Clasificación supervisada de imágenes de satélite en Google Earth Engine del municipio de Othón P. Blanco, Quintana Roo, México

Luis Antonio Mora-Tembre[1], Jorge Armando López Chan[2] y Cecilia Elizabeth Olivares Braus[3]

[1] *Licenciado en Biología, Estudiante de Maestría en Manejo de Zonas Costeras del TECNM/ITChetumal. Av. Insurgentes 330, Chetumal, Quintana Roo, México. C.P 77013. ORCID: 0000-0001-7648-2737 Correo E: M95390351@Chetumal.tecnm.mx*

[2] *Licenciado en Biología. Docente de la Licenciatura en Biología y Estudiante de la Maestría en Manejo de Zona Costera del TECNM/IT Chetumal. Av. Insurgentes 330, Chetumal, Quintana Roo, México. C.P 77013. ORCID: 0009-0007-5307-2429. Correo E: M95390351@Chetumal.tecnm.mx*

[3] *Ingeniera Geofísica. Candidata a obtener el grado de Maestría en Ingeniería Petrolera, estudiante de Maestría en Educación. Representante de la compañía EPYESA en Tabasco; Profesora de la Dirección de Ingeniería y Arquitectura de la Universidad Olmeca Carretera Villahermosa-Macuspana km 14, Poblado Dos Montes, C.P. 86280, Centro, Tabasco, México. ORCID 0009-0007-9592-9267. Correo E: colivares_braus@hotmail.com*

RESUMEN

El monitoreo de las coberturas terrestres es un tema de suma importancia en los ámbitos gubernamental, académico y social por las implicaciones que tienen los cambios de las coberturas y usos de suelo en temas como sustentabilidad, riesgos, conservación de la biodiversidad, servicios ambientales y el cambio climático. De acuerdo con el reporte especial del Grupo Intergubernamental de Expertos sobre el Cambio Climático del 2015, los cambios de cobertura de suelo a nivel mundial y regional de zonas por pérdida de cobertura forestal, limitan las trayectorias de mitigación global a 1.5 °C. El objetivo de la presente investigación es generar un mapa de coberturas para el municipio de Othón P. Blanco Quintana Roo, mediante el uso de imágenes satelitales y la plataforma *Google Earth Engine*, mediante el algoritmo «Random Forest». Se generó una imagen compuesta basado en la mediana de valores de pixel, del periodo de enero a diciembre del 2020, se utilizaron escenas de Sentinel-2 que abarcan los límites del municipio y se recortó por máscara para emplear la capa de los límites municipales en *Google Earth Engine*. Posteriormente se creó una malla de entrenamiento a partir de ortomosaicos obtenidos mediante vehículos aéreos No Tripulados. Para obtener la validación cuantitativa de la precisión se empleó una matriz de confusión para determinar la precisión de la predicción de clases empleadas, se determinó ejecutar la función.confusionMatrix que determina la precisión temática de un mapa de cobertura terrestre. Se obtuvo un mapa de cobertura con 16 clases, la presión de entrenamiento es de 0.99. Se recomienda realizar métodos de control en campo para revisión similar a los empleados para generar la malla de entrenamiento y corregir los valores del clasificador. El aprendizaje automático operado a partir de esta plataforma web es bastante accesible en comparación

Mora-Tembre, L. A., López-Chan, J.A., Olivares-Braus, C. E., 2023, Clasificación supervisada de imágenes de satélite en *Google Earth Engine* del municipio de Othón P. Blanco, Quintana Roo; México, Cuba: Geociencias UO. v. 6, núm. 1, enero-junio 2023. pp. 46-53.

con los flujos de trabajo tradicionales que implican contar con equipo de cómputo de alto rendimiento.

Palabras Claves: Sentinel; *Google Earth Engine*; clasificación supervisada; árbol de clasificación de bosques aleatorios; aprendizaje automático.

ABSTRACT

The monitoring of land cover is a topic of great importance in the governmental, academic and social spheres due to the implications that changes in land cover and land use have on issues such as sustainability, risks, biodiversity conservation, environmental services and the climate change. According to the special report of the Intergovernmental Panel on Climate Change of 2015, global and regional land cover changes in areas due to loss of forest cover limit global mitigation trajectories to 1.5 °C. The objective of this research is to generate a coverage map for the municipality of Othón P. Blanco Quintana Roo, through the use of satellite images and the Google Earth Engine platform, through the «Random Forest» algorithm. A composite image was generated based on median pixel values, for the period January to December 2020, using Sentinel-2 scenes spanning the city limits and clipped by mask using the City limits layer in Google Earth. Engine. A training mesh was created from orthomosaics obtained by Unmanned Aerial Vehicles, to obtain the quantitative validation of the precision, a confusion matrix was used to determine the precision of the prediction of the classes used, running the function confusionMatrix. which determines the thematic accuracy of a land cover map. A coverage map with 16 classes was obtained, the training pressure is 0.99. It is recommended to continue carrying out control methods in the field for review similar to those used to generate the training mesh and correct the values of the classifier. Machine learning powered from this web platform is quite accessible compared to traditional workflows that involve high-performance computing equipment.

Keywords: Sentinel; Google Earth Engine; supervised classification; random forest classification tree; machine learning.

RESUMO

El monitoreo de las coberturas terrestres es un tema de suma importancia en los ámbitos gubernamental, académico y social por las implicaciones que tienen los cambios de las coberturas y usos de suelo en temas como sustentabilidad, riesgos, conservación de la biodiversidad, servicios ambientales y el mudança climática. De acordo com o relatório especial do Painel Intergovernamental sobre Mudanças Climáticas de 2015, as mudanças globais e regionais na cobertura da terra em áreas devido à perda de cobertura florestal limitam as trajetórias globais de mitigação a 1.5 °C. O objetivo desta pesquisa é gerar um mapa de cobertura para o município de Othón P. Blanco Quintana Roo, através do uso de imagens de satélite e da plataforma Google Earth Engine, através do algoritmo «Random Forest». Uma imagem composta foi gerada com base nos valores médios de pixel, para o período de janeiro a dezembro de 2020, usando cenas do Sentinel-2 abrangendo os limites da cidade e recortadas por máscara usando a camada de limites da cidade no Google Earth. Engine. Uma malha de treinamento foi criada a partir de ortomosaicos obtidos por Veículos Aéreos Não Tripulados, para obter a validação quantitativa da precisão, uma matriz de confusão foi usada para determinar a precisão da predição das classes utilizadas, executando a função .confusionMatrix. que determina a precisão temática de um mapa de cobertura da terra. Obteve-se um mapa de cobertura com 16 aulas, a pressão de treinamento é de 0,99. Recomenda-se continuar realizando métodos de controle em campo para revisão semelhantes aos usados para gerar a malha de treinamento e corrigir os valores do classificador. O aprendizado de máquina alimentado por essa plataforma da Web é bastante acessível em comparação com os fluxos de trabalho tradicionais que envolvem equipamentos de computação de alto desempenho.

Palavras-chave: Sentinela; Motor Google Earth;

classificação supervisionada; árvore de classificação de floresta aleatória; aprendizado de máquina.

INTRODUCCIÓN

El monitoreo de las coberturas terrestres es un tema de suma importancia en los ámbitos gubernamental, académico y social por las implicaciones que tienen los cambios de las coberturas y usos de suelo en temas como sustentabilidad, riesgos, conservación de la biodiversidad, servicios ambientales y el cambio climático.

De acuerdo con el reporte especial del Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) del 2014, los cambios de cobertura de suelo a nivel mundial y regional de zonas por pérdida de cobertura forestal, limitan las trayectorias de mitigación global a 1.5 °C.

Los estudios territoriales, la percepción remota y los sistemas de información geográfica, al ser capaces de obtener estos datos de manera ágil y a bajo costo, se han convertido en tecnologías fundamentales. Estas tecnologías se integran a la planeación territorial de forma que permiten actualizar la base de datos conforme se observan los cambios en el territorio.

La Comisión Nacional Forestal ha implementado el sistema Satelital de Monitoreo Forestal (SAMOF) con el objetivo de procurar el manejo sustentable e integral de las superficies forestales. La elaboración de cartografía a partir de la integración de datos geoespaciales, geoprocesos, algoritmos de aprendizaje automático llevado a cabo por la validación de expertos en gabinete y campo; es complejo.

- Los procesos de levantamiento de imágenes aéreas con sistema de aeronave pilotada a distancia (RPAS) han sido abordados por diferentes autores. Martínez Beatle en el 2017 presenta un manual de procesos de levantamiento fotogramétrico con el uso de drones con grado de consumo y software de código libre como es OpenDroneMap, así como su interfaz gráfica WebODM, estas plataformas son de código libre. Presenta, por lo tanto, las bases teóricas para implementar flujos de trabajo efectivo; además de describir el funcionamiento de los algoritmos de postproceso empleados. El uso de fotografías aéreas de RPAS para la verificación de los sitios de cobertura permite enfrentar limitaciones de tiempo, de esta manera, el trabajo de campo se enfocó en áreas para la interpretación visual.

En la clasificación de imágenes se emplea el algoritmo Random Forest, (Breiman, 2001) el cual funciona a partir de un conjunto de árboles de decisiones independientes entre sí que se formulan a partir de un dato de vector aleatorio declarado previamente y se utilizan para entrenar y decidir sobre las muestras con múltiples niveles de árboles (**Figura 1**). Un conjunto de entidades de entrenamiento se clasifica a partir del conjunto de datos original mediante un método de muestreo y el tamaño de cada conjunto de entrenamiento es de una proporción aproximada de 2/3 del conjunto de datos original. Luego, se establecen árboles de clasificación para cada conjunto de entrenamiento para generar un árbol de decisiones compuesto por N árboles de clasificación. Durante el proceso de crecimiento de cada árbol, se seleccionan de manera aleatoria m atributos ($m \leq M$) de todas las M variables características y se selecciona el atributo predominante para la rama del nodo interno de acuerdo con el principio mínimo del coeficiente de Gini entre estos M atributos. Al final, se establecen los resultados de predicción de los árboles de clasificación N y la categoría de nuevas muestras se determina mediante votación. 1/3 de los datos de cada muestreo no se seleccionó y el error OOB se genera al utilizar los datos fuera del conjunto para estimar el error interno.

Para procesar grandes conjuntos de datos para aplicaciones globales se emplea *Google Earth Engine* (Gorelick *et al*., 2017) la potencia de computación en la nube de *Google Earth Engine* permite el procesamiento de petabytes de datos de imagen, combinados con otros datos vectoriales, dentro del entorno de la nube y elimina la necesidad de almacenar, procesar, y analizar los grandes volúmenes de datos satelitales en

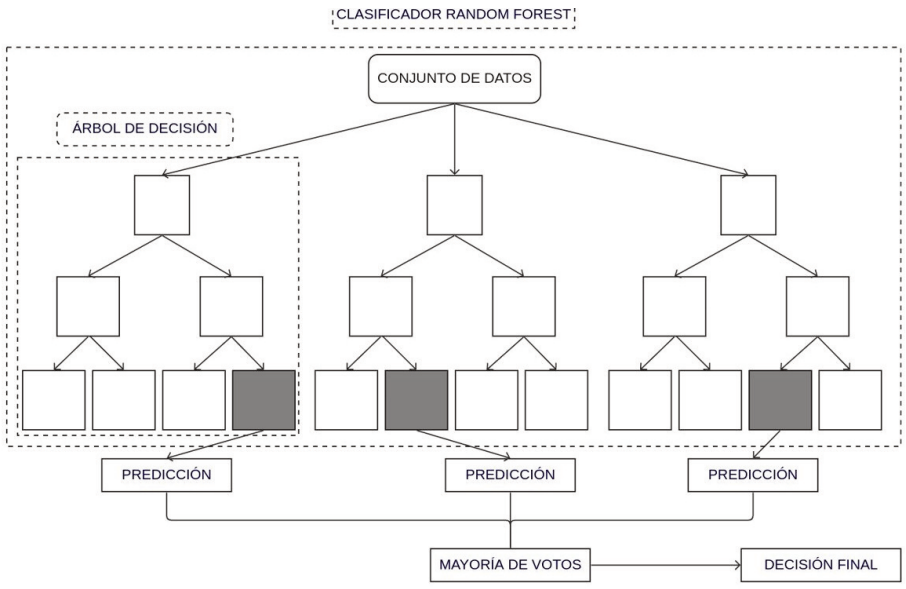

Figura 1. Diagrama del árbol de decisiones del clasificador Random forest. Tomado de Mbaadu (2022).

una computadora de oficina. para el análisis y la toma de decisiones finales (Kumar & Mutanga, 2018). El objetivo de la plataforma es reducir el tiempo invertido en el preprocesamiento y facilitar los análisis realizados con información geoespacial la plataforma es gratuita para uso académico y de investigación (Solórzona & Perilla, 2022).

El objetivo del presente documento es generar un mapa de coberturas para el municipio de Othón P. Blanco Quintana Roo, mediante el uso de imágenes satelitales y la plataforma *Google Earth Engine*.

Área de estudio

El municipio de Othón P. Blanco se encuentra al sur del estado de Quintana Roo, colinda al Este con el mar Caribe, al sur con el país de Belice y Guatemala, al oeste con el Estado de Campeche y al norte con el municipio de Bacalar. Cuenta con una extensión territorial de 1 179 017.66 hectáreas, es el municipio más grande del Estado de Quintana Roo y uno de los más grandes a nivel nacional (Pozo *et al*., 2011).

La mayor superficie del municipio es considerada una planicie costera, con algunas irregularidades que no sobrepasan pendientes de 5 %; existen humedales, cuerpos de agua permanentes y zonas inundables en los periodos de lluvia en la zona norponiente de la ciudad de Chetumal, suroriente de Subteniente López y norte de Huay-pix, por lo cual el crecimiento de la mancha urbana se orienta hacia el norte de la ciudad. El sistema principal lagunar-fluvial está formado entre la laguna de Bacalar y del Río Hondo, unida por una pequeña corriente de agua permanente conocida como el estero de Chaac. Al este de la ciudad de Chetumal se localiza la Bahía de Chetumal, cuya costa es baja y pedregosa; al sur se encuentra la desembocadura del Rio Hondo, cubierta en su mayoría por mangle y áreas pantanosas de tal manera que se limita el crecimiento urbano. De acuerdo con la elevación se conocen dos zonas: la baja, en la que se encuentra inmediatamente la Bahía con una elevación máxima de dos metros sobre el nivel del mar y se extiende tierra adentro con una elevación de entre seis a nueve metros (Pozo *et al*., 2011).

El clima es tropical cálido subhúmedo debido a la gran cantidad de ondas tropicales que acarrean gran cantidad de humedad en las estaciones de verano y principios de otoño. Durante el invierno, las masas de aire gélido se dirigen hacia el sur, un fenómeno que persiste hasta principios de marzo o incluso principios de abril. La temporada de secas más pronunciada es de abril y mayo. Las temperaturas oscilan entre los 18.8 ºC y 33.0 ºC en promedio. Las precipitaciones son de 1 100 a 1 400 mm anuales. A pesar de que se presentan fuer-

tes lluvias en verano y algunas sequías estacionales, existen precipitaciones pluviales en la mayor parte del año (Pozo *et al*., 2011).

MATERIALES Y MÉTODOS

Los tipos de vegetación identificados en el municipio Othón P. Blanco corresponden con las características de la selva alta subperennifolia, las selvas medianas subperennifolia y subcaducifolia, las selvas bajas subperennifolia y subcaducifolia, los tulares, las sábanas, los manglares y otros tipos de vegetación con menor extensión como los palmares, los petenes y la vegetación de dunas costeras (Pozo *et al*., 2011).

Para realizar la clasificación de las imágenes de satélite y garantizar la compatibilidad de las clases generadas, con los productos anteriores, se emplearon las 16 clases señaladas en los mapas de cobertura del suelo generados a través del sistema Satelital de Monitoreo Forestal y que están registradas para el municipio de Othón P. Blanco en la capa de la línea base 2016 **(Tabla 1**).

La clasificación de las imágenes se realiza en la plataforma *Google Earth Engine* con el algoritmo de aprendizaje automático, «Random Forest». Para crear la malla de entrenamiento para el algoritmo, se generó una red de puntos al interior de cada uno de los polígonos de las clases de diez puntos por cada una de las clases, separados a cinco kilómetros de distancia entre cada uno (**Figura 2**). Para esto se empleó el algoritmo «Puntos Aleatorios dentro de polígonos» del programa QGIS. De la malla de puntos generados, 20 % es empleada para la validación de las muestras de entrenamiento y el resto para generar el entrenamiento del algoritmo (Breiman, 2001).

Con la capa de la malla de entrenamiento generada, se realizaron salidas a campo para realizar la verificación de los puntos de entrenamiento mediante la generación de ortomosaicos derivados de vuelos de vehículos autónomos no tripulados (VANT). Con los ortomosaicos se verificó los cambios de cobertura de suelo existentes, para la actualización de las clases asignadas (Martínez Batlle, 2017) (**Figura 3**).

La clasificación sé realiza en *Google Earth Engine* mediante el paquete de clasificadores, al emplear *Random Forest* con la función ee.Classifier.smile*RandomForest*.

DESCRIPCIÓN SAMOF	Clases USV - INEGI	ESTRATO
Bosque Cultivado	BC y BI	C6
Selva Baja y Mediana Subperennifolia	SBQ, SBQP, SMQ, SG y VPN	C8
Manglar y Petén	VM y PT	C9
Selva Alta Subperennifolia	SAQ	C11
Selva Baja Caducifolia Subcaducifolia	SBC, SBK, SBS, MST y VPI	C12
Selva Mediana Caducifolia y Subcaducifolia	SMC y SMS	C13
Matorral Rosetófilo Costero	MRC	C20
Tular	VT	C23
Vegetación de Dunas Costeras	VU	C24
Pastizales Naturales	PN, PY, PH, VW, VS y VSI	C28
Pastizales Cultivado e Inducido	PC y PI	C280
Tierras Agrícolas	T, R y H Cultivos anuales	C29
Tierras Agrícolas	T, R y H Cultivos perennes	C290
Urbano y Construido	ZU y AH	C30
Suelo Desnudo	ADV y DV	C31
Agua	H2O, PRE y ACUI	C32

Tabla 1. Clases empleadas para la clasificación de imágenes de satélite

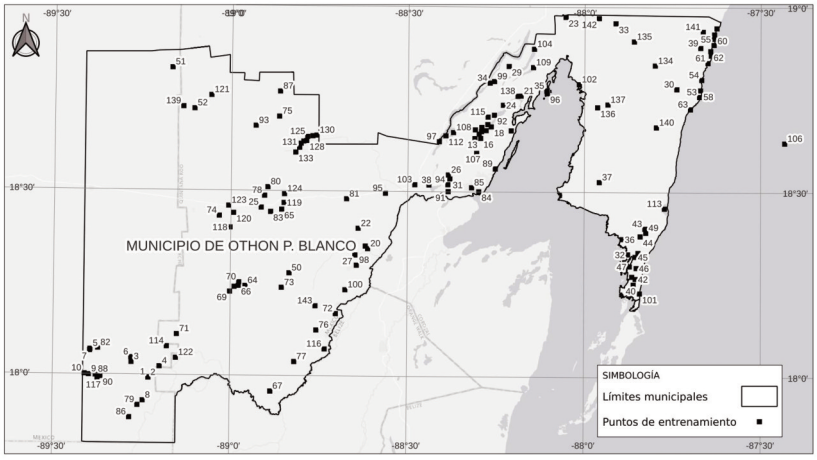

Figura 2. Distribución de los puntos de entrenamiento en el municipio de Othón P. Blanco para el 2021.

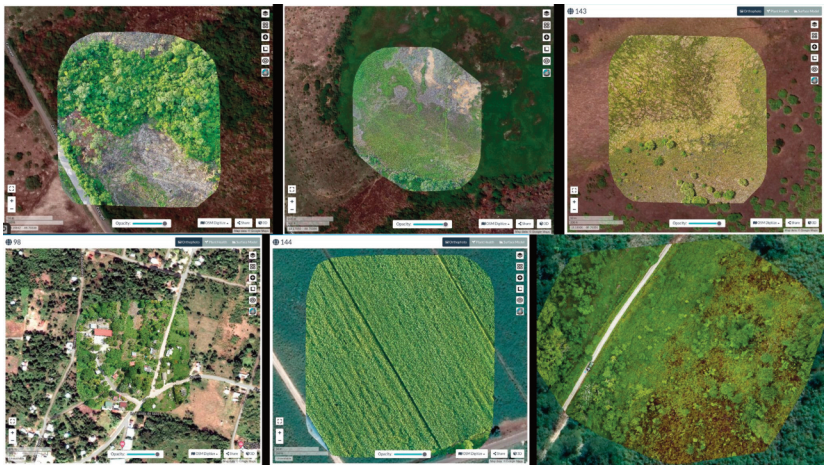

Figura 3. Ortomosaicos a partir de la malla de entrenamiento en el municipio de Othón P. Blanco para el 2021.

Se generó una imagen compuesta basado en la mediana de valores de píxel, del periodo de enero a diciembre del 2020, se empleó como proveedor de las imágenes a la Agencia Espacial Europea para utilizar las escenas de Sentinel-2 que abarcan los límites del municipio de Othón P. Blanco con una resolución espacial de 10 metros por pixel, se recortó el raster de la imagen compuesta mediante el uso de la capa de los límites municipales en *Google Earth Engine* con la función .clip(). Las nubes y sombras de nubes se enmascararon con una probabilidad de nubes inferior al 10 % y se filtraron las escenas mayores a un 20 % de cobertura de nubes (Gorelick *et al*., 2017) (**Figura 4**).

Sobre esta imagen sé aplica la malla de entrenamiento para el algoritmo ee.Classifier.smileRandomForest obteniéndose una capa vectoriales de polígonos de una hectárea para cada punto, validada de la supervisión en campo con los ortomosaicos a partir de VANT's. Sé generaron 93 puntos de cobertura terrestre, los puntos para validación de cuerpo de agua se realizaron con ayuda de los VANT's a la distancia sin generar los ortomosaicos, para reducir los tiempos de campo y de procesamiento (Martínez Batlle, 2017).

Validación del entrenamiento

Para obtener la validación cuantitativa de la precisión se empleó una matriz de confusión para determinar la

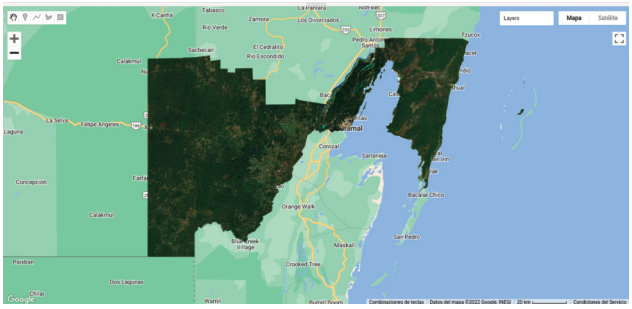

Figura 4. Imagen compuesta 2020 basada en mediana de valor de píxel.

precisión de la predicción de clases empleadas, al emplear la función .confusionMatrix() de Stehman,1997 que determina la precisión temática de un mapa de cobertura terrestre (**Tabla 2**).

RESULTADOS Y DISCUSIÓN

Se obtuvo un mapa de cobertura con las 16 clases (**Figura 5**). El resultado de la precisión general del entrenamiento: es de 0.99 lo cual señala una adecuada clasificación de las muestras que determina una suficiencia de las muestras de cada clase empleada para el entrenamiento del modelo clasificador (**Figura 6**).

Con respecto a la cuestión de la precisión de las clases, no es posible afirmar que la clasificación de la imagen compuesta sea suficientemente adecuada para representar con precisión las clases propuestas presentes en la región en cuestión. Para garantizar la precisión del período de clasificación de la imagen compuesta del 1 de enero al 31 de diciembre de 2020, es necesario realizar inspecciones visuales aleatorias del producto y verificar manualmente la coherencia de la asignación de clases entre la imagen y los datos de campo. La segunda fase del proyecto requiere una cantidad significativa de esfuerzo y recursos, al tiempo que enfatiza la importancia de obtener una comprensión integral del campo de estudio. Basándonos en nuestra observación inicial, parece que los resultados estimados para los bosques cultivados se han exagerado, ya que los bosques de tamaño mediano con cierto grado de degradación se han clasificado como bosques cultivados; por lo tanto, sugerimos realizar más inspecciones de campo similares a las utilizadas para generar la malla de entrenamiento para rectificar los valores del clasificador.

CONCLUSIÓN

La accesibilidad del aprendizaje automático a través de esta plataforma web es considerablemente de mayor capacidad en comparación con los flujos de trabajo

		REFERENCIAS															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLASES	0	620	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	950	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	1047	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	1050	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	1078	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	1052	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	1089	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	1060	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	320	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	974	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	0	1099	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	1072	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	1065	0	0
	14	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1062	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	829

Tabla 2. Matriz de error de reposición con representación de reposición de las clases mapeadas

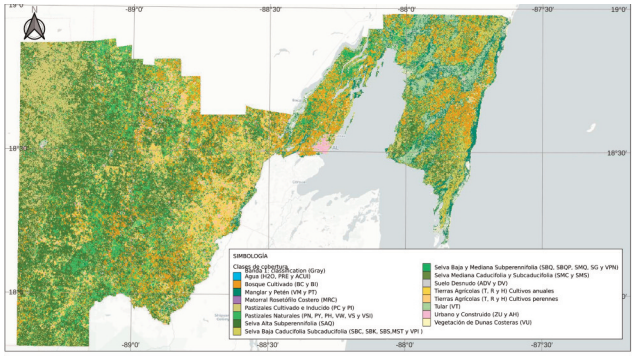

Figura 5. Mapa de cobertura generado del municipio de Othón P. Blanco para el año 2020.

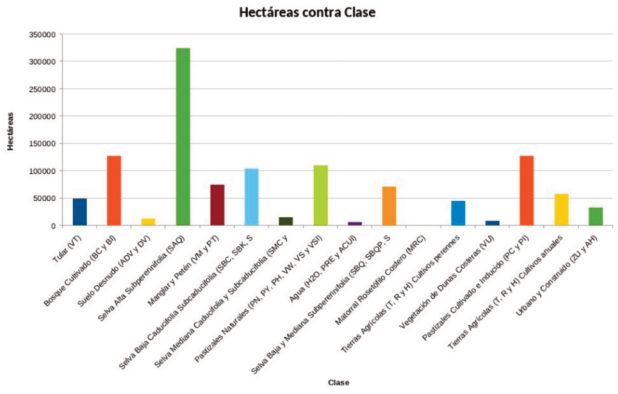

Figura 6. Valores de hectáreas por clase de cobertura de suelo.

convencionales que requieren el uso de equipos informáticos de alto rendimiento. Los drones sirven como un mecanismo de teledetección viable que nos permite obtener datos actuales y relevantes a un costo significativamente reducido, en comparación con la obtención de tiempo de satélite.

Pese a todas las opciones tecnológicas para reducir tiempos y costos, el trabajo para validar o calibrar en campo los datos obtenidos de sensores remotos es indispensable.

Repositorio del Código GEE

El proceso de clasificación se ejecuta mediante el uso de las librerías específicas del lenguaje de programación desarrollado para en *Google Earth Engine*, al que se accede como referencia utilizando la siguiente dirección web: https://code.earthengine.google.com/efa7a6b128383f63024d82ca38b7483f

REFERENCIAS BIBLIOGRÁFICAS

Breiman, L. 2001. Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

IPCC, 2014: Cambio climático 2014: Informe de síntesis. Contribución de los Grupos de trabajo I, II y III al Quinto Informe de Evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático [Equipo principal de redacción, R.K. Pachauri y L.A. Meyer (eds.)]. IPCC, Ginebra, Suiza, 157 págs.

Kumar, L., & Mutanga, O. 2018. Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. Remote Sensing, 10(10), 1509. https://doi.org/10.3390/rs10101509

Martínez Batlle, J. R. 2017. Introducción a la fotogrametría aplicada a imágenes tomadas desde sistemas de aeronaves pilotadas a distancia (RPAS) o drones. http://repository.humboldt.org.co/handle/20.500.11761/36058

Pozo, C., Armijo Canto, N. y Calmé, S. (editoras). 2011. Riqueza Biológica de Quintana Roo. Un análisis para su conservación, Tomo I. El Colegio de la Frontera Sur (Ecosur), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio), Gobierno del Estado de Quintana Roo y Programa de Pequeñas Donaciones (ppd). México, D. F.

Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment, 62(1), 77-89.

Fecha de recepción: 14 de abril de 2023
Fecha de arbitraje: 21 de junio de 2023
Fecha de aprobación: 25 de junio de 2023