

## Estimación de superficies agrícolas mediante procesamiento en la nube para el altiplano potosino

José de Jesús Hernández-Ramos<sup>1</sup>

José Pimentel-López<sup>1,§</sup>

Alejandro Amante-Orozco<sup>1</sup>

Esteban Salvador Osuna-Ceja<sup>2</sup>

1 Maestría en Ciencias en Innovación en Manejo de Recursos Naturales-Campus San Luis Potosí-Colegio de Postgraduados. Iturbide 73, Salinas de Hidalgo, San Luis Potosí, México. CP. 78600.

2 Campo Experimental Pabellón-INIFAP. Carretera Aguascalientes-Zacatecas km 32.5, Pabellón de Arteaga, Aguascalientes, México. CP. 20660.

Autor para correspondencia: josep@colpos.mx

### Resumen

Las imágenes satelitales ópticas son bancos de información poderosos para la estimación de superficies agrícolas. El objetivo de este estudio fue estimar las superficies agrícolas en los municipios de Salinas, Santo Domingo y Villa de Ramos, mediante el procesamiento en la nube de imágenes satelitales y su comparación con la tecnología de INEGI-tradicional. El trabajo se realizó limitado a la zona agrícola, la cual suma una superficie de 190 871 ha, de esta, el 86% son de temporal. El período de estudio fue de octubre 2020 a octubre 2021. Se aplicaron seis algoritmos de clasificación, tres para INEGI-tradicional: mínima distancia, máxima verosimilitud y spectral angle mapper en QGIS 3.18 y tres para el procesamiento en la nube: classification and regression trees, random forest y support vector machine con Google Earth Engine. Se estimaron las superficies de los principales cultivos (maíz, frijol, avena, alfalfa y chile) para la zona de estudio, a partir de 294 muestras de campo. Para el procesamiento de imágenes Sentinel-2, se utilizó una geomédiana libre de nubes. Los resultados de las matrices de confusión indicaron qué clasificaciones fueron más precisas, los valores fueron de: 89% para classification and regression trees y random forest, 59% para support vector machine, 48% para mínima distancia, 43% para máxima verosimilitud y 46% para spectral angle mapper. Los algoritmos classification and regression trees y random forest superaron en precisión a los demás clasificadores evaluados, estos estimaron las superficies agrícolas de maíz y frijol más cercanas entre sí (80 131 y 98 138 ha en maíz y 60 174 y 60 358 ha en frijol) en comparación a los clasificadores restantes.

### Palabras clave:

cultivos, google earth engine, QGIS.



## Introducción

El SIAP (2020) reporta que los municipios de Salinas, Villa de Ramos y Santo Domingo, cubren una superficie agrícola de 190 871 ha de las cuales el 86% son de temporal. Bajo esta condición, los principales cultivos que se siembran son maíz y frijol. Dada su importancia en la región, es necesario contar con información agrícola continua y espacialmente distribuida durante el ciclo de producción de dichos cultivos. Por ello, es conveniente utilizar, técnicas de percepción remota que han sido más aceptadas para el procesamiento de datos de teledetección, debido a los avances que ha tenido esta disciplina (Romero, 2016; Gallardo-Cruz *et al.*, 2019).

La combinación de la teledetección con la agronomía origina un tema de investigación específico, donde los especialistas tienen un ámbito de análisis, para un campo de aplicación muy variado, sobre todo, en los últimos años con la aplicación de infraestructuras, para el procesamiento en la nube de grandes volúmenes de datos (Aguilar, 2016). En el Instituto Nacional de Estadística y Geografía (INEGI) se aplica una metodología para la estimación de superficies agrícolas, que requiere de considerables recursos humanos, económicos e informáticos.

Sin embargo, el procesamiento de cómputo en la nube, combinado con las técnicas de percepción remota, puede ofrecer una mejor optimización del recurso humano e informático. Se ha reportado la combinación de disciplinas de procesamiento en la nube con otras, donde participan especialistas en uso de suelo y vegetación (Killough, 2018; Vega *et al.*, 2019; Venkatappa *et al.*, 2019; Aghababaei *et al.*, 2021).

Se conoció que hay pocos estudios registrados en la literatura especializada, referentes al procesamiento en la nube y zonas agrícolas, como los de German *et al.* (2019); Mananze *et al.* (2020); Amani *et al.* (2020). German *et al.* (2019), aplicaron algoritmos de support vector machine y random forest para estimar las superficies de hortalizas en la periferia de La Plata, Argentina, y obtuvieron una exactitud entre el 96% y 98%.

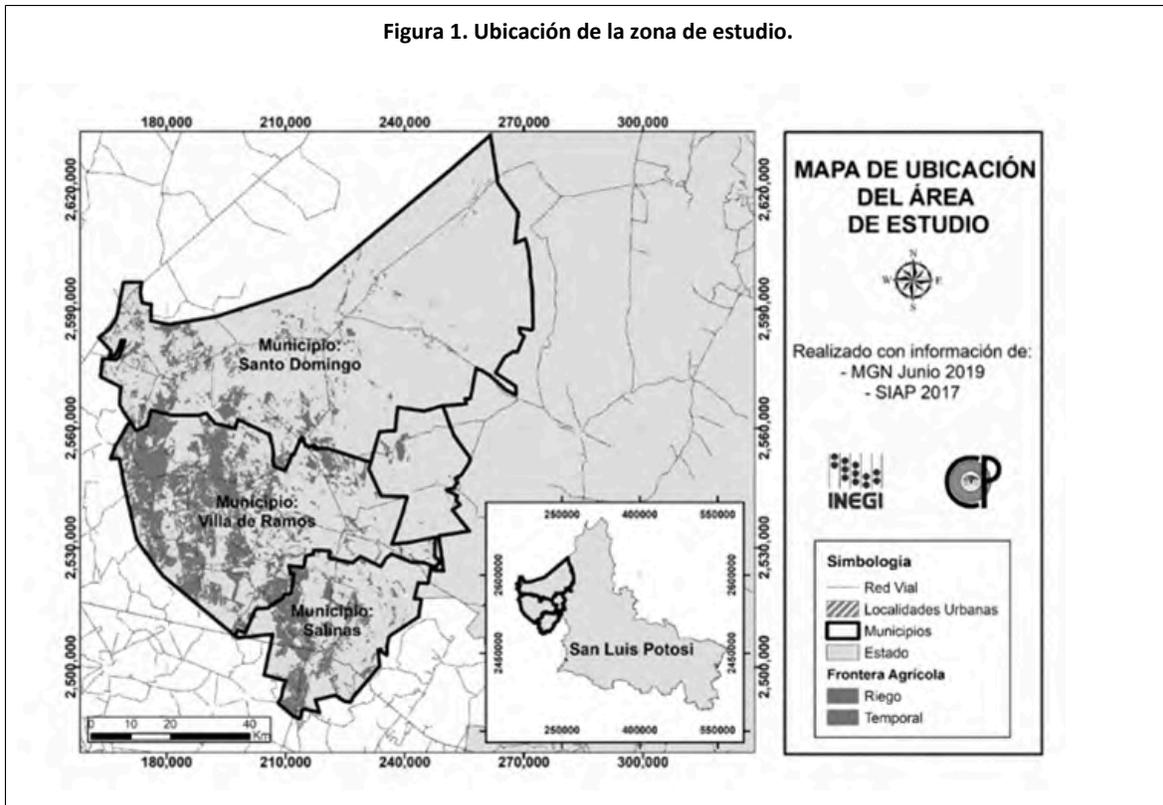
Los algoritmos aquí planteados pueden arrojar buenos resultados en cómputo en la nube, a pesar de que, en este caso, la estructura de las plantaciones de los cultivos de interés, son totalmente diferentes, ya que en su mayoría se concentran en zonas de temporal y por esta razón, cultivos como el frijol tienen una cobertura vegetal baja, así como los cultivos de avena y el maíz en etapas tempranas, por lo que predomina la exposición del suelo desnudo y es lo que finalmente capta la imagen satelital.

En cuanto al procesamiento en equipos de cómputo, existen estudios como el de TA *et al.* (2022), que trabajaron con ambientes ecológicos en Sinaloa, México, donde aplicaron el algoritmo de máxima verosimilitud y lograron obtener hasta un 87% de precisión. Con base en estos antecedentes, se realizó el presente trabajo con el objetivo de estimar las superficies agrícolas de tres municipios del Altiplano Noroeste Potosino, mediante el procesamiento en la nube de imágenes satelitales y su comparación con la tecnología INEGI-tradicional; así como evaluar los resultados de los seis algoritmos utilizados.

## Materiales y métodos

Los municipios estudiados se ubican en el Altiplano Noroeste Potosino, en los límites con el estado de Zacatecas (Figura 1). El clima predominante en la región es seco templado y semiseco templado de acuerdo con el INEGI (2008). La precipitación pluvial anual oscila entre los 300 y 400 mm y se tiene un régimen de lluvias en verano y precipitación invernal menor a 5% del registro anual (CONAGUA, 2020). La temperatura media anual varía entre 12 y 18 °C. La superficie municipal, que comprende las áreas de Salinas, Villa de Ramos y Santo Domingo alcanza las 845 250 ha, de la cual 190 871 ha son agrícola (SIAP, 2020), la mayor parte de esta superficie se encuentra en el municipio de Villa de Ramos con 104 760 ha.

Figura 1. Ubicación de la zona de estudio.

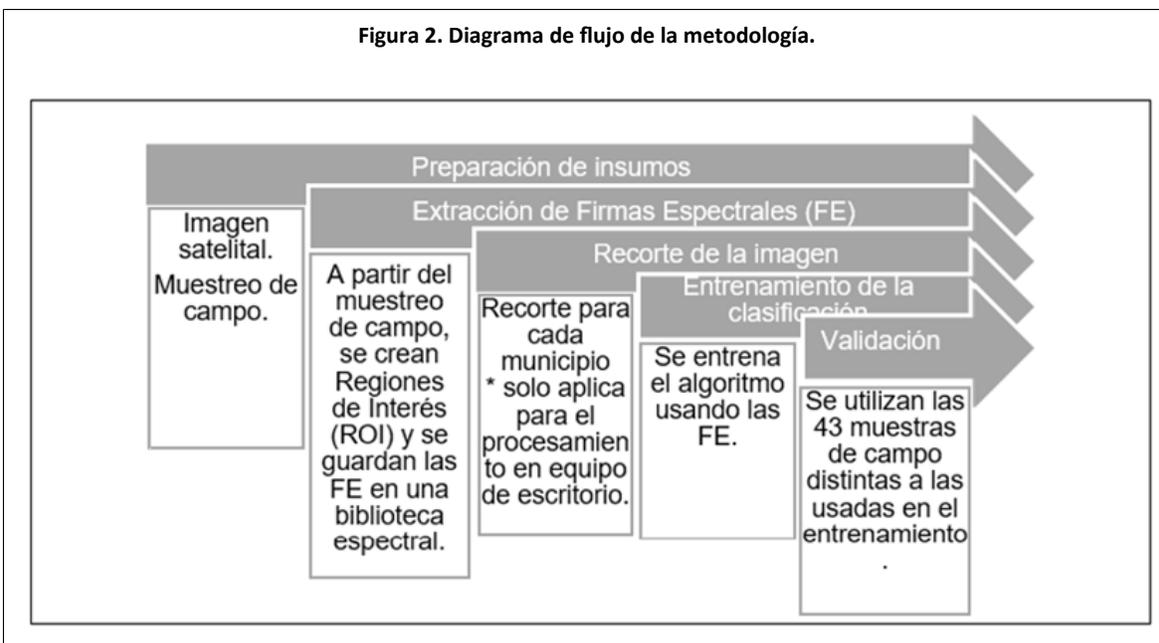


La información se obtuvo en el período comprendido del mes de octubre 2020 a octubre 2021, la cual fue utilizada para estimar la superficie agrícola en el ciclo primavera-verano (PV) 2020, en los tres municipios estudiados. En esta investigación se evaluaron los diferentes algoritmos que existen en la literatura para efectuar extracción de rasgos sobre regiones de interés en una imagen satelital y distintos algoritmos inteligentes para llevar a cabo la clasificación a partir de patrones, esto con el fin de proponer una mejora en la metodología que permita clasificar cultivos eficientemente.

Este trabajo se dividió en cinco etapas: la primera estuvo dedicada a la adquisición de las imágenes satelitales del sensor Sentinel-2 MSI multi spectral instrument, level-2A, de la plataforma de Copernicus. La resolución espacial utilizada, fue de 10 (B2, B3, B4 y B8) y 20 (B5, B6, B7, B8A, B11 y B12) metros a la par. El muestreo de campo se llevó a cabo del 19 al 23 de octubre de 2020, la segunda se enfocó en aplicar los algoritmos de extracción de las firmas espectrales, la tercera se dedicó al recorte de la imagen de cada municipio, la cuarta se orientó orientada al entrenamiento de los diferentes algoritmos inteligentes que permitirán clasificar los rasgos descriptivos obtenidos en la primera etapa y la última consistió en la validación del desempeño del clasificador mediante el uso de 43 muestras distintas a las utilizadas en el entrenamiento (Figura 2).



Figura 2. Diagrama de flujo de la metodología.



La información de campo que se levantó en el territorio estudiado tuvo como referencia el universo de 34 132 terrenos que incluyen propiedad ejidal y pequeña propiedad, para los tres municipios (INEGI, 2020a). Para el muestreo de campo se planeó caracterizar 380 parcelas de estos sitios, las cuales fueron seleccionadas al azar, mediante la herramienta random selection de QGIS v.2.18.

Por otra parte, en un proyecto de QField v.3.3.0 se levantaron los datos de campo, este fue dividido en cuatro apartados: ubicación de la muestra, detalle del cultivo, fotos y descripción general. En la ubicación, se incluyó el marco geoestadístico nacional (MGN), los nombres y claves geoestadísticas, dónde estuvo situado cada terreno, los cuales son datos no editables en el dispositivo móvil.

En el detalle del cultivo, solo se definió el tipo, su etapa fenológica y algún daño si existiera. En el apartado de fotos, se capturó un par de imágenes con la cámara del dispositivo móvil, estas estuvieron ligadas a cada terreno, una foto panorámica y otra de las plantas. Las muestras de campo sirvieron para definir las regions of interest (ROI) y entrenar las clasificaciones supervisadas. Primeramente, se trabajó en la plataforma de Google Earth Engine (GEE), con apoyo de la ingesta de archivos vectoriales para la digitalización de los ROI's, en este se usaron los algoritmos de random forest (RF) que funciona con árboles de decisión individuales, lo que crea para cada árbol un dato ligeramente distinto (Panagiotakis *et al.*, 2021).

El segundo algoritmo fue classification and regression trees (CART) que funciona a través de árboles de decisión para la regresión y clasificación de los datos (Strzelecka y Zawadzka, 2021) y finalmente el tercer algoritmo usado en GEE fue support vector machine (SVM), que funciona mediante la correlación de los datos en un espacio de grandes dimensiones, de tal forma que las regiones de entrenamiento se puedan categorizar (Sánchez-Pozo *et al.*, 2021; Kok *et al.*, 2021).

La misma imagen se clasificó en un equipo de cómputo con las siguientes características: procesador Intel® Core™ i7, RAM de 32 GB, sistema operativo Windows de 64 bits, donde se utilizaron los algoritmos: mínima distancia (MD), máxima verosimilitud (MV) y spectral angle mapper (SAM) en el software QGIS 3.18 y el complemento semi-automatic classification plugin (SCP). Para ejecutar el proceso se extrajeron las firmas espectrales (FE) de cada ROI, estas se almacenaron en una biblioteca espectral, para posteriormente aplicar las FE a la imagen satelital, para cada algoritmo.

Para la validación de los resultados se utilizaron 43 muestras de campo que se levantaron en la misma fecha y zona de estudio. Con estas muestras se crearon 43 ROIS, los cuales se ingresaron

al proyecto y se corrió el algoritmo de matriz de confusión de cada imagen clasificada *versus* los 43 ROIS. Posteriormente en el seguimiento de las precipitaciones, se usaron los productos climáticos mensuales que ofrece TerraClimate de la superficie terrestre. TerraClimate utiliza la interpolación asistida por el clima, que combina los datos climatológicos normalizados del conjunto de datos WorldClim (Abatzoglou *et al.*, 2018).

Para estimar la geomédiana, se utilizaron una serie de imágenes obtenidas principalmente de la malla 14QKL, ya que esta cubre la mayor parte de la zona de interés; además, de las imágenes: 14QKM, 13QHF, 13QHG 13QGF y 13QGG. La geomédiana se define como una estadística de alta dimensión que intercambia una serie temporal de imágenes de observación terrestre por un solo píxel compuesto de alta calidad con ruido espacial reducido que mantiene su consistencia espacial (Roberts *et al.*, 2017).

La geomédiana es particularmente útil cuando la distribución de probabilidad de los datos no es necesariamente normal multivariante o si hay valores atípicos en los datos (Roberts *et al.*, 2017; INEGI, 2020b). Finalmente se calculó el Índice de vegetación de diferencia normalizado (NDVI), por sus siglas en inglés en intervalos de 5 días durante el ciclo PV para los ROIS muestreados. Con la información obtenida, se analizó el comportamiento promedio y tendencia sostenida del NDVI en las áreas agrícolas de temporal y riego de los tres municipios bajo estudio, a partir del comportamiento del índice se identificó la fecha de mayor vigor de los cultivos.

## Resultados y discusión

El comportamiento de NDVI identificó el pico más alto en el desarrollo de la vegetación, en septiembre de 2020, se tomó este mes como la fecha óptima para detectar todos los cultivos en su mayor vigor. El análisis de resultados arrojó que, en la zona de estudio, el patrón de precipitación se caracteriza por una fuerte variabilidad interestacional e interanual, que a su vez rige el inicio de siembras de cultivos de temporal para el ciclo primavera-verano.

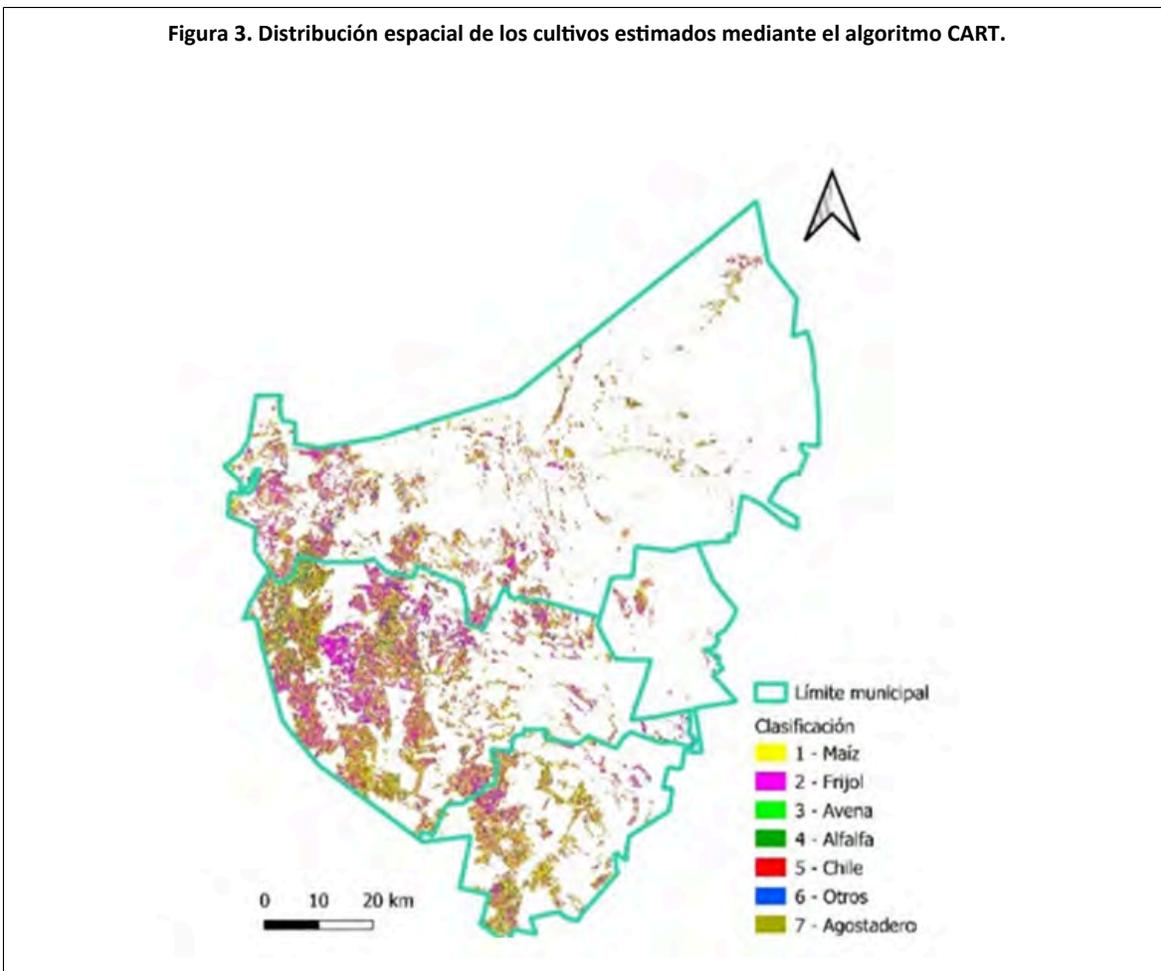
La magnitud estacional de la precipitación en el Altiplano Noroeste Potosino se registra, en promedio, como sigue: verano, otoño, invierno y primavera con 53.2%, 20.4%, 14.3 y 12.1% de la precipitación total anual, respectivamente. Se consideró que la gran variación existente en la precipitación total anual se origina en la fluctuación del número de eventos de precipitación más que en la magnitud de los mismos y es responsable de la ocurrencia y duración de la sequía intraestival o canícula, que se sabe ocurre en estas zonas semiáridas (Bravo *et al.*, 2006; Núñez-López *et al.*, 2007).

Por otra parte, los resultados del muestreo no fueron los esperados de acuerdo con lo planeado, se lograron obtener 294 muestras de las 380 indicadas en el tamaño de muestra, la razón de no cumplir el objetivo del muestreo fue el tiempo disponible para el muestreo, debido a las cosechas tempranas en la zona. De las muestras colectadas: 102 fueron de frijol, 101 de maíz, 7 de alfalfa, 5 de avena, 6 de chile y 35 de otros cultivos (entre los que se incluyeron principalmente la combinación de dos o más cultivos en la misma parcela) y finalmente 38 muestras de sin cultivo.

La distribución estimada de los diferentes cultivos mediante el algoritmo CART, procesados en GEE se muestra en la Figura 3. Puede notarse que los cultivos dominantes en la superficie sembrada son maíz y frijol, en estos casos se agruparon los cultivos del mismo tipo, tanto de riego como de temporal. No obstante, de acuerdo con lo observado durante el muestreo, los cultivos de avena, alfalfa y chile se distribuyen espacialmente en las zonas de riego.



Figura 3. Distribución espacial de los cultivos estimados mediante el algoritmo CART.



Se digitalizaron 294 ROIS que sirvieron para entrenar las clasificaciones en cada algoritmo. Estos fueron polígonos homogéneos de un mismo rango espectral, para disminuir la confusión que se pudiera dar entre las clases. En el Cuadro 1, se presentan los resultados de la superficie estimada para cada uno de los cultivos evaluados con los seis algoritmos trabajados. En general, se observó que las superficies estimadas con los tres algoritmos (CART, RF y SVM) aplicados en GEE fueron superiores que las estimadas con los tres algoritmos (MV, MD y SAM) en QGIS para todos los casos.

Cuadro 1. Superficies estimadas en hectáreas para los diferentes algoritmos por cultivo.

Cultivo	SVM	CART	RF	MV	SAM	MD	SIAP
Maíz	110 706	80 131	98 138	55 472	75 621	78 265	41 435
Frijol	52 565	60 174	60 358	77 794	46 686	48 553	89 605
Avena	2 988	1 921	1 538	1 914	1 724	3 341	6 850
Alfalfa	5 402	5 019	3 503	1 514	4 434	2 625	2 780
Chile	2 741	5 642	3 979	735	2 272	3 111	16 552
Otros	1 914	23 442	14 211	12 149	15 154	12 968	1 029
Sin cultivo	25 867	25 853	20 454	16 436	17 469	15 420	-
Total	202 182	202 182	202 182	166 015	163 359	164 282	158 251

MD= mínima distancia; MV= máxima verosimilitud; SAM= spectral angle mapper; CART= classification and regression trees; RF= random forest; SVM= support vector machine.

De acuerdo con estos resultados la superficie más alta estimada con todos los clasificadores fue la de maíz, luego con frijol y por último los cultivos de riego (alfalfa, chile y avena) además se incluyeron otros y sin cultivo. De manera que, los cultivos de maíz y frijol son los de mayor presencia y los más importantes en la zona de estudio, ya que suman una superficie promedio estimada de 140 744 ha para el ciclo estudiado. Respecto a la superficie reportada por el SIAP (2021) para el ciclo estudiado difiere mucho la superficie estimada con la mayoría de los algoritmos aplicados para todos los cultivos evaluados.

Probablemente las discrepancias observadas sobre todo en los cultivos de riego, se deba a la distribución espacial de las muestras que se tuvieron para entrenar los algoritmos. Los tiempos de procesamiento estimados para cada algoritmo se muestran en el Cuadro 2. En el caso del entrenamiento, abarcó la creación, modificación y depuración de los ROI's, mismos que sirvieron para entrenar los demás algoritmos dentro de la misma plataforma.

**Cuadro 2. Tiempos promedios (en horas) de procesamiento.**

	Entrenamiento	Ejecución de la clasificación	Descarga de la imagen	Cálculo de matriz de confusión	Total
GEE	5.25	0.25	0.25	0.25	6
QGIS	6	15	0	5	26

\* = comprende la creación de los ROI's, por única ocasión, ya que serán los mismos polígonos de entrenamiento para todos los algoritmos.

El tiempo de procesamiento en la nube, utilizado en los 294 ROI's, fue de 6 h por algoritmo, mientras que, el tiempo requerido para la ejecución de cada algoritmo, con los mismos ROI's, aplicado en software para equipo de escritorio (QGIS 3.18), fue de hasta 26 horas. Esta diferencia de 20 h es debido a los motores de GEE que funcionan como servidores para procesar. Este ahorro de tiempo coincide con lo reportado por Perilla y Mas (2020) en el sentido que GEE, es una herramienta poderosa, que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube.

Por lo cual, los tiempos de procesamiento fueron significativamente más rápidos en la nube que en el equipo de escritorio, lo que asegura el ahorro de recursos económicos, informáticos y humanos. La validación de los resultados estuvo determinada por la matriz de confusión de cada algoritmo, esta matriz se obtuvo al comparar los datos predichos *versus* datos observados que correspondieron a las 43 muestras que no se utilizaron en el entrenamiento. Esto identifica cuántos píxeles fueron clasificados de manera correcta y cuantos fueron confundidos con otro cultivo, eso se realizó para cada una de las 7 clases.

Las clases obtenidas se definieron de la siguiente manera: 1 para maíz, 2 para frijol, 3 para avena, 4 para alfalfa, 5 para chile, 6 para otros cultivos que están combinados y finalmente la clase 7 para parcelas sin cultivo o en descanso. Para el cálculo de las matrices de error que representan la probabilidad de qué áreas muestreadas en la imagen son clasificadas correctamente con la suma de aquellos píxeles que fueron bien clasificados. Esto significó, que el valor predicho y el valor observado corresponden a la misma clase, por lo que se tiene un número de píxeles clasificados correctamente.

A este valor se le calculó su cociente, se dividió entre el número total de píxeles y se obtuvo un porcentaje de exactitud para cada clasificación. Para el caso del algoritmo SVM, hubo una ligera confusión entre la clase 2 de los predichos y la 1 de los observados y entre las clases 6 y 1 (28 píxeles), en este caso se obtuvo una exactitud del 59% (Cuadro 3).



**Cuadro 3. Matriz de confusión para el algoritmo SVM.**

SVM	Valores predichos							Suma		
	1	2	3	4	5	6	7			
Valores observados	1	95	68	0	0	0	28	11	202	
	2	15	89	0	0	0	0	3	107	
	3	0	0	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	
	5	0	0	0	0	0	0	0	0	
	6	0	0	0	0	0	0	0	0	
	7	5	0	0	0	0	1	1	7	
Suma		115	157	0	0	0	29	15	185	
									VP+VN	185
									VP+FP+FN+VN	316
									Exactitud	59%

VP= verdadero positivos; VN= verdaderos negativos; FP= falso positivo; FN= falso negativo.

Para el caso del algoritmo CART, se detectó una confusión baja y fue entre las clases 1 y 2 principalmente. Este algoritmo arrojó una exactitud del 89% en 281 pixeles de los 316 (Cuadro 4).

**Cuadro 4. Matriz de confusión para el algoritmo CART.**

CART	Valores predichos							Suma		
	1	2	3	4	5	6	7			
Valores observados	1	102	7	0	0	0	4	2	115	
	2	8	148	0	0	0	4	3	163	
	3	0	0	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	
	5	1	0	0	0	0	0	0	1	
	6	3	2	0	0	0	21	0	26	
	7	1	0	0	0	0	0	10	11	
Suma		115	157	0	0	0	29	15	281	
									VP+VN	281
									VP+FP+FN+VN	316
									Exactitud	89%

VP= verdadero positivos; VN= verdaderos negativos; FP= falso positivo; FN= falso negativo.

Para el algoritmo RF, la confusión de la clase 1 fue mínima con la 2, 6 y 7, con este algoritmo se obtuvo una exactitud del 89% (Cuadro 5).

**Cuadro 5. Matriz de confusión para el algoritmo RF.**

RF	Valores predichos							Suma	
	1	2	3	4	5	6	7		
Valores observados	1	104	10	0	0	0	6	3	123
	2	7	147	0	0	0	2	0	156
	3	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0

RF	Valores predichos							Suma	
	1	2	3	4	5	6	7		
5	1	0	0	0	0	0	0	1	
6	2	0	0	0	0	21	2	25	
7	1	0	0	0	0	0	10	11	
Suma	115	157	0	0	0	29	15	282	
								VP+VN	282
								VP+FP+FN+VN	316
								Exactitud	89%

VP= verdadero positivos; VN= verdaderos negativos; FP= falso positivo; FN= falso negativo.

El algoritmo mínima distancia arrojó una exactitud del 48%, mientras que para el algoritmo MV se obtuvo una exactitud del 43% y para el algoritmo SAM una exactitud del 46%. Estos tres algoritmos presentaron los valores más bajos de exactitud.

De acuerdo con las matrices y los valores de exactitud al comparar los algoritmos ejecutados en GEE *versus* en QGIS, se encontró que los porcentajes más altos se obtuvieron con los algoritmos ejecutados en GEE, los cuales alcanzaron el 89% de precisión con excepción de SVM que tuvo un 59% mientras que, los estimados con los algoritmos de QGIS fueron más bajos, estos oscilaron entre el 40 y 50%. Con esto se corrobora que los mejores resultados estimados fueron a partir de la plataforma GEE, al menos eso se pudo calcular con las 43 muestras que se usaron para validar las clasificaciones.

No hay un trabajo previo de la zona que estime superficies agrícolas, sin embargo, la fuente de información (SIAP, 2021) oficial arroja un dato mensual del avance de siembras por cultivo, el cual es comparable con los resultados que se obtuvieron en este trabajo (Cuadro 1). Al cotejar la información se observó que se disparan las superficies, por ejemplo, el Chile, donde SIAP (2021) reportó hasta cuatro veces más de lo estimado en este trabajo. Sabiendo que esta metodología está basada en un modelo matemático. Es conocido que el resultado final estimado poseerá un porcentaje de error en comparación al dato real.

Por otro lado, durante el proceso metodológico, se detectó un error en la ejecución del muestreo de campo, que no fue cubierto en su totalidad de acuerdo a la planeación de este, la razón trascendental fue que el muestreo fue llevado a cabo en un periodo corto de tiempo y no es que no se pudiera prolongar por más días el muestreo, sino que cuando se visitaban los terrenos, algunos de estos ya se encontraban cosechados con restos del cultivo o los productores andaban recolectando sus cultivos y en el mejor de los casos aún estaba el cultivo de pie.

Aunado a esto, la distribución de las muestras estuvo acomodada de manera que, para llegar a la zona de muestreo, se consideraban hasta 40 km de distancia debido a que el personal de campo pernotó en la localidad de Salinas de Hidalgo. Por ejemplo, aquellos terrenos muestreados en la localidad de La Herradura, Santo Domingo, a una distancia de 45 km de Salinas. Un aspecto importante para la seguridad del personal es garantizar la integridad del personal en la zona y esto perturbo en los horarios de trabajo, razón por la cual el personal regresaba de las parcelas con tiempo calculando no llegar de noche al lugar de pernocta.

## Conclusiones

Los algoritmos que arrojaron la precisión más alta fueron CART y RF que están basados en el procesamiento en la plataforma de GEE, mismos que permiten, el ahorro de hardware y software, así como de tiempo. Esto se ve reflejado en las etapas de preprocesamiento, entrenamiento y ejecución de la clasificación supervisada.

Sin embargo, la etapa de muestreo de campo y todo lo que implica su desarrollo, no hay forma de suplirlo, ya que es requisito indispensable como insumos para la clasificación. La plataforma de GEE ha venido a revolucionar la forma de trabajar con datos de percepción remota, ya que fortalece la ayuda de los usuarios en el ahorro de recursos humanos, informáticos y económicos al ser una plataforma de libre acceso.

Como mejoras a la metodología usada en este trabajo se identifican dos que están relacionadas por la variable tiempo; la primera mejora podría ser, planear con más antelación el muestreo de campo para cumplir el 100% en forma y tiempo. La otra mejora es el seguimiento de las condiciones climáticas del lugar en el ciclo agrícola de estudio, para identificar plenamente el inicio de las siembras y la posterior colusión del ciclo, esto coadyuvará en identificar el mejor momento para llevar a cabo el muestreo de campo.

## Agradecimientos

Al Instituto Nacional de Estadística y Geografía por el tiempo y el apoyo prestado para llevar a cabo el muestreo de campo, así como la infraestructura informática para la corrida de las clasificaciones. Al Colegio de Postgraduados Campus San Luis Potosí, por el apoyo y asesoría durante el desarrollo de este trabajo.

## Bibliografía

- 1 Abatzoglou, J. T.; Dobrowski, S. Z.; Parks, S. A. and Hegewisch, K. C. 2018. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data. 5(1):1-12. Doi:10.1038/sdata.2017.191.
- 2 Aghababaei, M.; Ebrahimi, A.; Naghipour, A.; Asadi, E.; and Verrelst, J. 2021. Vegetation types mapping using multi-temporal Landsat images in the google earth engine platform. Remote Sensing. 13(22):4683-15. <https://doi.org/10.3390/rs13224683>.
- 3 Aguilar, L. J. 2016. Big data, análisis de grandes volúmenes de datos en organizaciones. Primera edición. Alfaomega Grupo Editor. México, DF.
- 4 Amani, M.; Kakooei, M.; Moghimi, A.; Ghorbanian, A.; Ranjgar, B.; Mahdavi, S.; Davidson, A.; Fiset, T.; Rollin, P.; Brisco, B. and Mohammadzadeh, A. 2020. Application of google earth engine cloud computing platform, sentinel imagery and neural networks for crop mapping in Canada. Remote Sens. 12(21):3561-18. <https://doi.org/10.3390/rs12213561>.
- 5 Bravo, L. A. G.; González, G. H. y Rumayor, R. A. 2006. Sequía: vulnerabilidad, impacto y tecnología para afrontarla en el Norte-Centro de México. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP)-CIRNOC-Campo Experimental Zacatecas. 2ª Edición. Libro técnico núm. 4. 297 p.
- 6 CONAGUA. 2020. Comisión Nacional del Agua. Estaciones meteorológicas automáticas (EMAS). <https://smn.conagua.gob.mx/es/observando-el-tiempo/estaciones-meteorologicas-automaticas-emas>.
- 7 Gallardo-Cruz, A; Fernández-Montes, O. A. y Rives, C. 2019. Detección de amenazas y oportunidades para la conservación en la cuenca baja del Usumacinta a partir de técnicas de percepción remota. Ecosistemas. 28(2):82-99. <https://doi.org/10.7818/ECOS.1611>.
- 8 German, L. A.; Vitale, J. P.; Waldman, C. P. y Castañeda N. 2019. Estimación de superficie de invernáculos en el partido de la plata, mediante dos algoritmos de inteligencia artificial en la plataforma Google Earth Engine. In: XI Congreso de AgrolInformática (CAI) JAIIO. 48:1-11 <http://sedici.unlp.edu.ar/handle/10915/88069>.
- 9 INEGI. 2008. Instituto Nacional de Estadística y Geografía. Unidades climáticas. <https://www.inegi.org.mx/temas/climatologia/>.
- 10 INEGI. 2020a. Instituto Nacional de Estadística y Geografía. Marco geoestadístico nacional (MGN) <https://www.inegi.org.mx/temas/mg/#Descargas>.
- 11 INEGI. 2020b. Instituto Nacional de Estadística y Geografía. Geomediana Landsat <https://www.inegi.org.mx/investigacion/geomediana/>.

- 12 Killough, B. 2018. Overview of the open data cube initiative. *In*: IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium. 8629-8632 pp.
- 13 Kok, Z. H.; Shariff, A. R. M.; Alfatni, M. S. M. and Kairunniza-Bejo, S. 2021. Support vector machine in precision agriculture: a review. *Comput. Electron. Agric.* 191:106546-12. <https://doi.org/10.1016/j.compag.2021.106546>.
- 14 Mananze, S.; Pôças, I. and Cunha, M. 2020. Mapping and assessing the dynamics of shifting agricultural landscapes using google earth engine cloud computing, a case study in Mozambique. *Remote Sensing.* 12(8):1279-23. <https://doi.org/10.3390/rs12081279>.
- 15 Núñez-López, D.; Muñoz-Robles, V. M.; Reyes-Gómez, I.; Velasco-Velasco, I. y Gadsden-Esperza, H. 2007. Caracterización de la sequía a diversas escalas de tiempo en México. *Agrociencia.* 41(3):253-262.
- 16 Panagiotakis, C.; Papadakis, H. and Fragopoulou, P. 2021. A dual hybrid recommender system based on SCoR and the random forest. *Computer Science and Information Systems.* 18(1):115-128. <https://doi.org/10.2298/CSIS200515046P>.
- 17 Perilla, G. A. y Mas, J. F. 2020. Google Earth Engine (GEE): una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube. *Investigaciones Geográficas.* 101:1-6. <https://doi.org/10.14350/rig.59929>.
- 18 Roberts, D.; Mueller, N. and McIntyre, A. 2017. High-dimensional pixel composites from earth observation time series. *IEEE Transactions on Geoscience and Remote Sensing.* 55(11):6254-6264. Doi: 10.1109/TGRS.2017.2723896.
- 19 Romero, F. S. 2006. La teledetección satelital y los sistemas de protección ambiental. *Quivera Revista de Estudios Territoriales.* 8(1):315-356.
- 20 Sánchez-Pozo, N.; Trilles-Oliver, S.; Solé-Ribalta, A.; Lorente-Leyva, L.; Mayorca-Torres, D. and Peluffo-Ordóñez D. 2021. Algorithms air quality estimation: a comparative study of stochastic and heuristic predictive models, lecture notes in computer science. 12886:293-304. <https://doi.org/10.1007/978-3-030-86271-8-25>.
- 21 SIAP. 2020. Servicio de Información Agroalimentaria y Pesquera. Frontera agrícola serie II. <http://infosiap.siap.gob.mx/gobmx/datosAbiertos.php>.
- 22 SIAP. 2021. Servicio de Información Agroalimentaria y Pesquera. Avance de siembras y cosechas. <https://nube.siap.gob.mx/cierreagricola/>.
- 23 Strzelecka, A. and Zawadzka, D. 2021. Application of classification and regression tree (CRT) analysis to identify the agricultural households at risk of financial exclusion. *Procedia Computer Science.* 192:4532-4541. <https://doi.org/10.1016/j.procs.2021.09.231>.
- 24 TA, V. G.; Douriet-Angulo, A.; Tirado-Ramírez, M. A.; López-Urquidez, G. A. and López-Orona, C. A. 2022. Root rot and wilt caused by *Fusarium Nygamai* of bean (*Phaseolus vulgaris*) in Sinaloa, México. *Plant Disease.* 106(10):2748-2. <https://doi.org/10.1094/pdis-01-22-0123-pdn>.
- 25 Vega, P. J. J.; Zárate-Gómez, R.; Minaya, V. R. J.; Brañas, M. M. y Benavides, R. J. E. 2019. Predicción de la pérdida de la cobertura vegetal por aumento de áreas urbanas en Iquitos, Perú. *Ciencia Amazónica.* 7(1):37-50. <https://doi.org/10.22386/ca.v7i1.263>.
- 26 Venkatappa, M.; Sasaki, N.; Shrestha, R. P.; Tripathi, N. K. and Ma, H. O. 2019. Determination of vegetation thresholds for assessing land use and land use changes in Cambodia using the google earth engine cloud-computing platform. *Remote Sensing.* 11(13):1514-30. <https://doi.org/10.3390/rs11131514>.

## Estimación de superficies agrícolas mediante procesamiento en la nube para el altiplano potosino

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 October 2024
Date accepted: 01 February 2025
Publication date: 12 March 2025
Publication date: Jan-Feb 2025
Volume: 16
Issue: 1
Electronic Location Identifier: e3369
DOI: 10.29312/remexca.v16i1.3369

### Categories

Subject: Artículo

### Palabras clave:

**Palabras clave:**

cultivos  
google earth engine  
QGIS.

### Counts

Figures: 3  
Tables: 5  
Equations: 0  
References: 26  
Pages: 0