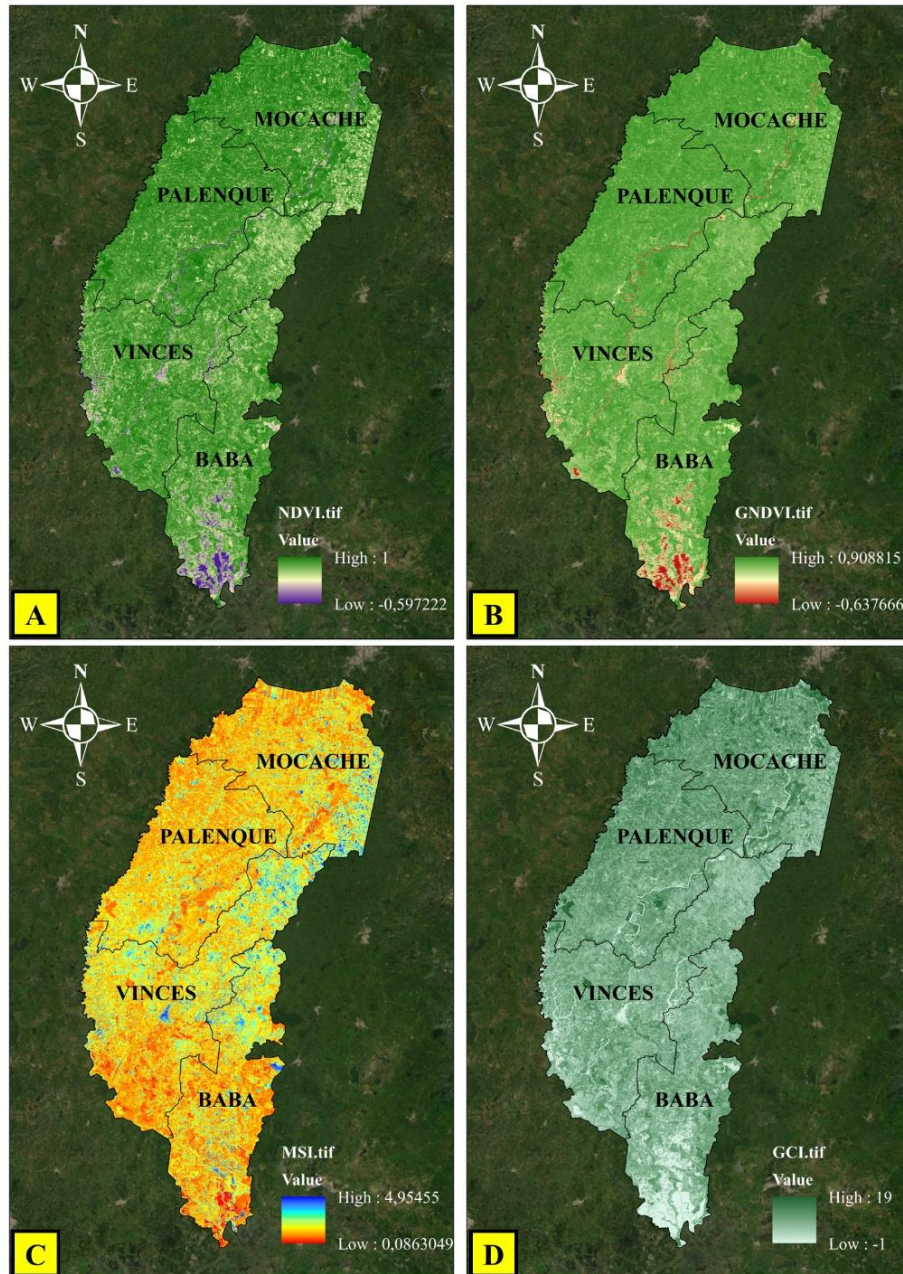
ELABORADO: AUTOR

De acuerdo a la figura 16, el NDVI obtuvo valores desde -0.277328 hasta 1.00262; los valores más altos representan la existencia de densa vegetación verde en el área de estudio. Por otro lado, el GNDVI obtuvo valores desde -0.338403 hasta 0.863415; los valores más altos del índice en la representación indican mayor actividad fotosintética de la vegetación de la zona. De la misma forma, el MSI generó valores desde 0.0753357 hasta 9.13544; se

evidencia en la representación gran cantidad de zonas con valores altos del índice (valores relacionados a la humedad). Finalmente, el GCI obtuvo valores desde -1 hasta 12; los valores más altos del índice representan mayor cantidad de clorofila.

Figura 17

Representación de índices espectrales de Sentinel-2: A(NDVI) – B(GNDVI) – C(MSI) – D(GCI).



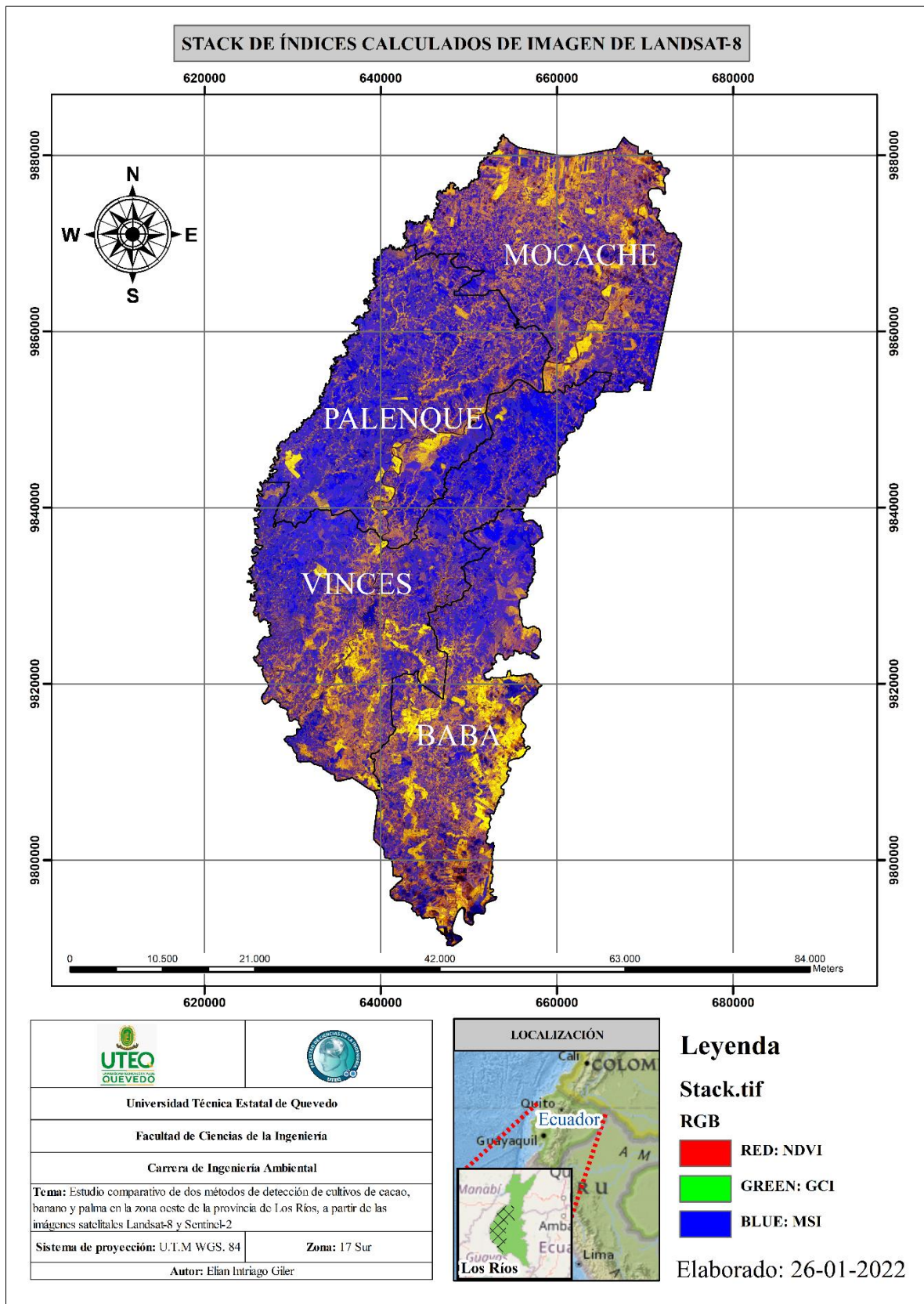
ELABORADO: AUTOR

En la figura 17 se observa que para la imagen de Sentinel-2, el NDVI obtuvo valores desde -0.597222 hasta 1. Por otro lado, el GNDVI obtuvo valores desde -0.637666 hasta 0.908815. De la misma forma, el MSI generó valores desde 0.0863049 hasta 4.95455. Finalmente, el GCI obtuvo valores desde -1 hasta 19.

4.3.2 Stack de índices.

Se generó un stack de índices calculados para la imagen de Landsat-8 y otro para la imagen de Sentinel-2; sobre dichos stacks, se crearon las regiones de interés para entrenar a los clasificadores RF y ML. Para formar el stack por cada imagen se crearon dos archivos ráster con cuatro bandas cada uno; dichas bandas fueron los índices espectrales calculados para cada imagen (Figura 16 y Figura 17). A pesar de que cada stack contiene cuatro bandas espectrales, en la representación gráfica de cada stack se muestra una combinación RGB (Figura 18 y Figura 19). En dicha combinación RGB, se le asignó la banda roja (R) al índice NDVI, la banda verde (G) al índice GCI y la banda azul (B) al índice MSI.

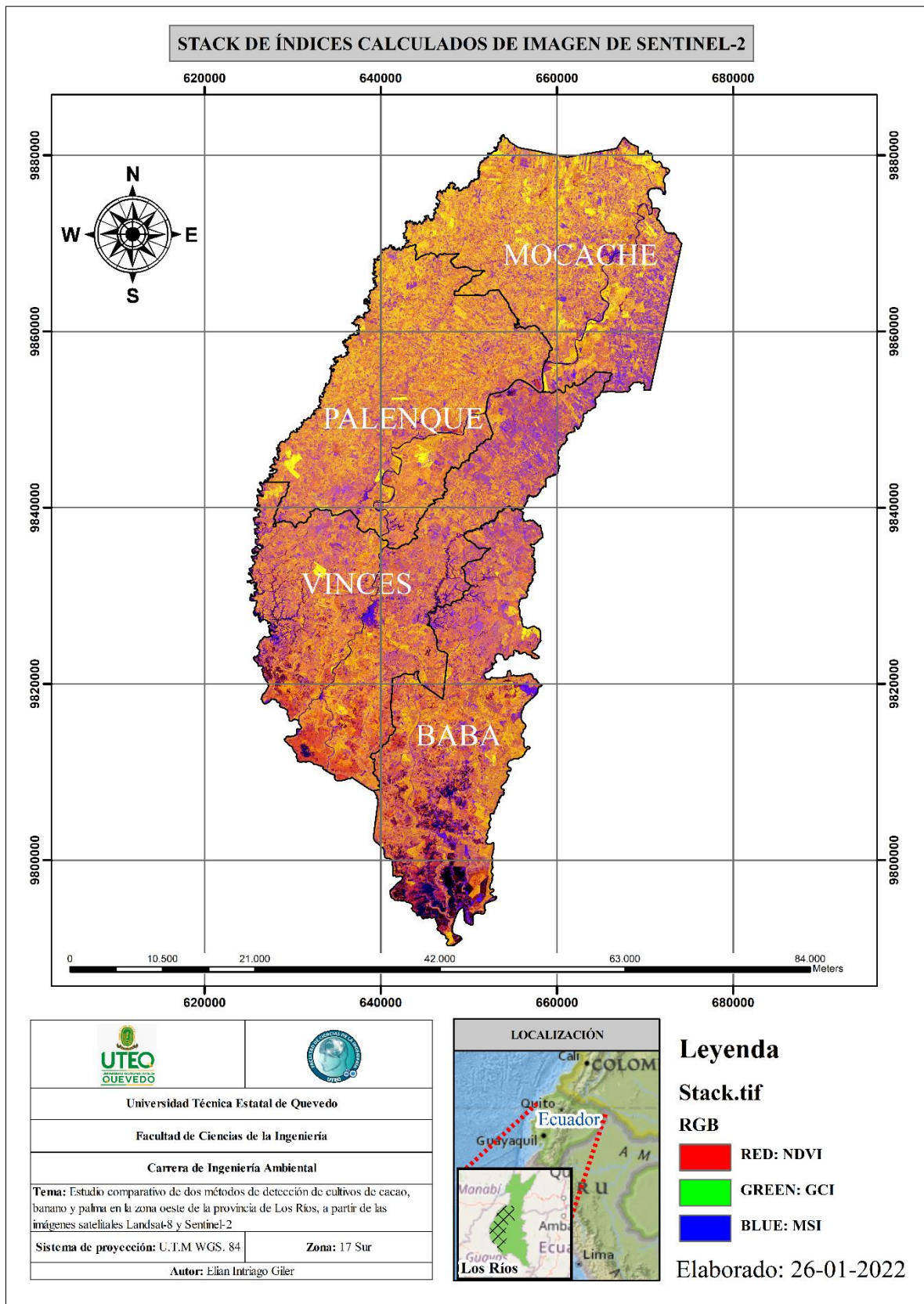
Figura 18
Stack de índices calculados de Landsat-8.



ELABORADO: AUTOR

Figura 19

Stack de índices calculados de Sentinel-2.



ELABORADO: AUTOR

4.4 Clasificación supervisada.

4.4.1 Creación de áreas de entrenamiento.

Para entrenar a los clasificadores (RF y ML) se crearon áreas de entrenamiento a partir de la identificación de zonas en el área de estudio con las siguientes coberturas de suelo: zonas con cultivos de cacao, zonas con cultivos de palma, zonas con cultivos de banano, zonas con cuerpos de agua, zonas con asentamientos humanos, zonas sin vegetación aparente y zonas con otras coberturas. Las coordenadas de las distintas coberturas identificadas se evidencian en la tabla 7.

Tabla 7
Áreas de entrenamiento

Categorías	Registro de coordenadas UTM		Procedencia	
	X	Y	Google Earth	Toma <i>in situ</i>
Zonas con cultivos de cacao	650570.6616 E	9873385.9137 S		X
	667229.8535 E	9878372.2623 S		X
	667271.4355 E	9878363.8339 S		X
	666701.5625 E	9874722.6160 S		X
	666800.1416 E	9877621.3162 S		X
	666349.8877 E	9874012.4390 S		X
	669008.4326 E	9867764.0027 S		X
Zonas con cultivos de palma	632648.4897 E	9832726.8738 S	X	
	633072.4329 E	9832336.7712 S	X	
	656305.7593 E	9826194.4836 S	X	
	656424.7668 E	9825797.9541 S	X	
	656683.4033 E	9825495.0281 S	X	
Zonas con cultivos de banano	671791.2634 E	9877341.8153 S	X	
	671798.2727 E	9876703.9262 S	X	

	641882.9443 E	9845368.7830 S	X
	642006.6559 E	9844460.3857 S	X
Zonas con cuerpos	647604.9976 E	9829989.4397 S	X
de agua	629451.6261 E	9819215.3271 S	X
	633243.9990 E	9818768.9380 S	X
	650422.1851 E	9849671.5161 S	X
Zonas con	646856.6144 E	9802892.8003 S	X
asentamientos	666224.1553 E	9868469.1565 S	X
humanos	638387.0984 E	9841427.6245 S	X
Zonas sin	632405.7088 E	9842766.0083 S	X
vegetación aparente	636168.0377 E	9842187.4500 S	X
	641739.9939 E	9848532.1802 S	X
	647257.1021 E	9851830.8936 S	X
	655030.8740 E	9850308.9063 S	X
	654189.9536 E	9849646.4038 S	X
Zonas con otras	657338.0164 E	9842128.2373 S	X
coberturas	652083.9111 E	9809407.8613 S	X
	635276.1740 E	9827659.1516 S	X
	635306.5356 E	9827989.5032 S	X
	635697.7982 E	9828126.2500 S	X
	631380.5814 E	9813386.1426 S	X
	630194.4502 E	9822240.4309 S	X
	628830.0838 E	9821807.8369 S	X
	631819.2810 E	9821807.8369 S	X
	633147.1722 E	9811956.5869 S	X
	632882.4683 E	9813301.9727 S	X
	635477.8101 E	9810403.5791 S	X
	639246.3354 E	9818751.7334 S	X

632473.6987 E	9827372.4561 S	X
630512.4646 E	9823129.2139 S	X






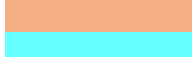


ELABORADO: AUTOR

Como se observa en la tabla 7, se definieron 45 coordenadas para las categorías de las áreas de entrenamiento, las mismas que fueron utilizadas para la creación de las regiones de interés (ver 4.4.2). Para la toma de coordenadas se utilizó principalmente Google Earth, identificando a partir de las imágenes reales captadas a nivel del suelo, zonas extensas con las coberturas necesarias según las categorías predefinidas. Por otro lado, se tomaron 7 coordenadas *in situ* para la categoría “Zonas con cultivos de cacao” debido a complicaciones para identificar áreas extensas de esta cobertura mediante Google Earth.

4.4.2 Creación de los ROIs.

Después de definir las áreas de entrenamiento, se crearon las regiones de interés en función de las categorías de la tabla 7, tanto para el stack de índices calculados de Landsat-8 (Figura 18) como para el stack de índices calculados de Sentinel-2 (Figura 19). En la tabla 8 se puede visualizar los píxeles que abarcó cada categoría.

Tabla 8
ROIs para Stacks de índices

Categoría	Color asignado	Número de píxeles por categoría	
		Landsat-8	Sentinel-2
Zonas con asentamientos humanos		716	1263
Zonas con cultivos de banano		387	1099
Zonas con cultivos de cacao		200	481
Zonas con cultivos de palma		219	1384
Zonas con cuerpos de agua		247	1344
Zonas sin vegetación aparente		963	1910
Zonas con otras coberturas		609	10869
Nubes		1481	No considerado

ELABORADO: AUTOR

De acuerdo a la tabla 8, el stack de índices calculados con mayor asignación de píxeles fue el de Sentinel-2. La mayor asignación de píxeles para el stack de Sentinel-2 obedece a una mayor resolución espacial con respecto al stack de Landsat-8, lo que derivó en una mayor cantidad de píxeles en un área menor. Debido a la presencia de nubes sobre parte de la imagen de Landsat-8, para la clasificación se agregó una categoría adicional de “Nubes” no considerada en las áreas de entrenamiento (Tabla 7).

4.4.3 Cálculo de separabilidad espectral de ROIs.

Una vez creadas las regiones de interés para cada stack de índices calculados, fue necesario verificar la separabilidad espectral que tienen dichos ROIs. Esta separabilidad espectral permitió identificar las categorías que mayor dificultad tienen para diferenciarse entre sí; por lo tanto, dicho parámetro es un indicador importante sobre la calidad de los ROIs creados y su funcionalidad/utilidad de acuerdo a los objetivos de detección. Los resultados de separabilidad pueden fluctuar entre 0 y 2, donde 2 es el indicador de mayor separabilidad espectral; es decir, mientras el resultado más se aleje de 2, menor separabilidad espectral habrá.

De acuerdo a la tabla 9, Se calcularon 28 valores de separabilidad espectral para las ROIs del Stack de Landsat-8. La menor separabilidad espectral se dio entre las categorías “Zonas con cultivos de cacao” y “Zonas con otras coberturas” con un valor de 1.47, mientras que la separabilidad más alta fue de 2.00 entre “Zonas con cultivos de cacao” y “Zonas sin vegetación aparente”, “Zonas con cultivos de cacao” y “Zonas con cuerpos de agua”, “Zonas con cultivos de palma” y “Zonas con cuerpos de agua”, “Zonas con cultivos de banano” y “Zonas con cuerpos de agua”, “Zonas con asentamientos humanos” y “Zonas con cultivos de banano”, “Zonas con asentamientos humanos” y “Zonas con cultivos de palma”, “Zonas con cuerpos de agua” y “Zonas sin vegetación aparente”, “Zonas con cultivos de palma” y “Zonas sin vegetación aparente”, y entre “Zonas con cultivos de banano” y “Zonas sin vegetación aparente”.

Tabla 9

Separabilidad espectral Jeffries-Matusita de los ROIs del stack de índices calculados de Landsat-8

Separabilidad espectral de ROIs – Landsat-8
Zonas con cultivos de cacao y Zonas con otras coberturas = 1.47134083
Zonas con asentamientos humanos y Nubes = 1.63593065
Zonas con cultivos de banano y Zonas con cultivos de palma = 1.86146314
Zonas con otras coberturas y Nubes = 1.88821816
Zonas con asentamientos humanos y Zonas con otras coberturas = 1.90394173
Zonas con cultivos de palma y Zonas con otras coberturas = 1.95365374
Zonas con cultivos de banano y Zonas con otras coberturas = 1.95751270
Zonas sin vegetación aparente y Zonas con otras coberturas = 1.98946556
Zonas con cultivos de cacao y Nubes = 1.99220436
Zonas con cultivos de cacao and Zonas con cultivos de palma = 1.99938979
Zonas con asentamientos humanos y Zonas sin vegetación aparente = 1.99944666
Zonas con cuerpos de agua y Nubes = 1.99963234
Zonas con cultivos de banano y Zonas con cultivos de cacao = 1.99978317
Zonas con asentamientos humanos y Zonas con cultivos de cacao = 1.99999721
Zonas sin vegetación aparente y Nubes = 1.99999791
Zonas con cultivos de banano y Nubes = 1.99999856
Zonas con cuerpos de agua y Otras coberturas = 1.99999980
Zonas con cultivos de palma y Nubes = 1.99999984
Zonas con asentamientos humanos y Zonas con cuerpos de agua = 1.99999996
Zonas con cultivos de cacao y Zonas sin vegetación aparente = 2.00000000
Zonas con cultivos de cacao y Zonas con cuerpos de agua = 2.00000000
Zonas con cultivos de palma y Zonas con cuerpos de agua = 2.00000000
Zonas con cultivos de banano y Zonas con cuerpos de agua = 2.00000000
Zonas con asentamientos humanos y Zonas con cultivos de banano = 2.00000000
Zonas con asentamientos humanos y Zonas con cultivos de palma = 2.00000000
Zonas con cuerpos de agua y Zonas sin vegetación aparente = 2.00000000
Zonas con cultivos de palma y Zonas sin vegetación aparente = 2.00000000
Zonas con cultivos de banano y Zonas sin vegetación aparente = 2.00000000

ELABORADO: AUTOR

De las tres categorías objetivo de detección (cacao, banano y palma) el cacao fue el que menor separabilidad espectral presentó; con “Zonas con otras coberturas” obtuvo una separabilidad de 1.47. La baja separabilidad indicada se debe a que el ROI de “Zonas con cultivos de cacao” tuvo pocos píxeles (área representativa pequeña) debido a la escasa identificación de áreas grandes y únicas con este cultivo (generalmente el cacao no se siembra en grandes extensiones, o se siembra junto a otro tipo de vegetación y/o cultivo); además, la baja calidad de la imagen dificultó la identificación del área con el cultivo al momento de crear la región de interés respectiva. Los demás resultados de separabilidad espectral de “Zonas con cultivos de cacao” con otros ROIs fueron excelentes, con valores de 1.99 hasta 2.00.

A diferencia del cacao, para los cultivos de banano y palma si se identificaron varias áreas extensas solo con este tipo de coberturas, por lo tanto, la separabilidad espectral mejoró. Para el cultivo de banano los valores más bajos de separabilidad fueron 1.86 (con zonas con cultivos de palma) y 1.95 (con zonas con otras coberturas); los demás valores de separabilidad para el banano oscilaron entre 1.99 y 2.00. De la misma forma, para los ROIs del cultivo de palma sus valores más bajos fueron de 1.86 (con zonas con cultivos de banano) y 1.95 (con zonas con otras coberturas).

Por otro lado, como se observa en la tabla 10, se calcularon 21 valores de separabilidad espectral para las categorías del Stack de Sentinel-2. La menor separabilidad espectral se dio entre las categorías “Zonas con cultivos de cacao” y “Zonas con otras coberturas” con un valor de 1.90. Por otro lado la separabilidad espectral más alta fue de 2.00 entre “Zonas con cultivos de palma” y “Zonas sin vegetación aparente”, “Zonas con cultivos de cacao” y “Zonas con cuerpos de agua”, “Zonas con asentamientos humanos” y “Zonas con cultivos de banano”, “Zonas con cuerpos de agua” y “Zonas sin vegetación aparente”, “Zonas con asentamientos humanos” y “Zonas con cultivos de palma”, “Zonas con cultivos de palma” y “Zonas con cuerpos de agua”, y entre “Zonas con cultivos de banano” y “Zonas con cuerpos de agua”.

Tabla 10

Separabilidad espectral Jeffries-Matusita de los ROIs del stack de índices calculados de Sentinel-2

Separabilidad espectral de ROIs – Sentinel-2
Zonas con cultivos de cacao y Zonas con otras coberturas = 1.90092264
Zonas con asentamientos humanos y Zonas sin vegetación aparente = 1.93646725
Zonas con cultivos de banano y Zonas con otras coberturas = 1.98960358
Zonas con cultivos de palma y Zonas con otras coberturas = 1.99724824
Zonas asentamientos humanos y Zonas con cuerpos de agua = 1.99906934
Zonas con cuerpos de agua y Zonas con otras coberturas = 1.99981115
Zonas con cultivos de cacao y Zonas con cultivos de palma = 1.99987623
Zonas con cultivos de cacao y Zonas sin vegetación aparente = 1.99990283
Zonas con asentamientos humanos y Zonas con otras coberturas = 1.99998437
Zonas con cultivos de banano y Zonas con cultivos de cacao = 1.99998534
Zonas con cultivos de banano y Zonas con cultivos de palma = 1.99999294
Zonas sin vegetación aparente y Zonas con otras coberturas = 1.99999522
Zonas con asentamientos humanos y Zonas con cultivos de cacao = 1.99999768
Zonas con cultivos de banano y Zonas sin vegetación aparente = 1.99999999
Zonas con cultivos de palma y Zonas sin vegetación aparente = 2.00000000
Zonas con cultivos de cacao y Zonas con cuerpos de agua = 2.00000000
Zonas con asentamientos humanos y Zonas con cultivos de banano = 2.00000000
Zonas con cuerpos de agua y Zonas sin vegetación aparente = 2.00000000
Zonas con asentamientos humanos y Zonas con cultivos de palma = 2.00000000
Zonas con cultivos de palma y Zonas con cuerpos de agua = 2.00000000
Zonas con cultivos de banano y Zonas con cuerpos de agua = 2.00000000

ELABORADO: AUTOR

De la misma forma, también se obtuvo varias separabilidades espectrales (13) que oscilaron entre 1.93 y 1.99. Es notorio que las separabilidades espectrales para el stack de índices Sentinel-2 fueron más altas que las obtenidas para el stack de Landsat-8; la razón principal de la mejora mencionada se debe al aumento de la calidad de la imagen de Sentinel-2, lo que permitió una mejor identificación de las categorías de los ROIs creados para dicha imagen.










4.4.4 Clasificación supervisada.

4.4.4.1 Clasificación con el método de Maximum Likelihood de la imagen de Landsat-8.

De acuerdo a la tabla 11, el clasificador ML para la imagen de Landsat-8 detectó que la mayor cantidad de área la ocupa la cobertura “Zonas con otras coberturas” con un total de 1512.96 km². Por otro lado, la cobertura con menor ocupación de área fue “Zonas con cuerpos de agua” con 21.21 km² (exceptuando lo que no se clasificó). Para los objetivos principales de detección (cacao, banano y palma) se reportó un total de 208.71 km², 74.83 km² y 24.19 km², respectivamente. El área total detectada fue de 2348.3241 km² a partir de la identificación de 2609249 píxeles con resolución espacial de 30 m x 30 m (900 m² por píxel). En la figura 20 se pueden visualizar las coberturas detectadas por el clasificador.

Tabla 11

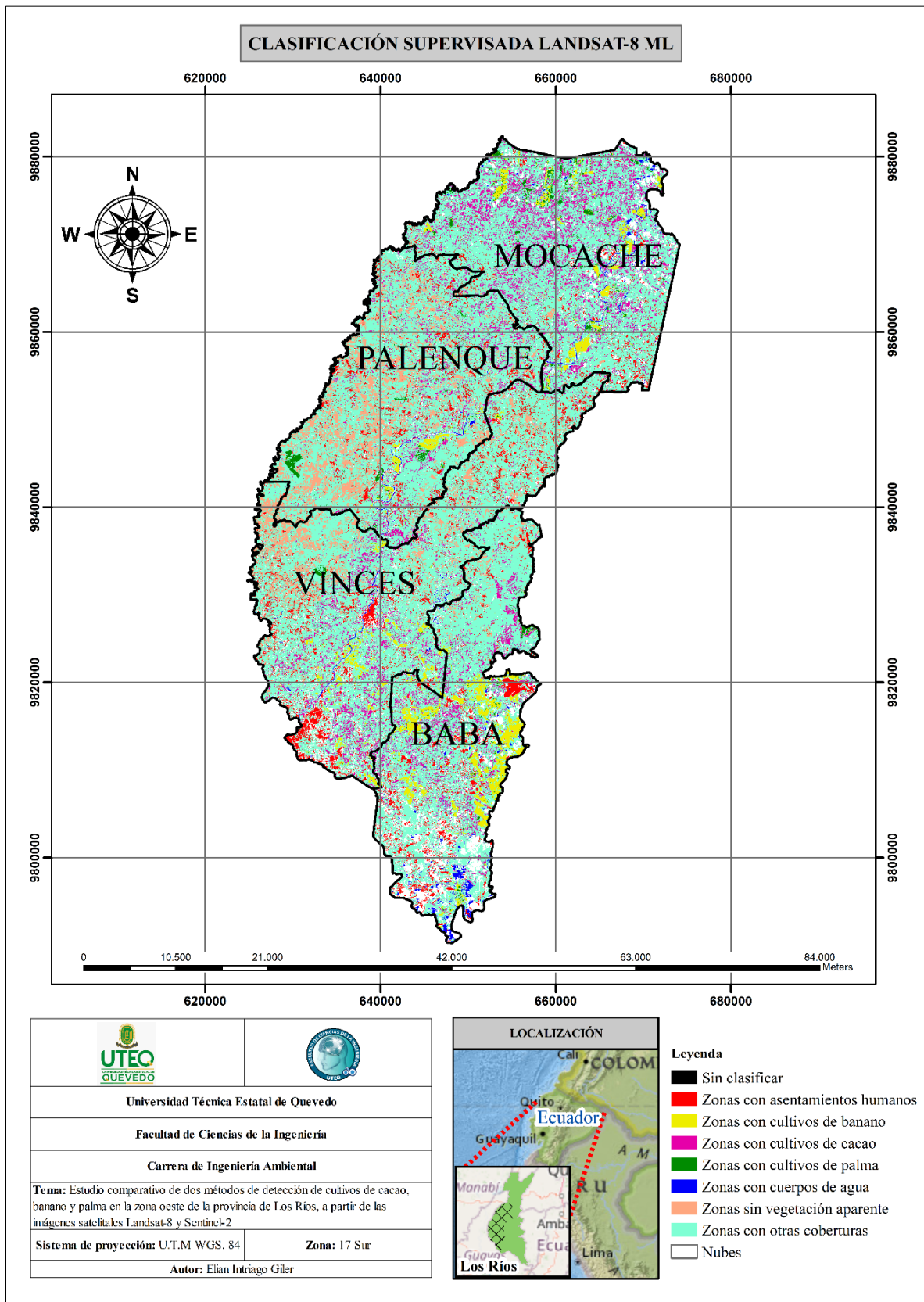
Extensión de coberturas identificadas con el clasificador Maximum Likelihood para la imagen de Landsat-8

Color id	Categoría	Conteo (píxeles)	Área (m ²)	Área (km ²)
	Zonas con asentamientos humanos	167491	150741900	150.7419
	Zonas con cultivos de banano	83140	74826000	74.826
	Zonas con cultivos de cacao	231897	208707300	208.7073
	Zonas con cultivos de palma	26880	24192000	24.192
	Zonas con cuerpos de agua	23566	21209400	21.2094
	Zonas sin vegetación aparente	183139	164825100	164.8251
	Zonas con otras coberturas	1681065	1512958500	1512.9585
	Nubes	212069	190862100	190.8621
	Sin clasificar	2	1800	0.0018
	Total	2609249	2348324100	2348.3241

ELABORADO: AUTOR

Figura 20

Clasificación supervisada con ML de la imagen de Landsat-8.











ELABORADO: AUTOR

4.4.4.2 Clasificación con el método de Maximum Likelihood de la imagen de Sentinel-2.

En la tabla 12, se observa que el clasificador ML para la imagen de Sentinel-2 detectó mayor cantidad de área para “Zonas con otras coberturas” con un total de 1091.29 km². Por otro lado, la cobertura con menor ocupación de área fue “Zonas con cultivos de palma” con 31.23 km². Para los otros dos objetivos principales de detección (cacao y banano), se reportó un total de 500.30 km² y 34.97 km², respectivamente. El área total detectada fue de 2348.3177 km² a partir de la identificación de 23483177 píxeles con resolución espacial de 10 m x 10 m (100 m² por píxel). En la figura 21 se pueden visualizar las coberturas detectadas por el clasificador.

Tabla 12

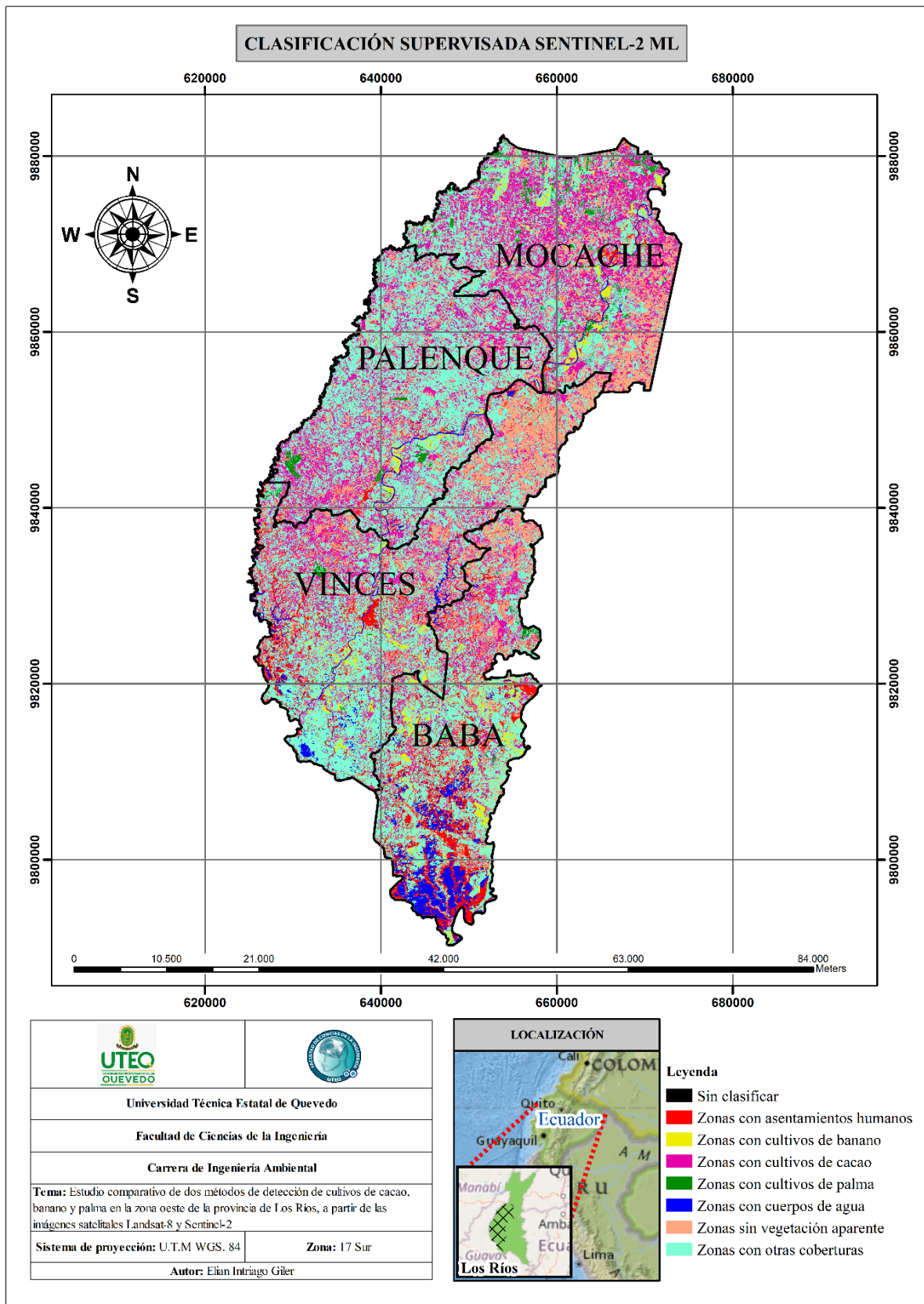
Extensión de coberturas identificadas con el clasificador Maximum Likelihood para la imagen de Sentinel-2

Color id	Categoría	Conteo (píxeles)	Área (m ²)	Área (km ²)
	Zonas con asentamientos humanos	2074305	207430500	207.4305
	Zonas con cultivos de banano	349646	34964600	34.9646
	Zonas con cultivos de cacao	5003043	500304300	500.3043
	Zonas con cultivos de palma	312343	31234300	31.2343
	Zonas con cuerpos de agua	729550	72955000	72.955
	Zonas sin vegetación aparente	4101510	410151000	410.151
	Zonas con otras coberturas	10912780	1091278000	1091.278
	Sin clasificar	0	0	0
	Total	23483177	2348317700	2348.3177

ELABORADO: AUTOR

Figura 21

Clasificación supervisada con ML de la imagen de Sentinel-2.












4.4.4.3 Clasificación con el método Random Forest de la imagen de Landsat-8.

Igual que el clasificador ML, el clasificador Random Forest detectó que la mayor cantidad de área la ocupa la cobertura “Zonas con otras coberturas” con un total de 1242.09 km². De la misma manera, la cobertura con menor ocupación de área fue “Zonas con cuerpos de agua” con 14.91 km². Para los objetivos principales de detección (cacao, banano y palma) se reportó un total de 222.68 km², 118.52 km² y 15.30 km², respectivamente. El área total detectada fue de 2348.3241 km² a partir de la identificación de 2609249 pixeles con resolución espacial de 30 m x 30 m, al igual que con el clasificador ML para el stack de Landsat-8 (Tabla 13). En la figura 22 se pueden observar las coberturas detectadas por el clasificador.

Tabla 13

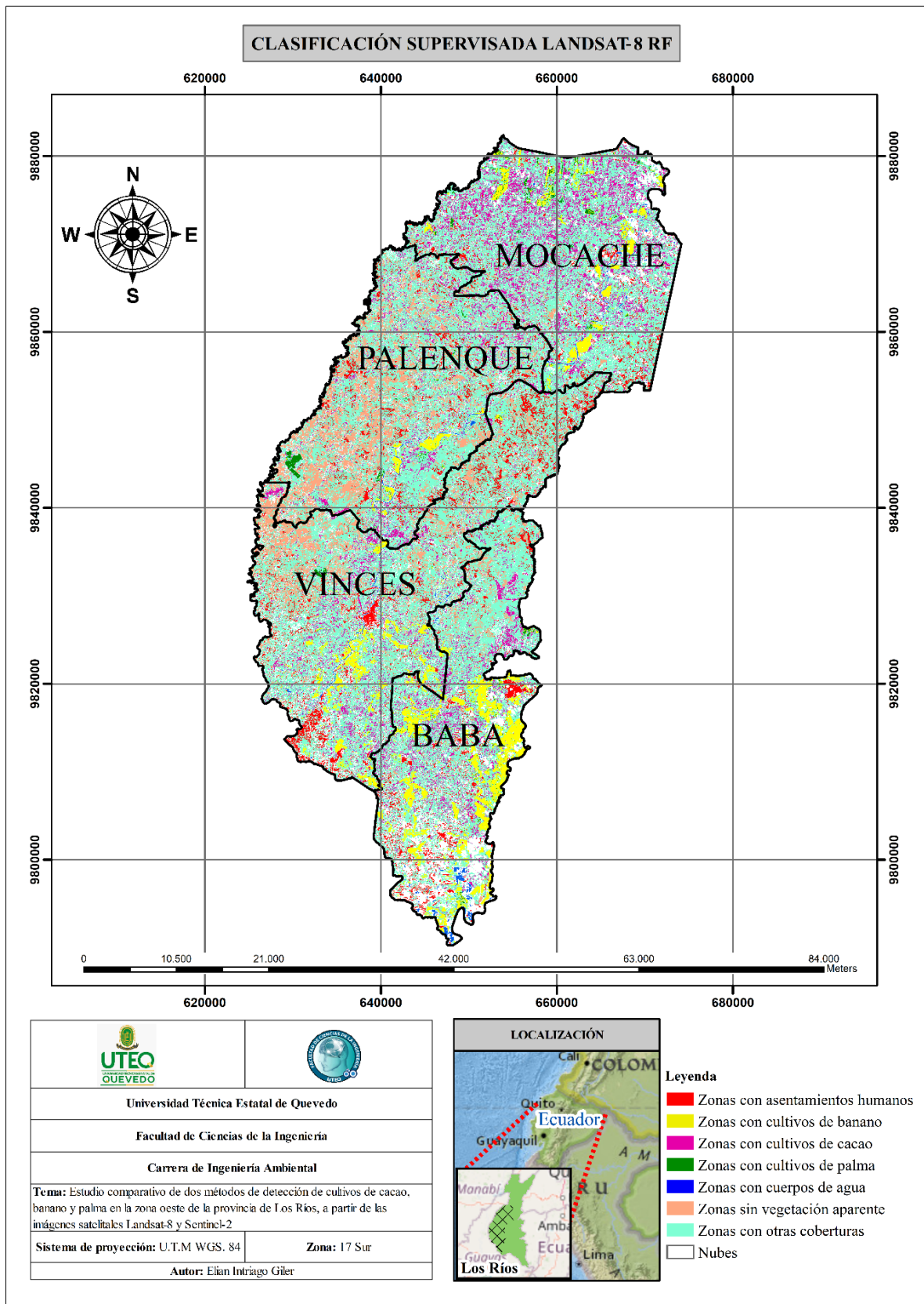
Extensión de coberturas identificadas con el clasificador Random Forest para la imagen de Landsat-8

Color id	Categoría	Conteo (pixeles)	Área (m ²)	Área (km ²)
	Zonas con asentamientos humanos	171786	154607400	154.6074
	Zonas con cultivos de banano	131683	118514700	118.5147
	Zonas con cultivos de cacao	247425	222682500	222.6825
	Zonas con cultivos de palma	17001	15300900	15.3009
	Zonas con cuerpos de agua	16569	14912100	14.9121
	Zonas sin vegetación aparente	249281	224352900	224.3529
	Zonas con otras coberturas	1380095	1242085500	1242.0855
	Nubes	395409	355868100	355.8681
	Sin clasificar	0	0	0
	Total	2609249	2348324100	2348.3241

ELABORADO: AUTOR

Figura 22

Clasificación supervisada con RF de la imagen de Landsat-8.







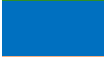



ELABORADO: AUTOR

4.4.4.4 Clasificación con el método Random Forest de la imagen de Sentinel-2.

De igual forma que las demás clasificaciones, el clasificador Random Forest para la imagen de Sentinel-2 detectó que la mayor cantidad de área pertenece a “Zonas con otras coberturas” con un total de 1631.51 km². Por otro lado, la cobertura con menor ocupación de área fue “Zonas con cultivos de banano” con 26.55 km². Para el cultivo de palma y cacao (los otros dos objetivos de detección) se reportó un total de 26.68 km² y 130.96 m² respectivamente. El área total detectada fue de 2348.3177 km² a partir de la identificación de 23483177 píxeles con resolución espacial de 10 m x 10 m, al igual que con el clasificador ML para el Stack de Sentinel-2 (Tabla 14). En la figura 23 se muestran las coberturas detectadas en la clasificación.

Tabla 14

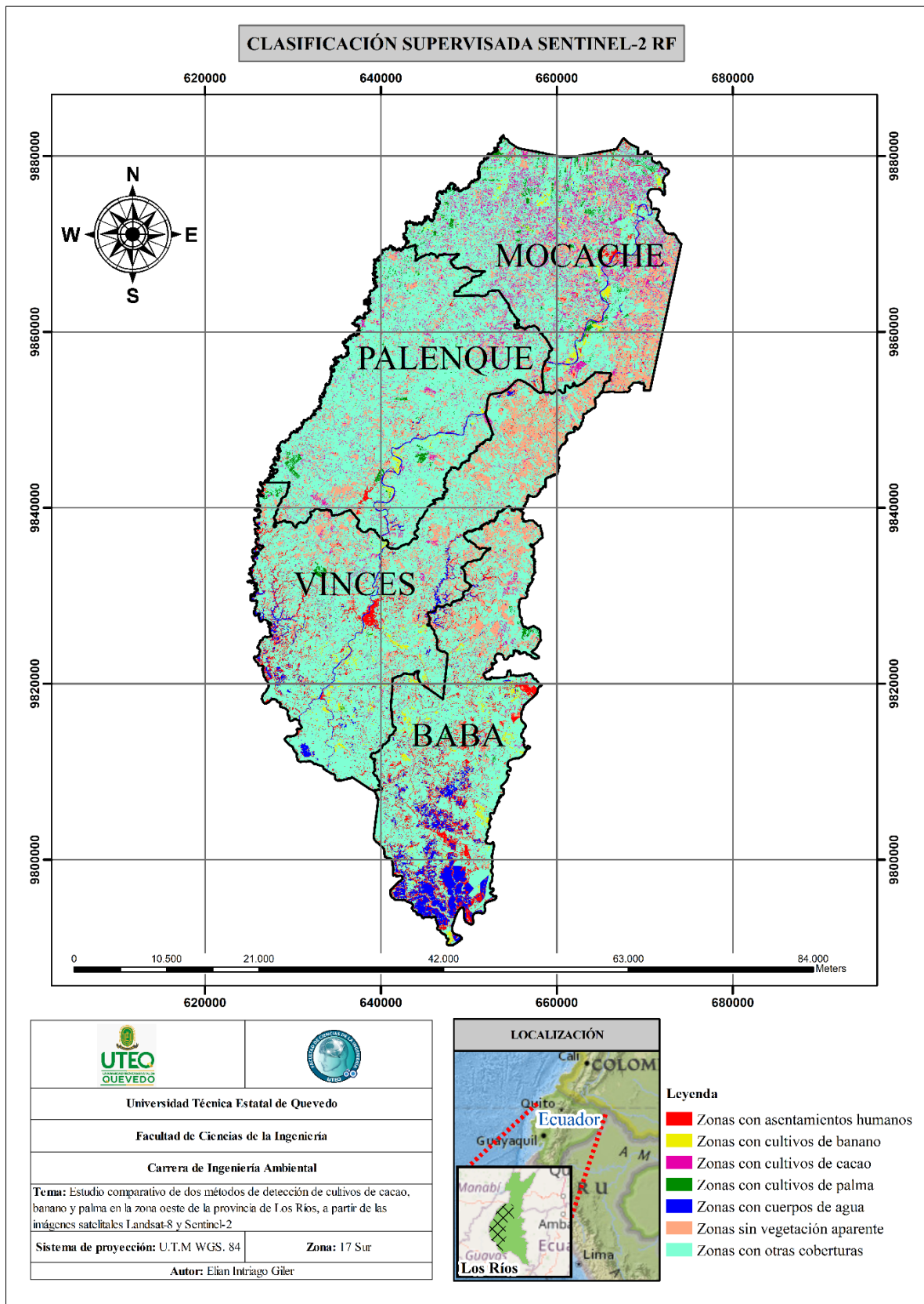
Extensión de coberturas identificadas con el clasificador Random Forest para la imagen de Sentinel-2

Color id	Categoría	Conteo (píxeles)	Área (m ²)	Área (km ²)
	Zonas con asentamientos humanos	1275608	127560800	127.5608
	Zonas con cultivos de banano	265459	26545900	26.5459
	Zonas con cultivos de cacao	1309600	130960000	130.96
	Zonas con cultivos de palma	266827	26682700	26.6827
	Zonas con cuerpos de agua	689621	68962100	68.9621
	Zonas sin vegetación aparente	3360949	336094900	336.0949
	Zonas con otras coberturas	16315113	1631511300	1631.5113
	Sin clasificar	0	0	0
	Total	23483177	2348317700	2348.3177

ELABORADO: AUTOR

Figura 23

Clasificación supervisada con RF de la imagen de Sentinel-2.



ELABORADO: AUTOR

4.5 Validación de resultados.

A continuación, se presenta el apartado de validación de resultados. Inicialmente, se crearon áreas de entrenamiento con coordenadas al azar (Tabla 15), las cuales permitieron crear nuevos ROIs. Con los nuevos ROIs se generaron las matrices de confusión (Figura 24) y se calcularon los coeficientes de Kappa (Tabla 16) en función de las 4 clasificaciones realizadas (ver 4.4.4).

4.5.1 Áreas de entrenamiento al azar.

La tabla 15 muestra las áreas de entrenamiento al azar creadas. Para este procedimiento se utilizó principalmente Google Earth; no obstante, para las categorías “Zonas con cultivos de banano” y “Zonas con otras coberturas”, se tomaron coordenadas *in situ*.

Tabla 15

Áreas de entrenamiento al azar para creación de nuevos ROIs y generación de matrices de confusión y cálculo de coeficientes de Kappa

Categorías	Registro de coordenadas UTM		Procedencia	
	X	Y	Google Earth	Toma <i>in situ</i>
Zonas con cultivos de cacao	664996.6944 E	9872461.3683 N		X
	664965.2071 E	9871412.5831 N		X
	665000.7763 E	9871350.9671 N		X
	665449.1340 E	9869618.8714 N		X
Zonas con cultivos de palma	630623.5254 E	9845577.0972 N	X	
	629873.0508 E	9844925.7471 N	X	
	630019.3103 E	9844226.3037 N	X	
Zonas con cultivos de banano	640254.7290 E	9835614.7021 N	X	
	640629.4922 E	9835938.4570 N	X	
Zonas con cuerpos de agua	644264.1028 E	9849648.7134 N	X	
	643669.8059 E	9849731.8188 N	X	

	669153.9258 E	9873361.6779 N	X	
	650444.4067 E	9849689.1748 N	X	
Zonas con asentamientos	639526.4291 E	9828965.4773 N	X	
humanos	633600.7269 E	9818598.2104 N	X	
Zonas sin vegetación	653022.1021 E	9849611.5161 N	X	
aparente	660226.7529 E	9854595.8716 N	X	
Zonas con otras	638748.7354 E	9849834.8877 N	X	
coberturas	657782.1826 E	9854103.8696 N	X	
	641996.8994 E	9822858.0371 N	X	
	649951.4355 E	9815725.2979 N	X	
	650446.0590 E	9801366.8240 N	X	
	640035.7434 E	9802178.0273 N	X	
	648870.2832 E	9817409.6924 N	X	
	651992.0825 E	9811764.1260 N	X	
	666601.3570 E	9874334.1664 N		X
	665205.0843 E	9870281.5024 N		X

ELABORADO: AUTOR

4.5.2 Matrices de confusión.

Una vez creados los ROIs en función de la tabla 15, se generaron las matrices de confusión. En la figura 24 se puede visualizar las cuatro matrices con los niveles de precisión general e individual (por categorías) de cada clasificación.

Figura 24

Matrices de confusión de clasificaciones realizadas: A(ML-L8) – B(ML-S2) – C(RF-L8) – D(RF-S2).



ELABORADO: AUTOR

De acuerdo a la figura 24, la mejor clasificación se obtuvo utilizando Random Forest con imagen de Sentinel-2 (precisión general de 89%), mientras que, se obtuvo una clasificación con menor precisión utilizando Maximum Likelihood con imagen de Sentinel-2 (precisión general de 76%). Al mismo tiempo, para los objetivos principales de detección (banano, cacao y palma) se obtuvieron porcentajes de acuerdo/exactitud individuales que fluctuaron entre 33.85% (más baja) y 98.63% (más alta).

El cultivo de cacao obtuvo los porcentajes más bajos de acuerdo/exactitud (entre los objetivos principales de detección); con ML-L8 obtuvo 50%, con RF-L8 41.67% y con RF-S2 33.85% (con ML-S2 obtuvo 54.56% pero el más bajo fue banano). Por lo tanto, las zonas con cultivos de cacao no se detectaron de forma precisa en ninguna clasificación. No obstante, las zonas con cultivos de palma se detectaron de forma muy precisa con ML-L8 (96.30%), ML-S2 (98.76%) y RF-L8 (96.17%); en ese mismo sentido, las zonas con cultivos de banano también se detectaron de forma muy precisa, pero solo con ML-L8 (98.63%) y con RF-L8 (96.58%).

La clasificación que de forma conjunta detectó con mayor precisión los tres objetivos principales de detección fue ML-L8, con 98.63% de acuerdo/exactitud para zonas con cultivos de banano, 50% para zonas con cultivos de cacao y 96.30% para zonas con cultivos de palma. Por el contrario, la clasificación con menor precisión para la detección de banano, cacao y palma fue RF-S2, con porcentajes de 64.47%, 33.85% y 72.84%, respectivamente; no obstante, fue la mejor clasificación general por su alta precisión en la detección de las demás categorías.

En relación a las demás categorías clasificadas, los cuerpos de agua obtuvieron los porcentajes más altos de acuerdo/exactitud (100% con ML-L8 y con RF-L8). En cambio, los porcentajes más bajos se obtuvieron en la categoría zonas con otras coberturas, con valores por debajo del 73% utilizando ML-L8, ML-S2 y RF-L8 (con RF-S2 se obtuvo 93.85%). De la misma forma, zonas sin vegetación aparente obtuvo porcentajes de acuerdo/exactitud del 67.10% con ML-L8 y 68.74% con RF-L8, mientras que, zonas con cuerpos de agua obtuvo 80.25% con ML-S2. Los demás porcentajes (la mayoría) fluctuaron entre 86.19% y 99.29%

4.5.3 Coeficientes de Kappa.

De la misma forma, se calcularon los coeficientes de Kappa para las cuatro clasificaciones realizadas (ver 4.4.4). La tabla 16 muestra los resultados.

Tabla 16

Coeficientes de Kappa para las clasificaciones realizadas con Maximum Likelihood y Random Forest

Clasificación	Coeficiente de Kappa	Precisión
Maximum Likelihood – Landsat-8	0.79	Considerable
Maximum Likelihood – Sentinel-2	0.61	Considerable
Random Forest – Landsat-8	0.80	Considerable
Random Forest – Sentinel-2	0.78	Considerable

ELABORADO: AUTOR

De acuerdo a la tabla 16, todas las clasificaciones realizadas se encuentran dentro del rango “Considerable”. La mejor clasificación se obtuvo utilizando Random Forest con imagen de Landsat-8 (0.80), mientras que, la clasificación con menor precisión se dio al utilizar Maximum Likelihood con Sentinel-2 (0.61) (igual que en la matriz de confusión). Según este coeficiente, las mejores clasificaciones se obtuvieron con imágenes de Landsat-8 (0.80 con RF-L8, y 0.79 con ML-L8); aunque, la clasificación de Sentinel-2 con Random Forest obtuvo un coeficiente similar (0.78). Dicha similitud en la precisión de estas tres clasificaciones (RF-L8, ML-L8 y RF-S2) también se observó con los resultados de la matriz de confusión, aunque en órdenes distintos y con precisiones de 82% (ML-L8), 84% (RF-L8) y 89% (RF-S2).

4.6 Análisis estadístico.

Una vez que se extrajeron los valores al azar y se descartó la normalidad de los datos, se ejecutó el análisis estadístico con la prueba no-paramétrica Wilcoxon-Mann-Whitney para muestras independientes. El resultado se puede observar dentro de la tabla 17.

Tabla 17

Estadísticos de prueba de Suma de Rangos de Wilcoxon para las comparaciones estadísticas entre métodos clasificadores. a: variable de agrupación MLIRF2. b: variable de agrupación MLIRF2

Estadísticos de prueba	^a Valores por puntos al azar	^b Valores por puntos al
	Landsat-8	azar Sentinel-2
U de Mann-Whitney	2728860.000	2047002,500
W de Wilcoxon	5486586.000	2047002,500
Z	-.667	-17,028
Sig. asintótica(bilateral)	.505	5,0514E-65

ELABORADO: AUTOR

Como se observa en la tabla 17, el valor de p (Sig. asintótica (bilateral)) de la comparación estadística entre las clasificaciones de Landsat-8 es 0.505. Por lo tanto, la hipótesis nula no es rechazada y se infiere que no existen diferencias estadísticas significativas entre los resultados obtenidos con Maximum Likelihood y Random Forest para la clasificación de la imagen de Landsat-8. Por el contrario, el valor de p de la comparación estadística entre las clasificaciones de Sentinel-2 es menor que 0.05. Por consecuencia, se rechaza la hipótesis nula y es posible afirmar que existen diferencias estadísticas entre los resultados obtenidos con ML Y RF para la clasificación de la imagen de Sentinel-2.

Los resultados del análisis estadístico respaldan lo reportado por las matrices de confusión y por los coeficientes de Kappa. En las clasificaciones de Landsat-8 la diferencia fue de 2% en las matrices de confusión y de 0.01 en los coeficientes de Kappa; por lo tanto, es lógico que no haya diferencias significativas entre las clasificaciones de Landsat-8 con los métodos ya mencionados. Por otro lado, en las clasificaciones de Sentinel-2 hubo diferencias considerables en las matrices de confusión (13%) y en los coeficientes de Kappa (0.18); esto fue respaldado por en análisis estadístico que sí demostró diferencias significativas en el reporte.

4.7 Discusión.

El método de Maximum Likelihood ha sido uno de los más utilizados desde hace décadas para la clasificación de imágenes multiespectrales; sin embargo, han surgido nuevos métodos basados en inteligencia artificial como Random Forest [6], [7]. Esto incluye la necesidad de comprobar la precisión de estos nuevos métodos de clasificación. Además, en la provincia Los Ríos, donde la producción agrícola es transcendental en las condiciones socioeconómicas (Ministerio de Agricultura Ganadería Acuicultura y Pesca, 2016), resulta imprescindible el mapeo actualizado de la distribución de los cultivos más importantes [63], ya que los mapas que ofrecen las entidades gubernamentales públicas suelen quedar obsoletos y es común que se engloben las áreas agrícolas dentro de macro categorías.

Para la detección de distintos cultivos es común utilizar índices espectrales de vegetación ya que aumenta las posibilidades de obtener una clasificación precisa [83], [84]. En varios estudios se han probado distintos índices de vegetación, como el NDVI y el índice de vegetación ajustado al suelo (SAVI), entre otros [85]. Precisamente, el NDVI es uno de los índices más utilizados; por ejemplo, se ha utilizado para la estimación del índice del área foliar de manglares [86], para determinar la dinámica ecológica de vegetación forestal [87], entre otros múltiples estudios alrededor del mundo. En este estudio se utilizaron el NDVI, GNDVI, GCI y MSI para formar una pila de índices (stack) por cada imagen y sobre estos crear los ROIs y generar las clasificaciones.

A pesar de que, con otros índices de vegetación y otros tipos de preprocesamientos, Maximum Likelihood y Random Forest han obtenido resultados excelentes en la clasificación de imágenes de Landsat-8 y Sentinel-2 [88]–[91], las clasificaciones realizadas en esta investigación, obtuvieron una precisión considerable según el coeficiente de Kappa, con porcentajes de precisión general entre 82% y 89% (según la matriz de confusión). Es necesario resaltar que, con mejores y más datos de entrenamiento, las precisiones generales pudieron haber oscilado entre 95% y 99% [63], [92].

Si bien, la precisión general de las clasificaciones fluctúa entre 82 % y 89%, la precisión individual de detección para dos (banano y palma) de los tres objetivos principales de detección fue excelente (por arriba del 96%). Por el contrario, el cacao no se detectó con

precisión en ninguna clasificación (valores de acuerdo/exactitud por debajo del 55%). Esto fue por la baja separabilidad espectral de las áreas con cultivos de cacao con la categoría “Zonas con otras coberturas”. La posible explicación consiste en que, en la zona de estudio, el cacao no es sembrado en grandes extensiones o es sembrado junto a otros cultivos. Por consecuencia, es posible afirmar que similar a otros casos [93], [94] la detección de este cultivo no fue precisa, porque la falta de muestras de campo para capacitar a los clasificadores fue la gran limitante.

Generalmente, Maximum Likelihood al igual que Random Forest, detectan de forma muy precisa áreas con cultivos agrícolas en imágenes de Landsat-8 y Sentinel-2 [50], [51], [53], [63], [88], [91], [95]. En este caso de estudio, la mejor clasificación solo para la detección de banano, cacao y palma fue la de Landsat-8 utilizando Maximum Likelihood, mientras que, la menos precisa fue la de Sentinel-2 utilizando Random Forest; aunque, RF-S2 en precisión general (englobando agua, uso urbano, sin vegetación y otras coberturas) fue la más precisa de todas y RF-L8 obtuvo una precisión general mejor que ML-L8. La baja precisión de RF-S2 para la detección de los tres cultivos de interés, se debe a dos razones principales. La primera razón fue por la temporada de la imagen de Sentinel-2 (estación lluviosa), lo que influye notoriamente en la precisión de detección de cultivos [96], [97] . La segunda razón fue por la falta de datos de campo para cultivos de cacao (como se mencionó previamente) [93], [94].

Esta potencial mejora (en la precisión general) de Random Forest frente a Maximum Likelihood es importante a tener en cuenta, sobre todo en las clasificaciones de la imagen de Sentinel-2, en donde se demostraron diferencias estadísticamente significativas. Por lo tanto, se sugiere que, en próximas investigaciones, se muestreen más y mejores coordenadas *in situ* de zonas con cultivos de cacao; además de, utilizar imágenes de Sentinel-2 en época seca, porque se presume que, la baja precisión en la detección de los cultivos objetivo en Sentinel-2 fue por la temporada de adquisición de la imagen (época lluviosa). Esto, con la finalidad de mejorar las condiciones favorables para los clasificadores y cuantificar la capacidad de mejora real del clasificador Random Forest frente a Maximum Likelihood en la clasificación de imágenes de Sentinel-2.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones.

- La menor resolución espacial del stack de índices calculados de Landsat-8 (30 m) influyó negativamente en la separabilidad espectral de los ROIs, por lo que, el de Sentinel-2 (mayor resolución espacial) proporcionó mejores resultados en este aspecto. En ese mismo sentido, las diferencias de valores en los índices calculados fue significativa por las diferencias estacionales de las imágenes (Landsat-8 temporada seca y Sentinel-2 temporada húmeda); en temporada lluviosa aumenta la vegetación, la actividad fotosintética y la clorofila (menos estrés hídrico de la vegetación). Esta diferencia en los valores de los índices influyó negativamente en la precisión de las clasificaciones; por lo tanto, es posible afirmar que, altos niveles de NDVI, GNDVI y GCI reducen la precisión en la clasificación de banano, cacao y palma utilizando imágenes de Landsat-8 y Sentinel-2.
- Las clasificaciones de Landsat-8 con ML y RF, así como las de Sentinel-2 con ambos clasificadores, obtuvieron resultados similares entre sí, aunque con ciertos matices. En las clasificaciones de Landsat-8 y en las de Sentinel-2, la categoría con mayor extensión espacial fue “Zonas con otras coberturas”, aunque, con Landsat-8 la menor fue “Zonas con cuerpos de agua” y con Sentinel-2 fue “Zonas con cultivos de palma”; esto se debió a la presencia de zonas inundadas en la imagen de Sentinel-2 (temporada lluviosa), lo que aumentó la extensión espacial de dicha categoría en la clasificación. Por otro lado, RF disminuyó (al comparar con ML) la extensión de la categoría “Zonas con cultivos de palma”, tanto en la imagen de Landsat-8 como en la de Sentinel-2; además, aumentó la cobertura de banano y cacao en Landsat-8 y redujo la extensión de las mismas categorías en las clasificaciones de Sentinel-2.
- Con las condiciones del presente proyecto de investigación, las zonas con cultivos de cacao no se pueden detectar de forma precisa en imágenes de Sentinel-2 y Landsat-8 utilizando ML y RF. Por el contrario, las zonas con cultivos de palma (precisión individual $> 96.17\%$ con ML-L8, ML-S2 y RF-L8) y banano (precisión individual $> 96.58\%$ con ML-L8 y RF-L8) se pueden detectar con precisiones

excelentes . De forma general, para clasificar todas las coberturas analizadas, la mejor combinación es Random Forest con imagen de Sentinel-2, mientras que, la combinación menos recomendada es Maximum Likelihood con Sentinel-2. En relación a las imágenes, Landsat-8 presentó mejores resultados, no obstante, su diferencia estacional con Sentinel-2 fue una condicionante importante; además, la clasificación de Sentinel-2 con Random Forest fue la mejor (de forma general según la matriz de confusión), lo que evidencia su gran potencial de clasificación.

- El análisis estadístico respaldó los resultados evidenciados en las matrices de confusión y en los coeficientes de Kappa. Las ajustadas diferencias en las precisiones de clasificación en Landsat-8 coincidieron con el análisis estadístico (sin diferencias significativas). Por otro lado, hubo amplias diferencias de precisión en las clasificaciones de Sentinel-2 según las matrices de confusión y los coeficientes de Kappa; el análisis con la prueba de Wilcoxon-Mann-Whitney sugiere que dichas diferencias son estadísticamente significativas.

5.2 Recomendaciones.

- Es recomendable que la imagen multiespectral a utilizar para la detección de cultivos agrícolas tenga una resolución espacial mayor a 30 m y no contenga nubes sobre el área a trabajar, aunque estas sean relativamente escasas. A mayor resolución espacial de la imagen multiespectral, mayor capacidad para definir buenas regiones de entrenamiento para los clasificadores, lo que aumenta la probabilidad de tener mayor separabilidad espectral y clasificaciones más precisas. Por otro lado, la presencia de nubes es susceptible a presentar baja separabilidad espectral con asentamientos humanos.
- Con coberturas complicadas de identificar mediante Google Earth, con áreas relativamente pequeñas (poco representativas) y que comúnmente suelen estar mezcladas, resulta fundamental la toma de muestras (coordenadas) en campo, con la finalidad de mejorar los ROIs y aumentar la separabilidad espectral para obtener clasificaciones más precisas.

CAPÍTULO VI

BIBLIOGRAFÍA

6.1 Bibliografía.

- [1] L. Miguel Ayala *et al.*, “Impact of agricultural expansion on water footprint in the Amazon under climate change scenarios,” *Sci. Total Environ.*, vol. 569–570, pp. 1159–1173, Nov. 2016, doi: 10.1016/j.scitotenv.2016.06.191.
- [2] X. Leng, X. Feng, and B. Fu, “Driving forces of agricultural expansion and land degradation indicated by Vegetation Continuous Fields (VCF) data in drylands from 2000 to 2015,” *Glob. Ecol. Conserv.*, vol. 23, p. e01087, Sep. 2020, doi: 10.1016/j.gecco.2020.e01087.
- [3] J. Volante *et al.*, “Expansión agrícola en Argentina, Bolivia, Paraguay, Uruguay y Chile. Caracterización espacial mediante series temporales de índices de vegetación,” *Rev. Investig. Agropecu.*, vol. 41, no. 2, pp. 179–191, 2015, [Online]. Available: <http://www.redalyc.org/articulo.oa?id=86441580011>.
- [4] G. Salmoral, K. Khatun, F. Llive, and C. M. Lopez, “Agricultural development in Ecuador: A compromise between water and food security?,” *J. Clean. Prod.*, vol. 202, pp. 779–791, Nov. 2018, doi: 10.1016/j.jclepro.2018.07.308.
- [5] Ministerio de Agricultura Ganaderia Acuacultura y Pesca MAGAP, *La politica agropecuaria ecuatoriana. Hacia el desarrollo territorial rural sostenible 2015-2025*. 2016.
- [6] F. Cánovas-García, F. Alonso, and F. Castillo, “Modificación del algoritmo Random Forest para su empleo en clasificación de imágenes de teledetección,” 2016, pp. 359–368.
- [7] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data*, Second. CRC Press, 2016.
- [8] H. Ngoma, J. Pelletier, B. P. Mulenga, and M. Subakanya, “Climate-smart agriculture, cropland expansion and deforestation in Zambia: Linkages, processes and drivers,” *Land use policy*, vol. 107, p. 105482, Aug. 2021, doi: 10.1016/j.landusepol.2021.105482.
- [9] ONU, “Transforming our world: The 2030 Agenda for Sustainable Development.”

United Nations, p. 41, 2015, [Online]. Available: sustainabledevelopment.un.org.

- [10] F. Sporchia, O. Taherzadeh, and D. Caro, “Stimulating environmental degradation: A global study of resource use in cocoa, coffee, tea and tobacco supply chains,” *Curr. Res. Environ. Sustain.*, vol. 3, p. 100029, 2021, doi: 10.1016/j.crsust.2021.100029.
- [11] L. Coltro and T. U. Karaski, “Environmental indicators of banana production in Brazil: Cavendish and Prata varieties,” *J. Clean. Prod.*, vol. 207, pp. 363–378, Jan. 2019, doi: 10.1016/j.jclepro.2018.09.258.
- [12] K. Rudolf, N. Hennings, M. A. Dippold, E. Edison, and M. Wollni, “Improving economic and environmental outcomes in oil palm smallholdings: The relationship between mulching, soil properties and yields,” *Agric. Syst.*, vol. 193, p. 103242, Oct. 2021, doi: 10.1016/j.agsy.2021.103242.
- [13] A. Castellanos-Navarrete, “Oil palm dispersal into protected wetlands: Human–environment dichotomies and the limits to governance in southern Mexico,” *Land use policy*, vol. 103, p. 105304, Apr. 2021, doi: 10.1016/j.landusepol.2021.105304.
- [14] L. Brenner, “Los impactos ambientales de las políticas públicas en los manglares de Chiapas, México,” *Gestión y Política Pública*, vol. 27, no. 1, pp. 237–267, 2018.
- [15] GADM, “Plan de Desarrollo y Ordenamiento Territorial del cantón Baba.” Gobierno Autónomo Descentralizado Municipal de Baba, p. 198, 2014.
- [16] GADM, “Plan de Desarrollo y Ordenamiento Territorial del cantón Mocache.” Gobierno Autónomo Descentralizado Municipal de Mocache, Mocache, 2020.
- [17] GADM, “Plan de Desarrollo y Ordenamiento Territorial del cantón Palenque.” Gobierno Autónomo Descentralizado Municipal de Palenque, Palenque, p. 181, 2014.
- [18] GADM, “Plan de Desarrollo y Ordenamiento Territorial del cantón Vines.” Gobierno Autónomo Descentralizado Municipal de Vines, Vines, p. 249, 2015.
- [19] GADPR, “Plan de Desarrollo y Ordenamiento Territorial 2015-2019.” Gobierno Autónomo Descentralizado Provincial de Los Ríos, p. 300, 2015.

- [20] S. F. Elbeih, “Evaluation of agricultural expansion areas in the Egyptian deserts: A review using remote sensing and GIS,” *Egypt. J. Remote Sens. Sp. Sci.*, Oct. 2021, doi: 10.1016/j.ejrs.2021.10.004.
- [21] O. Viteri-Salazar and L. Toledo, “The expansion of the agricultural frontier in the northern-Amazon region of Ecuador, 2000–2011: Process, causes, and impact,” *Land use policy*, vol. 99, p. 104986, Dec. 2020, doi: 10.1016/j.landusepol.2020.104986.
- [22] S. Aksoy, A. Yildirim, T. Gorji, N. Hamzehpour, A. Tanik, and E. Sertel, “Assessing the performance of machine learning algorithms for soil salinity mapping in Google Earth Engine platform using Sentinel-2A and Landsat-8 OLI data,” *Adv. Sp. Res.*, Oct. 2021, doi: 10.1016/j.asr.2021.10.024.
- [23] M. García, J. Évora, and M. Arbelo, *Satélites de Teledetección para la Gestión del Territorio*. Canarias: Litografía Romero, 2012.
- [24] M. Benabent Fernández de Córdoba and L. Vivanco Cruz, “The experience of the Municipal Spatial Planning in Ecuador,” *Estoa*, vol. 8, no. 15, pp. 133–144, 2019, doi: 10.18537/est.v008.n015.a11.
- [25] R. Martínez, “Sistemas de producción agrícola sostenible,” *Tecnol. en Marcha*, vol. 22, no. 2, pp. 23–39, 2009.
- [26] X. Chen, Y. Liu, F. Li, X. Li, and X. Jia, “Remote sensing image recommendation based on spatial–temporal embedding topic model,” *Comput. Geosci.*, Dec. 2021, doi: 10.1016/j.cageo.2021.104935.
- [27] T. Pei *et al.*, “GIScience and remote sensing in natural resource and environmental research: Status quo and future perspectives,” *Geogr. Sustain.*, vol. 2, no. 3, pp. 207–215, Sep. 2021, doi: 10.1016/j.geosus.2021.08.004.
- [28] J. Li and Z. Liu, “Multispectral Transforms Using Convolution Neural Networks for Remote Sensing Multispectral Image Compression,” *Remote Sens.*, vol. 11, no. 7, p. 759, Mar. 2019, doi: 10.3390/rs11070759.
- [29] S. Vura, P. Patil, and S. B. Patil, “A study of different compression algorithms for

- multispectral images,” *Mater. Today Proc.*, Jun. 2021, doi: 10.1016/j.matpr.2021.06.175.
- [30] D. Rodríguez, S. Noela, J. Domínguez, and C. Santa Marta, *Cuestiones de Teledetección*, Primera. Madrid: Universidad Nacional de Educación a Distancia Madrid 2015, 2015.
- [31] Y. Yang, Y. Chen, K. Yang, J. Cermak, and Y. Chen, “High-resolution aerosol retrieval over urban areas using sentinel-2 data,” *Atmos. Res.*, vol. 264, p. 105829, Dec. 2021, doi: 10.1016/j.atmosres.2021.105829.
- [32] A. Shebl, M. Abdellatif, M. Hissen, M. Ibrahim Abdelaziz, and Á. Csámer, “Lithological mapping enhancement by integrating Sentinel 2 and gamma-ray data utilizing support vector machine: A case study from Egypt,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 105, p. 102619, Dec. 2021, doi: 10.1016/j.jag.2021.102619.
- [33] ESA, “Sentinel-2,” *Sentinel-2*. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>.
- [34] I. Sola *et al.*, “Assessment of atmospheric correction methods for Sentinel-2 images in Mediterranean landscapes,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 73, pp. 63–76, Dec. 2018, doi: 10.1016/j.jag.2018.05.020.
- [35] Q. Qin, D. Xu, L. Hou, B. Shen, and X. Xin, “Comparing vegetation indices from Sentinel-2 and Landsat 8 under different vegetation gradients based on a controlled grazing experiment,” *Ecol. Indic.*, vol. 133, p. 108363, Dec. 2021, doi: 10.1016/j.ecolind.2021.108363.
- [36] D. P. Roy *et al.*, “Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity,” *Remote Sens. Environ.*, vol. 185, pp. 57–70, Nov. 2016, doi: 10.1016/j.rse.2015.12.024.
- [37] EOS, “Earth Observing System,” *Landsat 8*, 2013. <https://eos.com/es/find-satellite/landsat-8/>.
- [38] ESRI, “Análisis espacial,” *¿Qué es la clasificación de imagen?*, 2019. <https://desktop.arcgis.com/es/arcmap/10.7/extensions/spatial-analyst/image->

classification.

- [39] M. Misra, D. Kumar, and S. Shekhar, “Assessing Machine Learning Based Supervised Classifiers For Built-Up Impervious Surface Area Extraction From Sentinel-2 Images,” *Urban For. Urban Green.*, vol. 53, Aug. 2020, doi: 10.1016/j.ufug.2020.126714.
- [40] N. T. Ha, M. Manley-Harris, T. D. Pham, and I. Hawes, “A Comparative Assessment of Ensemble-Based Machine Learning and Maximum Likelihood Methods for Mapping Seagrass Using Sentinel-2 Imagery in Tauranga Harbor, New Zealand,” *Remote Sens.*, vol. 12, no. 3, p. 355, Jan. 2020, doi: 10.3390/rs12030355.
- [41] G. Rodríguez and R. Chica, “Clasificación de imágenes de satélite mediante software libre: Nuevas tendencias en algoritmos de Inteligencia Artificial,” Universidad de Granada, 2012.
- [42] O. Díaz, “Índices espectrales - NDWI,” 2021. <https://n9.cl/k2mr4>.
- [43] M. N. U. Sarker, F. Ulfat, M. Z. A. Siddique, and T. Farah, “Unsupervised Learning-based Vegetation change detection with Landsat 8 data,” *IEEE Xplore*, pp. 368–372, Mar. 2021, doi: 10.1109/ICAIS50930.2021.9395869.
- [44] I. C. Rhew, A. Vander Stoep, A. Kearney, N. L. Smith, and M. D. Dunbar, “Validation of the Normalized Difference Vegetation Index as a Measure of Neighborhood Greenness,” *Ann. Epidemiol.*, vol. 21, no. 12, pp. 946–952, Dec. 2011, doi: 10.1016/j.annepidem.2011.09.001.
- [45] W. Chen *et al.*, “Possibility of using multiscale normalized difference vegetation index data for the assessment of total suspended solids (TSS) concentrations in surface water: A specific case of scale issues in remote sensing,” *Environ. Res.*, vol. 194, p. 110636, Mar. 2021, doi: 10.1016/j.envres.2020.110636.
- [46] ESRI, “Green Normalized Difference Vegetation Index (GNDVI).” <https://pro.arcgis.com/es/pro-app/2.7/arcpy/image-analyst/gnvdi.htm>.
- [47] A. Dieser, “Auravant,” *¿Cuál es la diferencia entre el índice NDVI y GNDVI?*, 2022. <https://help.auravant.com/>.

- [48] Gis&Beers, “Cálculo del índice de clorofila GCI,” 2019. <http://www.gisandbeers.com/calculo-del-indice-de-clorofila-gci/>.
- [49] P. Welikhe, J. E. Quansah, S. Fall, and W. McElhenney, “Estimation of Soil Moisture Percentage Using LANDSAT-based Moisture Stress Index,” *J. Remote Sens. GIS*, vol. 06, no. 01, 2017, doi: 10.4172/2469-4134.1000200.
- [50] F. Santiago Bazan, H. Mallqui Meza, and R. Rios Recra, “Mapeo de la cobertura vegetal en la subcuenca Quillcay (Ancash – Perú) con el clasificador de Árbol de decisiones,” *Aporte Santiaguino*, vol. 14, no. 1, pp. 78–91, Jun. 2021, doi: 10.32911/as.2021.v14.n1.761.
- [51] X.-P. Song, W. Huang, M. C. Hansen, and P. Potapov, “An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping,” *Sci. Remote Sens.*, vol. 3, p. 100018, Jun. 2021, doi: 10.1016/j.srs.2021.100018.
- [52] B. Ai, K. Huang, J. Zhao, S. Sun, Z. Jian, and X. Liu, “Comparison of Classification Algorithms for Detecting Typical Coastal Reclamation in Guangdong Province with Landsat 8 and Sentinel 2 Images,” *Remote Sens.*, vol. 14, no. 2, p. 385, Jan. 2022, doi: 10.3390/rs14020385.
- [53] M. Campos-Taberner, F. J. García-Haro, B. Martínez, and M. A. Gilabert, “Deep learning para la clasificación de usos de suelo agrícola con Sentinel-2,” *Rev. Teledetección*, no. 56, p. 35, Nov. 2020, doi: 10.4995/raet.2020.13337.
- [54] USGS, “Earth Explorer.” <https://earthexplorer.usgs.gov/>.
- [55] L. Congedo, “Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS,” *J. Open Source Softw.*, vol. 6, no. 64, p. 3172, Aug. 2021, doi: 10.21105/joss.03172.
- [56] M. R. three decade assessment of forest cover changes in N. dwip national park using remote sensing and G. Islam, M. N. I. Khan, M. Z. Khan, and B. Roy, “A three decade assessment of forest cover changes in Nijhum dwip national park using remote sensing and GIS,” *Environ. Challenges*, vol. 4, Aug. 2021, doi: 10.1016/j.envc.2021.100162.

- [57] V. Belenok, T. Noszczyk, L. Hebryn-Baidy, and S. Kryachok, “Investigating anthropogenically transformed landscapes with remote sensing,” *Remote Sens. Appl. Soc. Environ.*, vol. 24, Nov. 2021, doi: 10.1016/j.rsase.2021.100635.
- [58] B. Moraes Rocha, A. Ueslei da Fonseca, H. Pedrini, and F. Soares, “Automatic detection and evaluation of sugarcane planting rows in aerial images,” *Inf. Process. Agric.*, Apr. 2022, doi: 10.1016/j.inpa.2022.04.003.
- [59] H. J. O. Al-Masaodi and H. A. M. Al-Zubaidi, “Spatial-temporal changes of land surface temperature and land cover over Babylon Governorate, Iraq,” *Mater. Today Proc.*, May 2021, doi: 10.1016/j.matpr.2021.05.179.
- [60] X. Tian, M. Zhang, C. Yang, and J. Ma, “FusionNDVI: A Computational Fusion Approach for High-Resolution Normalized Difference Vegetation Index,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5258–5271, Jun. 2021, doi: 10.1109/TGRS.2020.3014698.
- [61] J. Bian, A. Li, G. Lei, Z. Zhang, and X. Nan, “Global high-resolution mountain green cover index mapping based on Landsat images and Google Earth Engine,” *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 63–76, Apr. 2020, doi: 10.1016/j.isprsjprs.2020.02.011.
- [62] EOS, “6 Spectral Indexes To Make Vegetation Analysis Complete,” *Agriculture*, 2019. <https://eos.com/blog/6-spectral-indexes-on-top-of-ndvi-to-make-your-vegetation-analysis-complete/>.
- [63] J. Borràs, J. Delegido, A. Pezzola, M. Pereira, G. Morassi, and G. Camps-Valls, “Clasificación de usos del suelo a partir de imágenes Sentinel-2,” *Rev. Teledetección*, no. 48, p. 55, Jun. 2017, doi: 10.4995/raet.2017.7133.
- [64] S. Howell and J. Yackel, “The Jeffries–Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data,” *Int. J. Remote Sens.*, vol. 35, no. 19, pp. 6859–6873, 2014, doi: 10.1080/01431161.2014.960614.
- [65] N. Vera, “Análisis de uso de suelo a partir de imágenes satelitales Sentinel 2 en el cantón Buena Fe, provoncia de Los Ríos,” Universidad Técnica Estatal de Quevedo,

2020.

- [66] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999.
- [67] N. Karasiak, “lennepkade/dzetsaka: Fix bug in processing provider with vector files (v3.5.1),” *Zenodo*. Zenodo, 2019, doi: 10.5281/zenodo.3463523.
- [68] A. W. Sejati, I. Buchori, S. Kurniawati, Y. C. Brana, and T. I. Fariha, “Quantifying the impact of industrialization on blue carbon storage in the coastal area of Metropolitan Semarang, Indonesia,” *Appl. Geogr.*, vol. 124, Nov. 2020, doi: 10.1016/j.apgeog.2020.102319.
- [69] V. Santarsiero, G. Nolè, A. Lanorte, B. Tucci, G. Cillis, and B. Murgante, “Remote Sensing and Spatial Analysis for Land-Take Assessment in Basilicata Region (Southern Italy),” *Remote Sens.*, vol. 14, no. 7, p. 1692, Mar. 2022, doi: 10.3390/rs14071692.
- [70] Y. Ju and G. Bohrer, “Classification of Wetland Vegetation Based on NDVI Time Series from the HLS Dataset,” *Remote Sens.*, vol. 14, no. 9, p. 2107, Apr. 2022, doi: 10.3390/rs14092107.
- [71] E. B. Palafox-Juárez, J. O. López-Martínez, J. L. Hernández-Stefanoni, and H. Hernández-Nuñez, “Impact of Urban Land-Cover Changes on the Spatial-Temporal Land Surface Temperature in a Tropical City of Mexico,” *ISPRS Int. J. Geo-Information*, vol. 10, no. 2, p. 76, Feb. 2021, doi: 10.3390/ijgi10020076.
- [72] Sentinel-Hub, “Custom Scripts,” *Collection of custom scripts*. <https://custom-scripts.sentinel-hub.com/>.
- [73] J. John, *Introductory Digital Image Processing: A Remote Sensing Perspective*. Englewood Cliffs, Nueva Jersey, EE. UU: Prentice-Hall, 1986.
- [74] J. J. Pinto-Hidalgo and J. A. Silva-Centeno, “AmazonCRIME: un conjunto de datos y punto de referencia de Inteligencia Artificial Geoespacial para la clasificación de áreas potenciales vinculadas a Crímenes Ambientales Transnacionales en la Selva Amazónica,” *Rev. Teledetección*, no. 59, pp. 1–21, Jan. 2022, doi:

10.4995/raet.2022.15710.

- [75] R. Valbuena, M. Maltamo, and P. Packalen, "Classification of forest development stages from national low-density lidar datasets: a comparison of machine learning methods," *Rev. Teledetección*, no. 45, p. 15, Feb. 2016, doi: 10.4995/raet.2016.4029.
- [76] K. Rouibah and M. Belabbas, "Applying Multi-Index approach from Sentinel-2 Imagery to Extract Urban Area in dry season (Semi-Arid Land in North East Algeria)," *Rev. Teledetección*, no. 56, pp. 89–101, Nov. 2020, doi: 10.4995/raet.2020.13787.
- [77] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.
- [78] J. Cerda Lorca and L. Villarroel, "Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa," *Rev. Chil. Pediatr.*, vol. 79, no. 1, pp. 54–58, 2008, doi: 10.4067/s0370-41062008000100008.
- [79] I. D. Ávila-Pérez, E. Ortiz-Malavassi, C. Soto-Montoya, Y. Vargas-Solano, H. Aguilar-Arias, and C. Miller-Granados, "Evaluación de cuatro algoritmos de clasificación de imágenes satelitales Landsat-8 y Sentinel-2 para la identificación de cobertura boscosa en paisajes altamente fragmentados en Costa Rica," *Rev. Teledetección*, no. 57, pp. 37–49, Dec. 2020, doi: 10.4995/raet.2020.13340.
- [80] A. Stammas, A. Hassan, G. Spanos, and Z. Touloudi, "Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy." *Mail*, p. 78, 2021.
- [81] Z. Smida, L. Cucala, A. Gannoun, and G. Durif, "A Wilcoxon-Mann-Whitney spatial scan statistic for functional data," *Comput. Stat. Data Anal.*, vol. 167, Mar. 2022, doi: 10.1016/j.csda.2021.107378.
- [82] R. Sánchez, "Prueba de Wilcoxon-Mann-Whitney: mitos y realidades," *Rev. Mex. Endocrinol. Metab. y Nutr.*, vol. 2, pp. 18–21, 2015, [Online]. Available: <http://www.endocrinologia.org.mx/>.

- [83] S. X. Fei, C. H. Shan, and G. Z. Hua, "Remote Sensing of Mangrove Wetlands Identification," *Procedia Environ. Sci.*, vol. 10, pp. 2287–2293, 2011, doi: 10.1016/j.proenv.2011.09.357.
- [84] D. Denis Ávila, E. A. Curbelo, L. J. Madrigal-Roca, and R. D. Pérez-Lanyau, "Variación espacio-temporal de la respuesta espectral en manglares de La Habana, Cuba, evaluada con sensores remotos," *Rev. Biol. Trop.*, vol. 68, no. 1, pp. 321–335, 2020.
- [85] C. Munyati, "Detecting the distribution of grass aboveground biomass on a rangeland using Sentinel-2 MSI vegetation indices," *Adv. Sp. Res.*, vol. 69, no. 2, pp. 1130–1145, Jan. 2022, doi: 10.1016/j.asr.2021.10.048.
- [86] X. Guo, M. Wang, M. Jia, and W. Wang, "Estimating mangrove leaf area index based on red-edge vegetation indices: A comparison among UAV, WorldView-2 and Sentinel-2 imagery," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 103, p. 102493, Dec. 2021, doi: 10.1016/j.jag.2021.102493.
- [87] R. Prăvălie *et al.*, "NDVI-based ecological dynamics of forest vegetation and its relationship to climate change in Romania during 1987–2018," *Ecol. Indic.*, vol. 136, p. 108629, Mar. 2022, doi: 10.1016/j.ecolind.2022.108629.
- [88] M. Z. Ali, W. Qazi, and N. Aslam, "A comparative study of ALOS-2 PALSAR and landsat-8 imagery for land cover classification using maximum likelihood classifier," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 21, pp. S29–S35, Jul. 2018, doi: 10.1016/j.ejrs.2018.03.003.
- [89] A. Axelsson, E. Lindberg, H. Reese, and H. Olsson, "Tree species classification using Sentinel-2 imagery and Bayesian inference," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 100, Aug. 2021, doi: 10.1016/j.jag.2021.102318.
- [90] B. R. Shivakumar and S. V. Rajashekararadhya, "Investigation on land cover mapping capability of maximum likelihood classifier: A case study on North Canara, India," *Procedia Comput. Sci.*, vol. 143, pp. 579–586, 2018, doi: 10.1016/j.procs.2018.10.434.
- [91] G. Ge, Z. Shi, Y. Zhu, X. Yang, and Y. Hao, "Land use/cover classification in an

- arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms,” *Glob. Ecol. Conserv.*, vol. 22, Jun. 2020, doi: 10.1016/j.gecco.2020.e00971.
- [92] R. Sharma, K. Hara, and R. Tateishi, “High-Resolution Vegetation Mapping in Japan by Combining Sentinel-2 and Landsat 8 Based Multi-Temporal Datasets through Machine Learning and Cross-Validation Approach,” *Land*, vol. 6, no. 3, p. 50, Jul. 2017, doi: 10.3390/land6030050.
- [93] M. Weiss, F. Jacob, and G. Duveiller, “Remote sensing for agricultural applications: A meta-review,” *Remote Sens. Environ.*, vol. 236, p. 111402, Jan. 2020, doi: 10.1016/j.rse.2019.111402.
- [94] M. E. D. Chaves, M. C. A. Picoli, and I. D. Sanches, “Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review,” *Remote Sens.*, vol. 12, no. 18, p. 3062, Sep. 2020, doi: 10.3390/rs12183062.
- [95] S. Pareeth, P. Karimi, M. Shafiei, and C. De Fraiture, “Mapping Agricultural Landuse Patterns from Time Series of Landsat 8 Using Random Forest Based Hierarchical Approach,” *Remote Sens.*, vol. 11, no. 5, p. 601, Mar. 2019, doi: 10.3390/rs11050601.
- [96] M. I. Rodríguez-Valero and F. Alonso-Sarria, “Clasificación de imágenes Landsat-8 en la Demarcación Hidrográfica del Segura,” *Rev. Teledetección*, no. 53, pp. 33–44, Jun. 2019, doi: 10.4995/raet.2019.11016.
- [97] F. Gomariz-Castillo, F. Alonso-Sarría, and F. Cánovas-García, “Improving Classification Accuracy of Multi-Temporal Landsat Images by Assessing the Use of Different Algorithms, Textural and Ancillary Information for a Mediterranean Semiarid Area from 2000 to 2015,” *Remote Sens.*, vol. 9, no. 10, p. 1058, Oct. 2017, doi: 10.3390/rs9101058.

CAPÍTULO VII

ANEXOS

Anexo 1

Metadatos completos de la imagen adquirida de Landsat-8

Landsat Product Identifier	LC08_L1TP_011061_20161127_20180130_01_T1
Landsat Scene Identifier	LC80110612016332LGN02
Acquisition Date	27/11/2016
Collection Category	T1
Collection Number	1
WRS Path	11
WRS Row	61
Target WRS Path	11
Target WRS Row	61
Nadir/Off Nadir	NADIR
Roll Angle	-0.001
Date L-1 Generated	30/01/2018
Start Time	2016:332:15:32:57.4016260
Stop Time	2016:332:15:33:29.1716230
Station Identifier	LGN
Land Cloud Cover	8.99
Scene Cloud Cover	10.72
Ground Control Points Model	156
Ground Control Points Version	4
Geometric RMSE Model (meters)	8.516
Geometric RMSE Model X	6.697
Geometric RMSE Model Y	5.260
Image Quality	9
Processing Software Version	LPGS_13.0.0
Sun Elevation L1	5.934.443.739
Sun Azimuth L1	13.200.750.888
TIRS SSM Model	FINAL
Data Type Level-1	OLI_TIRS_L1TP
Sensor Identifier	OLI_TIRS
Panchromatic Lines	15481
Panchromatic Samples	15181
Reflective Lines	7741
Reflective Samples	7591
Thermal Lines	7741
Thermal Samples	7591
Map Projection Level-1	UTM
UTM Zone	17
Datum	WGS84
Ellipsoid	WGS84
Grid Cell Size Panchromatic	15.00
Grid Cell Size Reflective	30.00

Grid Cell Size Thermal	30.00
Bias Parameter File Name OLI	LO8BPF20161127151527_20161127155742.01
Bias Parameter File Name TIRS	LT8BPF20161114203648_20161203213854.01
Calibration Parameter File	LC08CPF_20161001_20161231_01.01
RLUT File Name	LC08RLUT_20150303_20431231_01_12.h5
Center Latitude	1°26'47.22"S
Center Longitude	80°20'29.33"W
UL Corner Lat	0°23'56.33"S
UL Corner Long	81°21'56.77"W
UR Corner Lat	0°23'55.72"S
UR Corner Long	79°19'10.85"W
LL Corner Lat	2°29'58.96"S
LL Corner Long	81°21'57.96"W
LR Corner Lat	2°29'55.21"S
LR Corner Long	79°19'05.27"W
Center Latitude dec	-144.645
Center Longitude dec	-8.034.148
UL Corner Lat dec	-0.39898
UL Corner Long dec	-8.136.577
UR Corner Lat dec	-0.39881
UR Corner Long dec	-7.931.968
LL Corner Lat dec	-249.971
LL Corner Long dec	-8.136.610
LR Corner Lat dec	-249.867
LR Corner Long dec	-7.931.813

FUENTE: USGS Science for a changing world [54]

Anexo 2

Metadatos completos de la imagen adquirida de Sentinel-2

Data Set Attribute	Attribute Value
Entity ID	L1C_T17MPU_A005814_20180417T154048
Acquisition Start Date	2018-04-17T15:40:48.043Z
Acquisition End Date	2018-04-17T15:45:47.488Z
Tile Number	T17MPU
Cloud Cover	431.830
Agency	ESA
Platform	SENTINEL-2B
Vendor	MTI_
Vendor Tile ID	L1C_T17MPU_A005814_20180417T154048

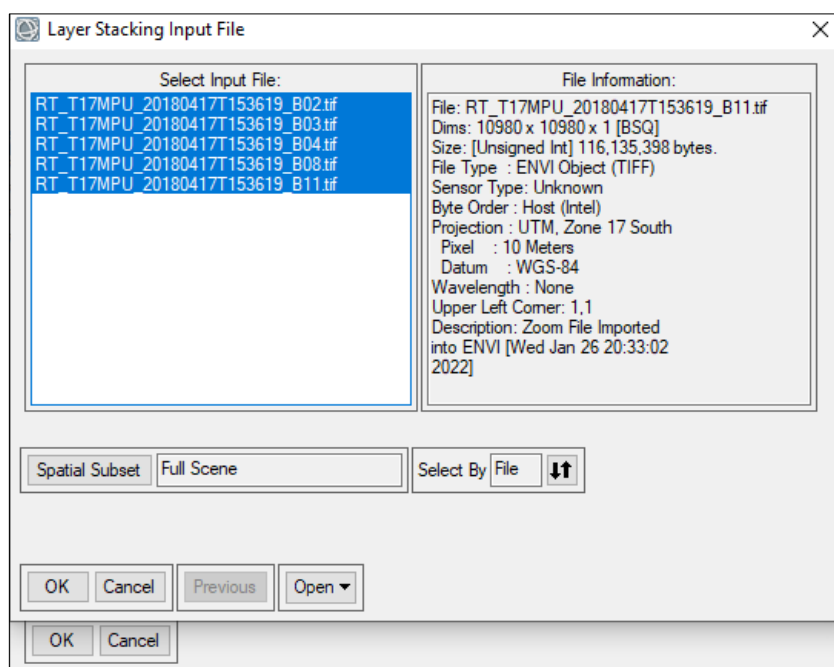
Orbit Number	68
Orbit Direction	Descending Orbit
Vendor Software Version	02.06
Production Date	2018-04-17T20:22:31.000000Z
Vendor Product ID	S2B_MSIL1C_20180417T153619_N0206_R068_T17MPU_20180417T202231
Archiving Center	MTI_
Datatake Type	INS-NOBS
Datatake Identifier	GS2B_20180417T153619_005814_N02.06
Data Type	UINT16
Product Type	S2MSI1C
Product Format	JPEG2000
Processing Level	LEVEL-1C
Datastrip ID	S2B_OPER_MSI_L1C_DS_MTI__20180417T202231_S20180417T154048_N02.06
Datum	WGS84
Map Projection	UTM
UTM Zone	17S
EPSG Code	32717
Resolution	10, 20, 60
Units	METER
Sun Zenith Angle Mean	263.580.995.492.743
Sun Azimuth Angle Mean	623.337.596.928.815
Quantification	10000
Image_Referencing	U
Center Latitude	1°24'02.33"S
Center Longitude	79°36'27.64"W
NW Corner Lat	0°54'15.32"S
NW Corner Long	80°06'04.50"W
NE Corner Lat	0°54'13.95"S
NE Corner Long	79°06'52.88"W
SE Corner Lat	1°53'48.20"S
SE Corner Long	79°06'50.03"W
SW Corner Lat	1°53'51.08"S
SW Corner Long	80°06'03.14"W
Center Latitude dec	-14.006.479
Center Longitude dec	-796.076.774
NW Corner Lat dec	-0.9042567
NW Corner Long dec	-801.012.512
NE Corner Lat dec	-0.903876
NE Corner Long dec	-791.146.897
SE Corner Lat dec	-18.967.232

SE Corner Long dec	-79.113.896
SW Corner Lat dec	-18.975.225
SW Corner Long dec	-801.008.726

FUENTE: USGS Science for a changing world [54]

Anexo 3

Unión de bandas necesarias para la imagen de Sentinel-2 con la opción “Layer Stacking”



ELABORADO: AUTOR

Anexo 4

Informe completo de separabilidad espectral de ROIs de Landsat-8

```

Input File: Landsat.tif
ROI Name: (Jeffries-Matusita, Transformed Divergence)

Asentamientos humanos:
  Banano: (2.00000000 2.00000000)
  Cacao: (1.99999721 2.00000000)
  Palma: (2.00000000 2.00000000)
  Cuerpos de agua superficiales: (1.99999996 2.00000000)
  Sin vegetación aparente: (1.99944666 2.00000000)
  Otras coberturas: (1.90394173 1.99683123)
  Nubes : (1.63593065 1.73151196)

Banano:
  Asentamientos humanos: (2.00000000 2.00000000)
  Cacao: (1.99978317 2.00000000)
  Palma: (1.86146314 1.96715800)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Sin vegetación aparente: (2.00000000 2.00000000)
  Otras coberturas: (1.95751270 2.00000000)
  Nubes : (1.99999856 2.00000000)

Cacao:
  Asentamientos humanos: (1.99999721 2.00000000)
  Banano: (1.99978317 2.00000000)
  Palma: (1.99938979 2.00000000)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Sin vegetación aparente: (2.00000000 2.00000000)
  Otras coberturas: (1.47134083 1.99673340)
  Nubes : (1.99220436 2.00000000)

Palma:
  Asentamientos humanos: (2.00000000 2.00000000)
  Banano: (1.86146314 1.96715800)
  Cacao: (1.99938979 2.00000000)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Sin vegetación aparente: (2.00000000 2.00000000)
  Otras coberturas: (1.95365374 2.00000000)
  Nubes : (1.99999984 2.00000000)

Cuerpos de agua superficiales:
  Asentamientos humanos: (1.99999996 2.00000000)
  Banano: (2.00000000 2.00000000)
  Cacao: (2.00000000 2.00000000)
  Palma: (2.00000000 2.00000000)
  Sin vegetación aparente: (2.00000000 2.00000000)
  Otras coberturas: (1.99999980 2.00000000)
  Nubes : (1.99963234 1.99999126)

Sin vegetación aparente:
  Asentamientos humanos: (1.99944666 2.00000000)
  Banano: (2.00000000 2.00000000)
  Cacao: (2.00000000 2.00000000)
  Palma: (2.00000000 2.00000000)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Otras coberturas: (1.98946556 2.00000000)
  Nubes : (1.99999791 2.00000000)

Otras coberturas:
  Asentamientos humanos: (1.90394173 1.99683123)
  Banano: (1.95751270 2.00000000)
  Cacao: (1.47134083 1.99673340)
  Palma: (1.95365374 2.00000000)
  Cuerpos de agua superficiales: (1.99999980 2.00000000)
  Sin vegetación aparente: (1.98946556 2.00000000)
  Nubes : (1.88821816 1.99131089)

Nubes :
  Asentamientos humanos: (1.63593065 1.73151196)
  Banano: (1.99999856 2.00000000)
  Cacao: (1.99220436 2.00000000)
  Palma: (1.99999984 2.00000000)
  Cuerpos de agua superficiales: (1.99963234 1.99999126)
  Sin vegetación aparente: (1.99999791 2.00000000)
  Otras coberturas: (1.88821816 1.99131089)

Pair Separation (least to most);
Cacao and Otras coberturas - 1.47134083
Asentamientos humanos and Nubes - 1.63593065
Banano and Palma - 1.86146314
Otras coberturas and Nubes - 1.88821816
Asentamientos humanos and Otras coberturas - 1.90394173
Palma and Otras coberturas - 1.95365374
Banano and Otras coberturas - 1.95751270
Sin vegetación aparente and Otras coberturas - 1.98946556
Cacao and Nubes - 1.99220436
Cacao and Palma - 1.99938979
Asentamientos humanos and Sin vegetación aparente - 1.99944666
Cuerpos de agua superficiales and Nubes - 1.99963234
Banano and Cacao - 1.99978317
Asentamientos humanos and Cacao - 1.99999721
Sin vegetación aparente and Nubes - 1.99999791
Banano and Nubes - 1.99999856
Cuerpos de agua superficiales and Otras coberturas - 1.99999980
Palma and Nubes - 1.99999984
Asentamientos humanos and Cuerpos de agua superficiales - 1.99999996
Cacao and Sin vegetación aparente - 2.00000000
Cacao and Cuerpos de agua superficiales - 2.00000000
Palma and Cuerpos de agua superficiales - 2.00000000
Banano and Cuerpos de agua superficiales - 2.00000000
Asentamientos humanos and Banano - 2.00000000
Asentamientos humanos and Palma - 2.00000000
Cuerpos de agua superficiales and Sin vegetación aparente - 2.00000000
Palma and Sin vegetación aparente - 2.00000000
Banano and Sin vegetación aparente - 2.00000000
  
```

ELABORADO: AUTOR

Anexo 5

Informe completo de separabilidad espectral de ROIs de Sentinel-2

```

Input File: Sentinel.tif
ROI Name: (Jeffries-Matusita, Transformed Divergence)

Asentamientos humanos:
  Banano: (2.00000000 2.00000000)
  Cacao: (1.99999768 2.00000000)
  Palma: (2.00000000 2.00000000)
  Cuerpos de agua superficiales: (1.99906934 1.99998795)
  Sin vegetación aparente: (1.93646725 1.99995171)
  Otras coberturas: (1.99998437 2.00000000)

Banano:
  Asentamientos humanos: (2.00000000 2.00000000)
  Cacao: (1.99998534 2.00000000)
  Palma: (1.99999294 1.99999776)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Sin vegetación aparente: (1.99999999 2.00000000)
  Otras coberturas: (1.98960358 2.00000000)

Cacao:
  Asentamientos humanos: (1.99999768 2.00000000)
  Banano: (1.99998534 2.00000000)
  Palma: (1.99987623 2.00000000)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Sin vegetación aparente: (1.99990283 2.00000000)
  Otras coberturas: (1.90092264 1.99999999)

Palma:
  Asentamientos humanos: (2.00000000 2.00000000)
  Banano: (1.99999294 1.99999776)
  Cacao: (1.99987623 2.00000000)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Sin vegetación aparente: (2.00000000 2.00000000)
  Otras coberturas: (1.99724824 2.00000000)

Cuerpos de agua superficiales:
  Asentamientos humanos: (1.99906934 1.99998795)
  Banano: (2.00000000 2.00000000)
  Cacao: (2.00000000 2.00000000)
  Palma: (2.00000000 2.00000000)
  Sin vegetación aparente: (2.00000000 2.00000000)
  Otras coberturas: (1.99981115 2.00000000)

Sin vegetación aparente:
  Asentamientos humanos: (1.93646725 1.99995171)
  Banano: (1.99999999 2.00000000)
  Cacao: (1.99990283 2.00000000)
  Palma: (2.00000000 2.00000000)
  Cuerpos de agua superficiales: (2.00000000 2.00000000)
  Otras coberturas: (1.99999522 2.00000000)

Otras coberturas:
  Asentamientos humanos: (1.99998437 2.00000000)
  Banano: (1.98960358 2.00000000)
  Cacao: (1.90092264 1.99999999)
  Palma: (1.99724824 2.00000000)
  Cuerpos de agua superficiales: (1.99981115 2.00000000)
  Sin vegetación aparente: (1.99999522 2.00000000)

Pair Separation (least to most);

Cacao and Otras coberturas - 1.90092264
Asentamientos humanos and Sin vegetación aparente - 1.93646725
Banano and Otras coberturas - 1.98960358
Palma and Otras coberturas - 1.99724824
Asentamientos humanos and Cuerpos de agua superficiales -
1.99906934
Cuerpos de agua superficiales and Otras coberturas - 1.99981115
Cacao and Palma - 1.99987623
Cacao and Sin vegetación aparente - 1.99990283
Asentamientos humanos and Otras coberturas - 1.99998437
Banano and Cacao - 1.99998534
Banano and Palma - 1.99999294
Sin vegetación aparente and Otras coberturas - 1.99999522
Asentamientos humanos and Cacao - 1.99999768
Banano and Sin vegetación aparente - 1.99999999
Palma and Sin vegetación aparente - 2.00000000
Cacao and Cuerpos de agua superficiales - 2.00000000
Asentamientos humanos and Banano - 2.00000000
Cuerpos de agua superficiales and Sin vegetación aparente -
2.00000000
Asentamientos humanos and Palma - 2.00000000
Palma and Cuerpos de agua superficiales - 2.00000000
Banano and Cuerpos de agua superficiales - 2.00000000

```

ELABORADO: AUTOR

Anexo 6

Toma de coordenadas para ROIs de zonas con coberturas de cacao



ELABORADO: AUTOR

Anexo 7

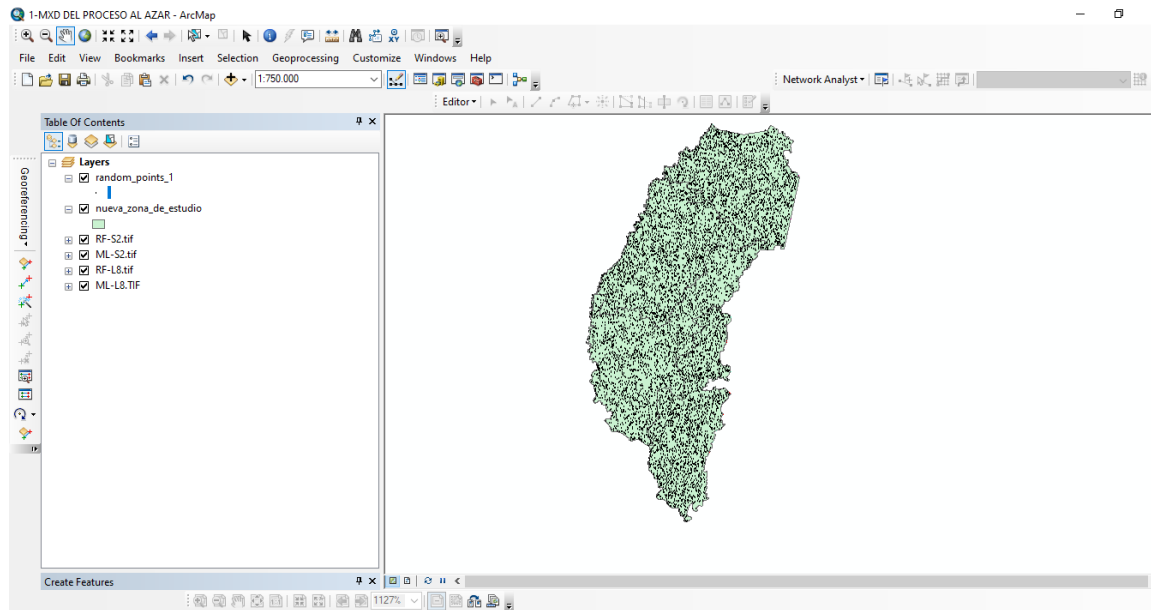
Toma de coordenadas para ROIs (de validación) de zonas con coberturas de cacao y zonas con otras coberturas



ELABORADO: AUTOR

Anexo 8

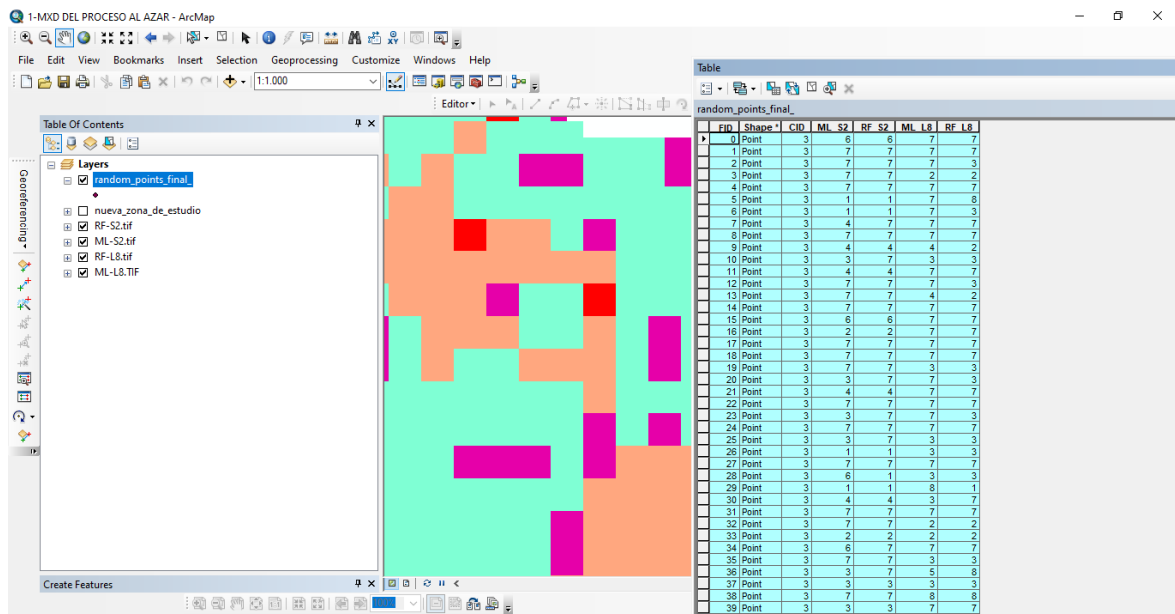
Generación de puntos al azar para análisis de diferencias estadísticas con la herramienta “Create Random Points”



ELABORADO: AUTOR

Anexo 9

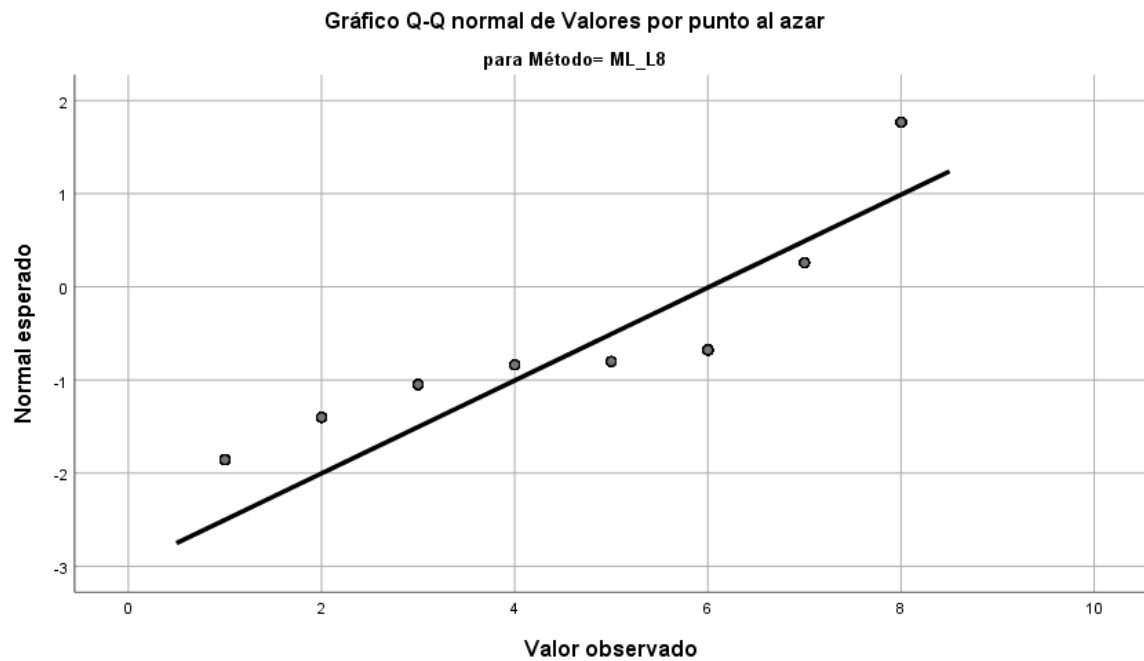
Extracción de puntos al azar para análisis de diferencias estadísticas con la herramienta “Extract Multi Values to Point”



ELABORADO: AUTOR

Anexo 10

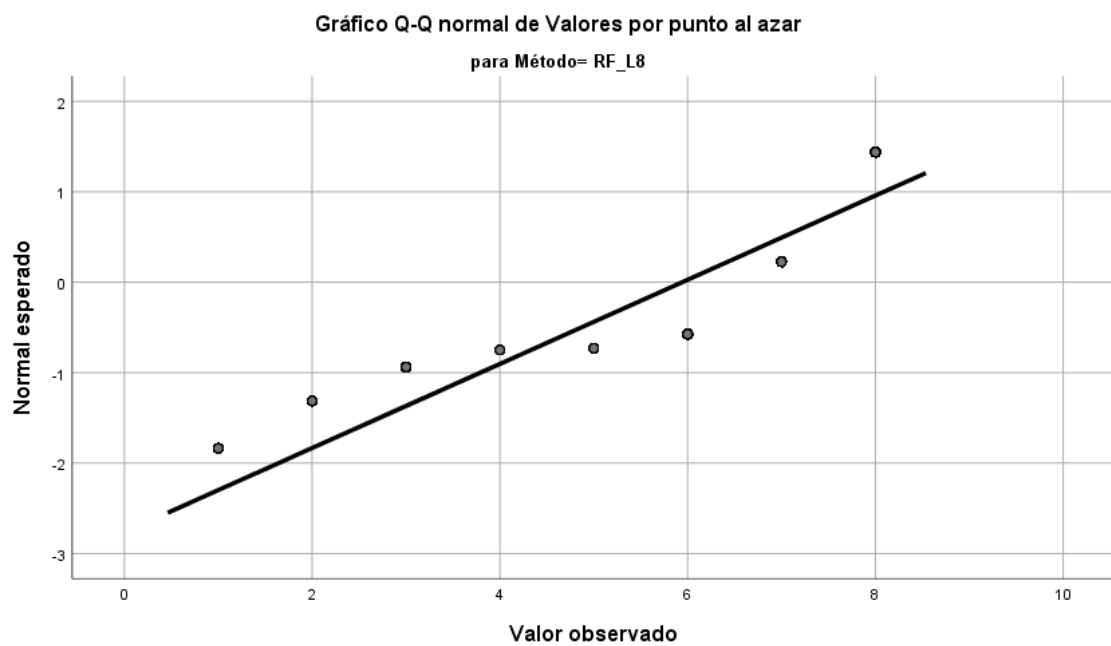
Gráfico Q-Q para análisis de normalidad de los datos de ML-L8



ELABORADO: AUTOR

Anexo 11

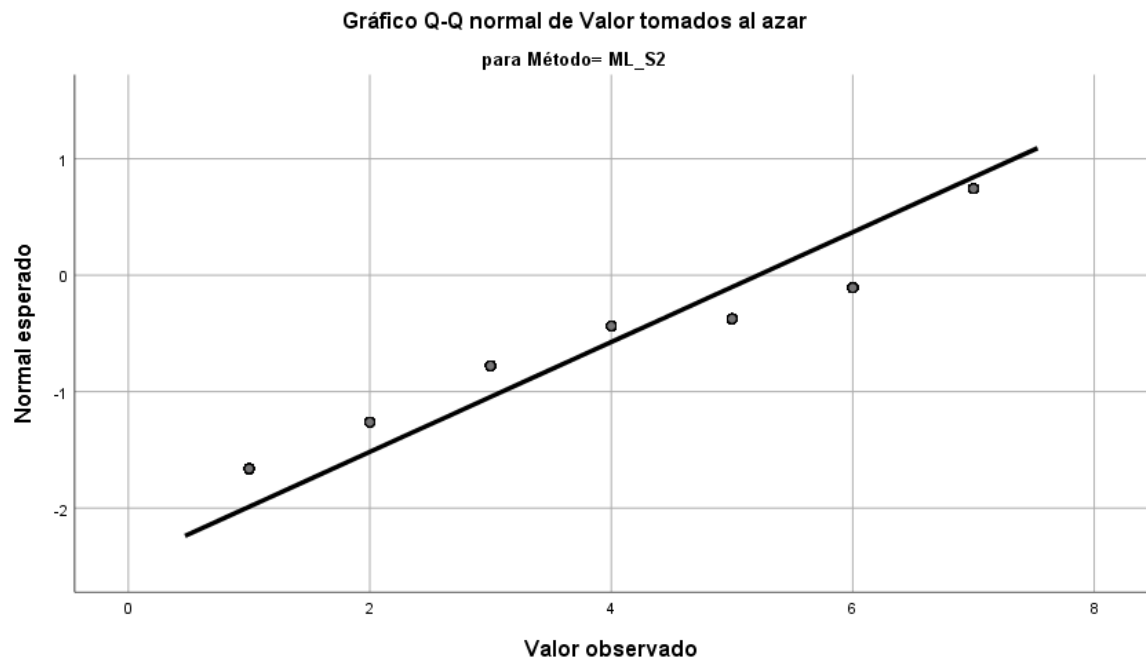
Gráfico Q-Q para análisis de normalidad de los datos de RF-L8



ELABORADO: AUTOR

Anexo 12

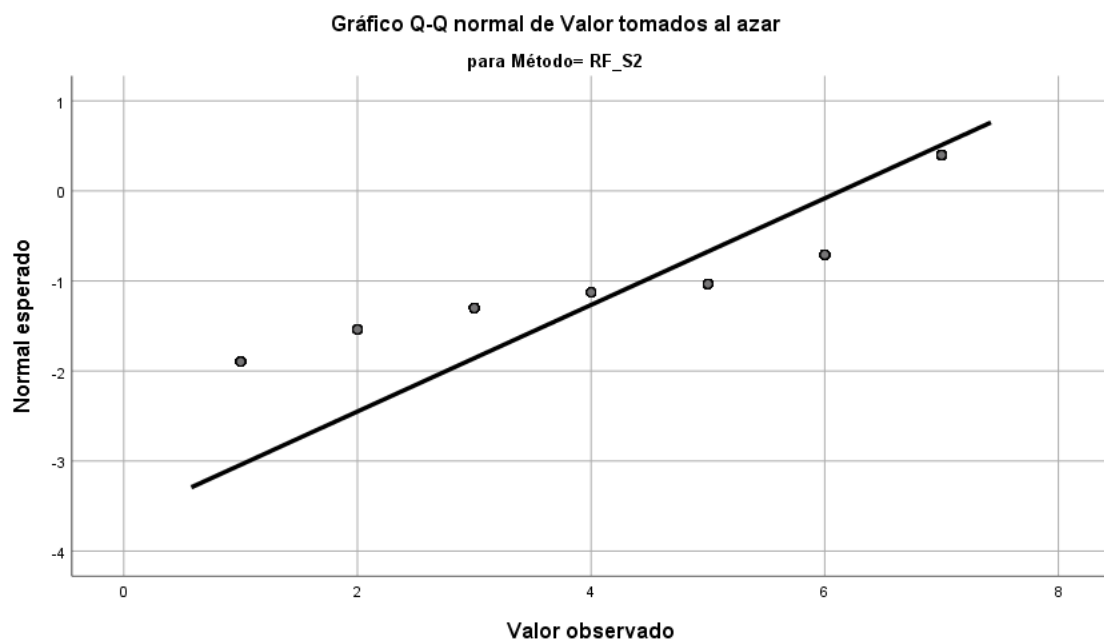
Gráfico Q-Q para análisis de normalidad de los datos de ML-S2



ELABORADO: AUTOR

Anexo 13

Gráfico Q-Q para análisis de normalidad de los datos de RF-S2



ELABORADO: AUTOR