

Estimación de superficie de invernáculos en el Partido de La Plata, mediante dos algoritmos de Inteligencia Artificial en la Plataforma Google Earth Engine

German, Leonardo¹, Vitale, Juan¹, Waldman, Cynthia¹ and Castañeda Natalia¹

Instituto de Clima y Agua, CNIA. INTA Castelar,
german.leonardo@inta.gob.ar

Abstract. Las hortalizas altamente percederas son producidas por establecimientos que se encuentran en las periferias de los centros urbanos. En los últimos 30 años el cinturón hortícola del partido de La Plata ha pasado a ser el principal productor de hortalizas de la Provincia de Buenos Aires y del país. La incorporación y expansión de los invernáculos al cultivo de hortalizas a partir de los años 80 demanda la necesidad de conocer su superficie y su distribución. El objetivo de este trabajo fue estimar la superficie con invernáculos existente en el año 2018 en el partido de La Plata mediante dos algoritmos de inteligencia artificial, Support Vector Machine y Random Forest, en la plataforma Google Earth Engine. Ambas clasificaciones resultaron confiables con una exactitud general del 98% para SVM y del 96% para el RF. La superficie estimada de invernáculos fue de aproximadamente 4000 hectáreas.

Keywords: cultivos, Sentinel 2B-1c, Cinturón verde, hortalizas.

1 Introducción

La actividad hortícola Argentina se caracteriza por desarrollarse a lo largo de todo el país y por la amplia diversidad de especies que se cultivan [1]. Buenos Aires es la provincia más destacada por la variedad de su producción hortícola, ocupando el área más extensa del país con el 19,7% de acuerdo al Censo Nacional Agropecuario 2002.

Un cinturón verde se caracteriza por presentar lotes de entre 2 y 15 hectáreas, donde se producen verduras de hoja y frutos de estación que proveen a las ciudades de hortalizas frescas en determinadas épocas del año [2, 3].

El Cinturón Verde de La Plata ha crecido en los últimos 25 años en productividad e importancia a nivel regional, provincial y nacional. Los lotes productivos tienen superficies de entre 2 y 10 hectáreas, con gran variedad de cultivos sobre todo hortalizas de hoja. Uno de los motivos que impulsó este gran crecimiento fue la implementación de la tecnología de invernáculo [2]. A partir de

la década del 90 se produce una fuerte expansión de esta práctica permitiendo generalizar el desarrollo de nuevas formas de cultivo [4].

El invernáculo, es una construcción agrícola cuya estructura es de madera, hierro u otros materiales de altura variable, con una cobertura plástica que permiten acondicionar el ambiente generando un microclima controlado que se ajusta a las necesidades de cada especie [2]. La producción de cultivos en invernáculo, llamados también cultivos protegidos, es una estrategia que permite no solo estabilizar factores meteorológicos sino también controlar el manejo del riego localizado, plagas y densidad de plantación. Mediante su implementación se logra obtener productos de mejor calidad, aumentando el rendimiento, adelantar o retrasar el momento de la cosecha y garantizar una producción hortícola permanente en el tiempo [4].

A través de la plataforma Google Earth Engine (GEE) es posible acceder de manera libre a un catálogo histórico de imágenes satelitales de diversos sensores. Mediante los recursos computacionales disponibles en la plataforma pueden realizarse análisis geoespaciales a distintas escalas. Los algoritmos de aprendizaje supervisado tales como Random Forest [5] y Support Vector Machine [6] no suponen una distribución normal de los datos. Por esta razón, se han vuelto cada vez más utilizados para el análisis de los datos espectrales los cuales no suelen seguir una distribución normal, generando clasificaciones más estables [5]. A partir de la identificación de las clases a diferenciar, los algoritmos se entrenan para predecir la clase de una nueva muestra y se ha observado que su implementación es eficiente en bases de datos grandes como las asociadas a imágenes satelitales de áreas extensas [7].

A la fecha existen pocos datos actualizados acerca de la producción hortícola a escala nacional o regional. La última información disponible a nivel nacional data del Censo Nacional Agropecuario 2002 y a nivel provincial el Censo Frutihortícola 2005, constituyendo los datos estadísticos más actualizados [2].

El objetivo de este trabajo fue comparar y evaluar dos algoritmos de inteligencia artificial en la estimación de la superficie de invernáculos en el cinturón verde de La Plata, Provincia de Buenos Aires en la Plataforma GEE.

2 Materiales y Métodos

El estudio se llevó a cabo sobre el partido de La Plata, provincia de Buenos Aires (ver Fig. 1).

Para el trabajo se utilizaron imágenes del satélite Sentinel 2B-1C de la Agencia Espacial Europea (ESA), debido a su mayor resolución temporal y espacial con respecto a otras imágenes disponibles en la plataforma (5 días y 10 m respectivamente). Se seleccionaron un total de 50 imágenes adquiridas en el período 28/01/2018 al 28/06/2018 y en la clasificación se incluyeron todas sus bandas [8].

Un esquema de la metodología en este estudio se presenta en la Fig. 2.

de estudio.jpg

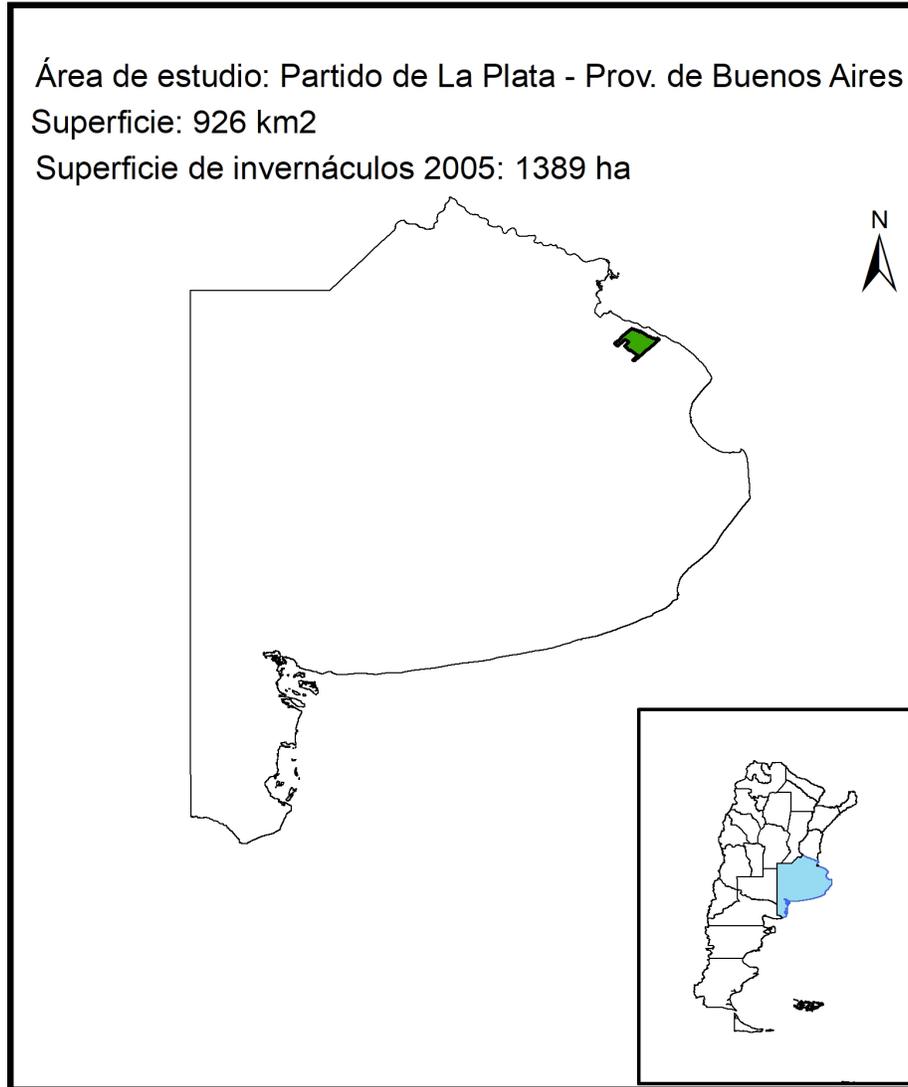


Fig. 1. Partido de La Plata, Provincia de Buenos Aires.

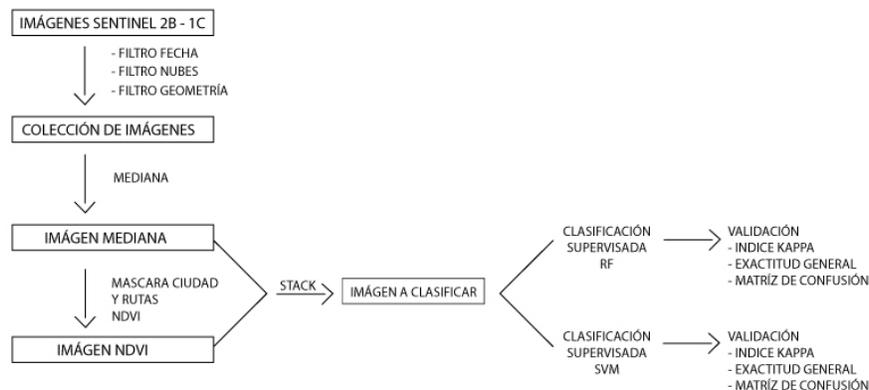


Fig. 2. Metodología aplicada en la plataforma Google Earth Engine para la clasificación de tipos de suelo en el Partido de La Plata.

2.1 Preparación de la imagen a clasificar

A partir de la colección de imágenes obtenida en el período temporal descripto anteriormente se seleccionaron aquellas que presentaron un porcentaje de nubes menor al 10%. Con esta colección de imágenes se generó una sola imagen con igual número de bandas que las imágenes originales, siendo el valor de cada pixel la mediana de los valores de los pixeles de cada banda en su ubicación [9].

A la imagen obtenida se le calculo el índice de vegetación de diferencia normalizada (NDVI) o índice verde [9], cuya expresión es:

$$NDVI = (NIR - VIS) / (NIR + VIS) \quad (1)$$

donde NIR es infrarojo cercano y VIS es rojo visible.

Como último paso se generó una imagen compuesta adicionando la imagen mediana a la del NDVI.

2.2 Clasificación

Para este trabajo se seleccionaron los algoritmos de clasificación Random Forest y Support Vector Machine, disponibles en la plataforma GEE.

Support Vector Machine (SVM): este algoritmo genera un hiperplano óptimo entre las clases a separar utilizando solo algunas de las muestras de entrenamiento que se encuentran en el borde de las distribuciones de clase, a las que se llama vectores soporte. Esto permite la selección de las muestras más informativas para la clasificación [6].

Random Forest: es un algoritmo de aprendizaje automático que construye múltiples árboles de decisión a partir de un subconjunto aleatorio de los datos

de entrenamiento. Cada árbol en el conjunto actúa como un clasificador para determinar la clase de un dato no clasificado. Esto se hace por mayoría de votos, donde cada clasificador emite un voto para su clase predicha, luego la clase con la mayoría de los votos se usa para clasificar el dato [5].

Para Random Forest se dejó libre la cantidad de variables/split y no se restringió la profundidad de los árboles ni el número de individuos/hoja. Esta configuración se consideró óptima debido a la capacidad del algoritmo de reducir sesgo y varianza. Para SVM se utilizó un separador lineal, con un procedimiento de decisión por voting y el tipo de clasificación fue c.SVC.

Para la clasificación se utilizó una máscara del área de estudio, con el fin de excluir de la clasificación las zonas urbanas y rutas, la cual fue generada previamente en el software libre **QGIS** [10]. Se realizó un muestreo a campo registrándose puntos de control aleatorios correspondientes a las clases invernáculos, vegetación arbórea, vegetación no arbórea (huertas y pastizal) y suelo. Se calculó la superficie en hectáreas de la clase invernáculos. Se compararon los datos obtenidos con datos censales de los Censos Hortícolas de Buenos Aires de los años 1998 y 2001 (CHBA 1998 y CHBA 2001 respectivamente) [11], [12], del Censo Nacional agropecuario del año 2002 [13] y los del Censo Hortiflorícola 2005 (CHFBA 2005) [14].

2.3 Medidas de rendimiento de los algoritmos

Se evaluaron las precisiones de los mapas obtenidos través de la matriz de confusión, la precisión general y el índice Kappa para ambas clasificaciones [15]. La matriz de confusión es una matriz de $n \times n$ donde las filas representan las clases reales de la clasificación y las columnas las predicciones de cada clase. La diagonal refleja la suma de todas las predicciones correctas obtenidas por el modelo. Se calculó también la precisión general del modelo para determinar la proporción total de predicciones que fueron correctamente clasificadas y el índice Kappa que es un índice de concordancia que tiene en cuenta la coincidencia entre la imagen clasificada y la verdad de campo debida solo a la exactitud de la clasificación, eliminando la concordancia que cabría esperar solo por el azar.

3 Resultados

La aplicación de los dos clasificadores supervisados dio como resultado una imagen raster con 4 clases: invernáculos, vegetación arbórea, vegetación no arbórea y suelo. La estimación de la superficie de invernáculos en el partido de La Plata dio una superficie aproximada de 4328HA para el clasificador Random Forest (Fig. 3) y de 4090HA para el clasificador Support Vector Machine (Fig. 4). La precisión general e índice Kappa fue del 96% y 0,83 para el primer clasificador y 98% y 0,89 para el segundo (Tabla. 1).

Al graficar los datos censales históricos del período 1998-2005 se observa un aumento constante en la superficie de invernáculos en el partido de La Plata, con una mayor expansión entre los años 2001 y 2002. (Fig. 5).

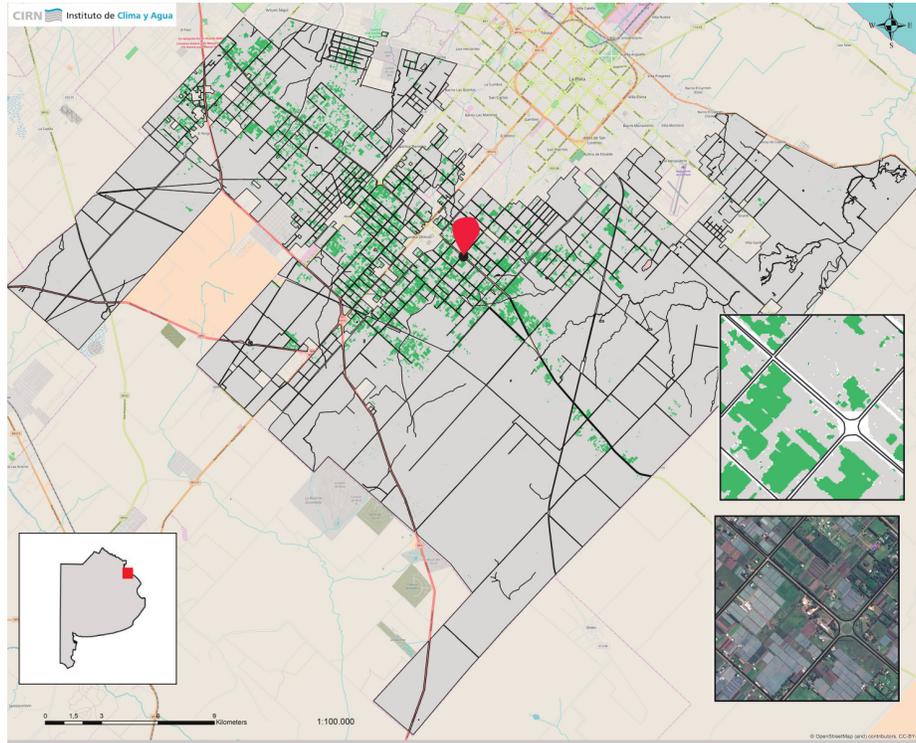


Fig. 3. Mapa de invernáculos en el Cinturón verde de La Plata obtenido con el clasificador Random Forest.

Table 1. Superficies en hectáreas de invernáculos calculadas con los algoritmos Support Vector Machine (SVM) y Random Forest (RF).

	Superficie(HA.)	EG	Indice Kappa
RF	4328	96%	0.83
SVM	4090	98%	0.89

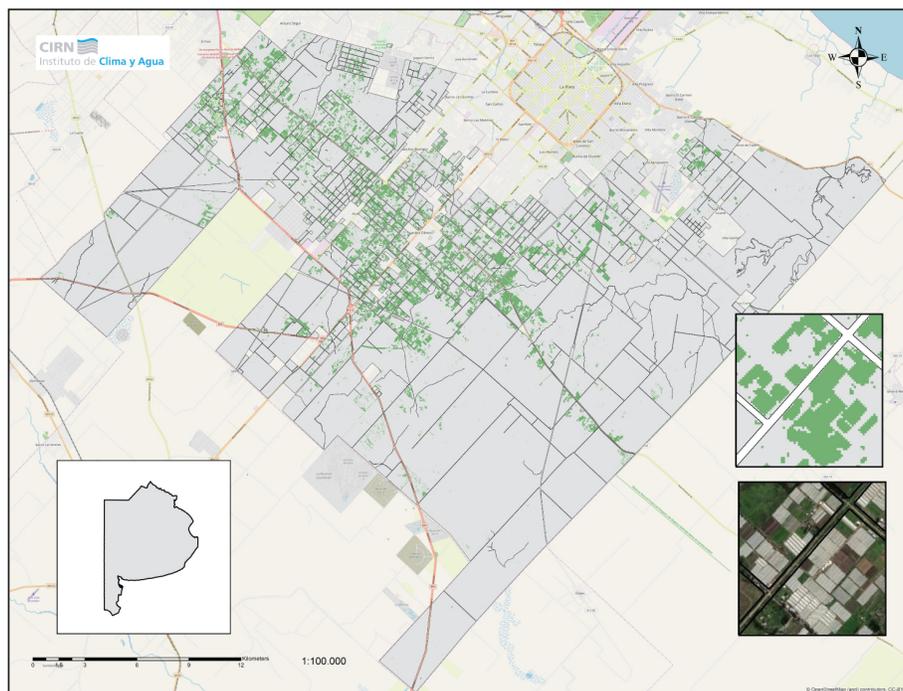


Fig. 4. Mapa de invernáculos en el Cinturón verde de La Plata obtenido con el clasificador Support Vector Machine.

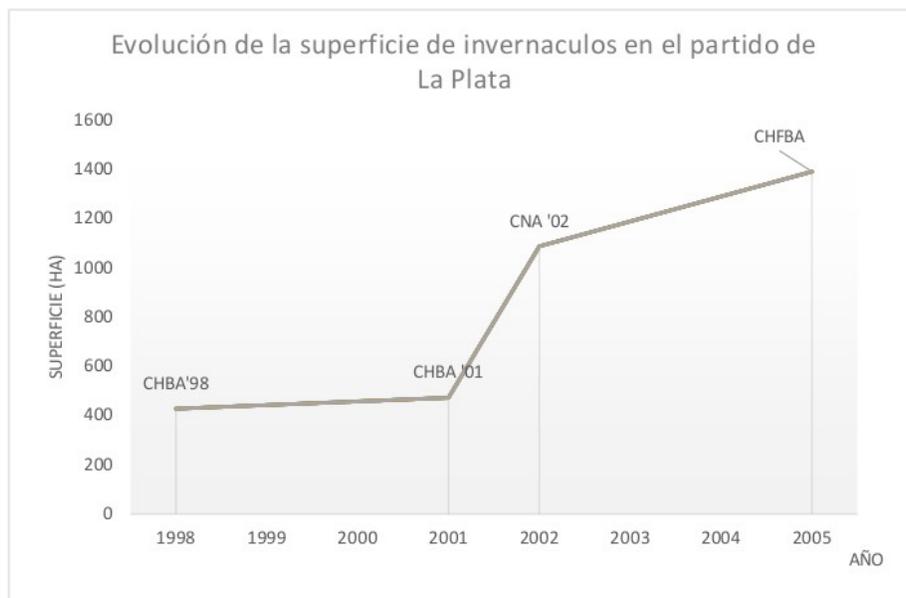


Fig. 5. Superficie de invernáculos en el cinturón verde de la ciudad de La Plata (Fuente: elaboración propia a partir de CHBA1998-2005. CHBA'98: Censo Hortícola de Buenos Aires 1998; CHBA'01: Censo Hortícola de Buenos Aires 2001; CNA'02: Censo nacional Agropecuario 2002; CHFBA'05: Censo Hortiflorícola 2005).

3.1 Matriz de Confusión

En la matriz de confusión para ambos clasificadores en la clase invernáculos se observa poca dispersión de los datos con respecto a la diagonal. Para el clasificador SVM el error de omisión fue de 2% es decir que solo en dos casos los techos fueron clasificados como otra cobertura y en 10 casos la vegetación fue clasificada como techo, dando un error de comisión del 8%. (Tabla 2)

Table 2. Matriz de Confusión clasificación SVM.

		Campo			
		invernáculo	Veg. no arbórea	Veg. arbórea	Suelo Desnudo
Mapa	Invernáculos	119	10	0	0
	Veg. no arbórea	1	2327	9	0
	Veg. arbórea	0	2	165	0
	Suelo desnudo	1	40	0	0

Para el clasificador RF el error de omisión fue de 13% ya que 12 veces los techos fueron clasificados como vegetación y 3 veces como suelo desnudo y en 19 casos la vegetación fue clasificada como techo, dando un error de comisión del 16% (Tabla 3).

Table 3. Matriz de Confusión clasificación RF.

		Campo			
		invernáculo	Veg. no arbórea	Veg. arbórea	Suelo Desnudo
Mapa	invernáculo	100	19	0	0
	Veg. no arbórea	12	2287	14	24
	veg. arbórea	0	3	164	0
	Suelo desnudo	3	2	0	18

4 Conclusiones

El monitoreo de los tipos de suelo a través de los datos adquiridos por sensores remotos resulta de gran utilidad para la clasificación de grandes extensiones. Mediante la plataforma Google Earth Engine es posible clasificar y evaluar distintos algoritmos gracias a su gran potencia y al libre acceso a colecciones de imágenes.

La clasificación de invernáculos mediante los algoritmos SVM y RF produjo resultados confiables, generando datos actualizados sobre la la superficie de está práctica de cultivo.

A partir de los resultados obtenidos, se observa que la incorporación de la tecnología de invernáculo ha tenido un crecimiento constante desde sus inicios a diferente tasa en cada periodo. La metodología aplicada codificada en un script dentro de la plataforma Google Earth Engine puede ser reproducida para otras zonas y mayores extensiones.

References

1. La Horticultura en la Argentina, Informe Final. INET, Ministerio de Educacion (2010)
2. García, M.: El cinturón hortícola platense: ahogándonos en un mar de plásticos. Un ensayo acerca de la tecnología, el ambiente y la política. Revista Theomai N^o 23, 35-53 (2011)
3. Rocco, R., Ruiz Arregui, J.: Logística de Cinturón Hortícola Bonaerense. Facultad de Ingeniería, Universidad Nacional de La Plata (2016)
4. Adlercreutz, E., Huarte, R.: Producción Hortícola bajo cubierta, Ediciones INTA (2013)
5. Belgiu, M., Drăguț, L.: Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31 (2016)
6. Mathur, A., Giles, M. F.: Crop classification by support vector machine with intelligently selected training data for an operational application. International Journal of Remote Sensing 29(8):2227-2240 · April (2008)
7. Nolasco, M., Bocco, M., Scavuzzo, C.: Mapeo de cultivos agrícolas en Córdoba: una comparación del desempeño de tres algoritmos de clasificación. CAI, Congreso Argentino de AgroInformática (2017)
8. Zhang, T., Su, J., Liu, C., and Chen, W.: Band Selection in Sentinel-2 Satellite for Agriculture Applications. Proceedings of the 23rd International Conference on Automation & Computing, University of Huddersfield, Huddersfield, UK, 7-8 September 2017
9. Ganem, K.: Classificação da cobertura do solo na Caatinga a partir do imagens do Landsat-8 E da Ferramenta Google Earth Engine: Uma comparação entre dados com e sem correção atmosférica - Dissertação de maestrando n113, Universidade de Brasília instituto de Geociencias (2017)
10. QGIS 3.4.x LTR, QGIS Development Team, (2019). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
11. Censo Hortícola de Buenos Aires 1998 (CHBA 98). Ministerio de Asuntos Agrarios de la Prov. de Buenos Aires, Instituto Nacional de Estadística y Censos y Secretaría de Agricultura, Ganadería y Pesca de la Nación (1998)
12. Censo Hortícola Bonaerense 2001 (CHBA 01). Ministerio de Agricultura, Ganadería y Alimentación de la Prov. de Buenos Aires, INDEC y Secretaría de Agricultura, Ganadería y Pesca de la Nación. (2001)
13. Censo Nacional Agropecuario 2002, https://www.indec.gob.ar/cna_index.asp
14. Censo Hortiflorícola de Buenos Aires 2005 .Ministerio de Asuntos Agrarios y Ministerio de Economía de la Prov. de Buenos Aires. (2005)
15. Lu, D., and Weng, Q.: A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: <https://www.tandfonline.com/loi/tres20> (2007)

16. García, M.: Horticultura de La Plata (Buenos Aires). Modelo productivo irracionalmente exitoso. Rev. Fac. Agron. La Plata (2015) Vol 114 (Núm. Esp.1) Agricultura Familiar, Agroecología y Territorio: 190-201 (2015)