Land-Use Database Updating with Remote Sensing and GIS:

Object-Based Classification and Change Detection

CZU - Emerging GIS Tools for Agriculture | *Technical Review*

ABSTRACT
This technical review presents a GIS-based workflow for updating a land-use cartographic database. The method integrates cadastral parcel data, land-use tables, high-resolution imagery, object-based feature extraction with FETEX, Sentinel-2 NDVI time series from Google Earth Engine, and Random Forest classification in Weka. The workflow selects 2,107 parcels, defines land-use classes, builds training samples, extracts spectral, texture, shape, and temporal descriptors, and compares predicted classes with cadastral records to identify potential changes. The validation shows practical operational value, while also highlighting class confusion between adult crops and tree-covered parcels.

1. OBJECTIVE AND STUDY AREA

The objective is to update a cartographic land-use database through the interpretation of high-resolution imagery and auxiliary cadastral information. The selected work area contains a diverse agricultural landscape and satisfies the operational requirement of more than 2,000 parcels.

A total of **2,107 parcels** were exported into a separate vector layer. A unique polygon identifier was created with the QGIS field calculator using the row number variable. This identifier became the common key for later joins with feature-extraction outputs and classification results.

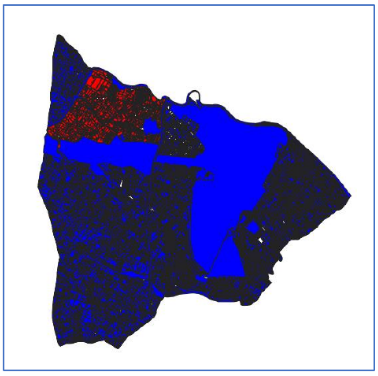


Figure 1. *Selected work area used for the parcel-based land-use update. The red sector marks the selected zone inside the provided cartographic dataset.*

2. DATA PREPARATION AND LAND-USE STRUCTURE

The cadastral tables **RUSUBPARCELA** and **RUCULTIVO** were joined to the parcel shapefile. The CC field was used for table-to-table matching, while a concatenated key derived from MASA, PARCELA, and SUBPARCE enabled the join with the polygon layer.

The final working layer contained **42 attribute columns** and **2,107 parcel records**. This enriched vector layer provided the basis for defining reference classes, selecting training samples, and attaching FETEX, Google Earth Engine, and Weka outputs.

Land-use categories present in the study area

Irrigated citrus; dry carob; riverbank trees; irrigated fruit trees; dry fruit trees; natural hydrography; unproductive land; irrigated arable land; dry arable land; scrubland; pasture; public road domain.

Figure 2. *Spatial distribution of land-use categories by parcel. The map shows the strong dominance of agricultural parcels and smaller areas assigned to non-crop or infrastructure classes.*

3. TRAINING SAMPLE DESIGN IN QGIS

Training samples were selected manually using visual interpretation in QGIS. A binary field separated training parcels from the rest of the layer: **1** for training samples and **0** for non-training parcels.

The modelling classes were simplified into five operational groups: **adult crop**, **young crop**, **unproductive land**, **bare soil**, and **tree-covered land**. This reduction helped translate a complex cadastral taxonomy into classes that could be interpreted from imagery and numerical descriptors.

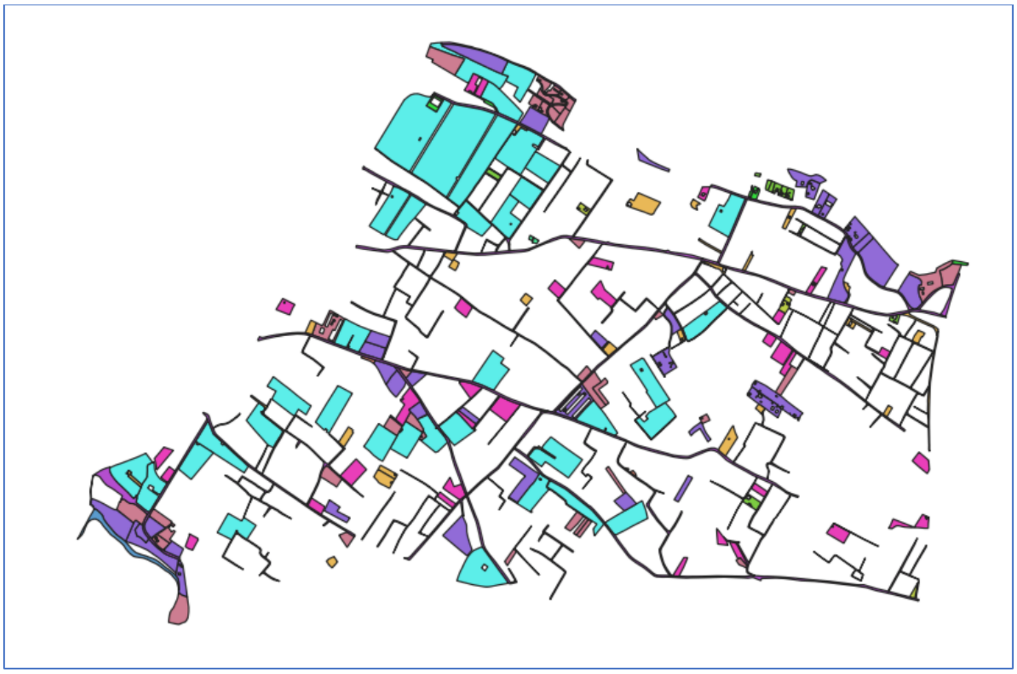


Figure 3. *Spatial distribution of selected training parcels. The samples cover different parts of the study area and represent the main modelling classes used for classification.*

4. FEATURE EXTRACTION AND REMOTE SENSING INPUTS

4.1 FETEX object-based descriptors

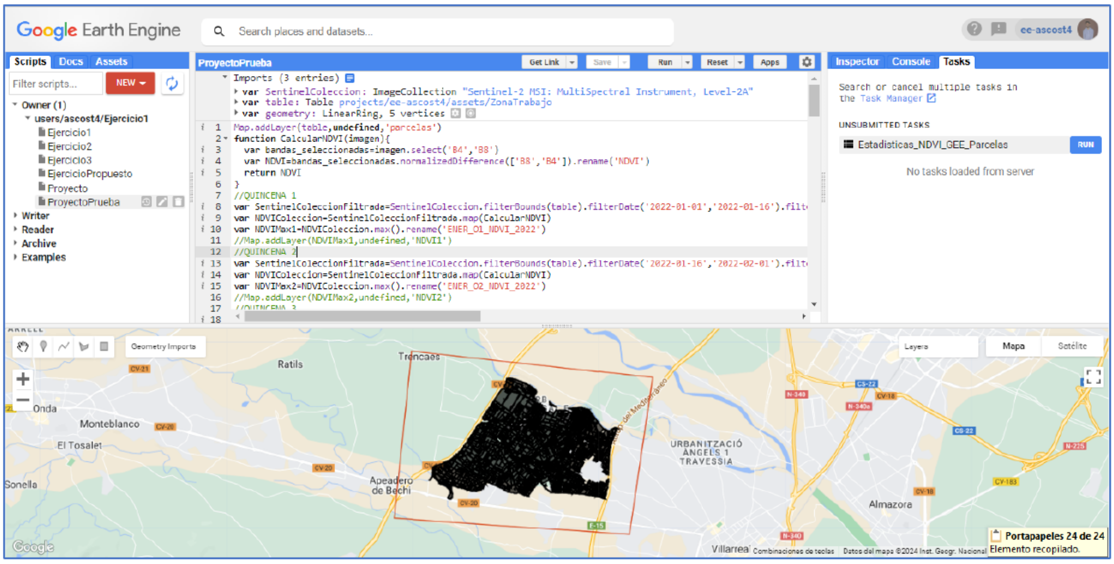
FETEX 2.0 was used to extract numerical descriptors from the cadastral parcel polygons and the input raster imagery. The extracted variables included spectral, textural, shape, and structural characteristics. The resulting DBF output was opened in Excel and converted to CSV for later processing.

4.2 Google Earth Engine NDVI time series

Google Earth Engine was used to derive parcel statistics from **Sentinel-2 MSI Level-2A** imagery. NDVI was calculated from the red band B4 and near-infrared band B8. Two observations were collected per month, one for each half-month period, producing **24 temporal variables** for the analysis year.

The exported CSV tables were joined back to the parcel layer in QGIS using the unique polygon identifier. Field prefixes such as **FE_** and **GEE_** were used to preserve the origin of each variable and keep the attribute table readable.

Tool / source	Role in the workflow	Output
QGIS	Parcel preparation, table joins, sample selection, and final map production.	Enriched vector layer and validation samples.
FETEX 2.0	Automatic extraction of parcel-level descriptors from imagery.	Spectral, texture, shape, and structural variables.
Google Earth Engine	Sentinel-2 NDVI time-series extraction at parcel level.	24 NDVI variables, two per month.
Weka	Machine-learning classification with Random Forest.	Predicted class and evaluation report.



```
Imports (3 entries)
 var SentinelColeccion: ImageCollection "Sentinel-2 MSI: MultiSpectral Instrument, Level-2A"
 var table: Table projects/ee-ascost4/assets/ZonaTrabajo
 var geometry: LinearRing, 5 vertices
Map.addLayer(table,undefined,'parcelas')
function CalcularNDVI(imagen){
  var bandas_seleccionadas=imagen.select('B4','B8')
  var NDVI=bandas_seleccionadas.normalizedDifference(['B8','B4']).rename('NDVI')
  return NDVI
}
//QUINCENA 1
var SentinelColeccionFiltrada=SentinelColeccion.filterBounds(table).filterDate('2022-01-01','2022-01-16').filt
var NDVIColeccion=SentinelColeccionFiltrada.map(CalcularNDVI)
var NDVIMax1=NDVIColeccion.max().rename('ENER_01_NDVI_2022')
//Map.addLayer(NDVIMax1,undefined,'NDVI1')
//QUINCENA 2
var SentinelColeccionFiltrada=SentinelColeccion.filterBounds(table).filterDate('2022-01-16','2022-02-01').filt
var NDVIColeccion=SentinelColeccionFiltrada.map(CalcularNDVI)
var NDVIMax2=NDVIColeccion.max().rename('ENER_02_NDVI_2022')
//Map.addLayer(NDVIMax2,undefined,'NDVI2')
```

Search or cancel multiple tasks in the Task Manager

UNSUBMITTED TASKS

Estadisticas_NDVI_GEE_Parcelas

No tasks loaded from server

Figure 4. *Google Earth Engine workspace used to load the parcel shapefile, calculate NDVI variables, and export parcel-level statistics for later joining in QGIS.*

5. RANDOM FOREST CLASSIFICATION

After feature extraction, the attribute table was filtered to keep the fields relevant to classification. The polygon identifier was removed before modelling so it would not influence the classification process. The resulting dataset was loaded into Weka.

The classification was performed with **Random Forest** and evaluated by **10-fold cross-validation**. Several trials were tested, including configurations with 200 and 300 trees. The best reported result reached **61.64%** correctly classified instances over 219 labelled samples.

Metric	Value
Correctly classified instances	135 / 219 (61.64%)
Incorrectly classified instances	84 / 219 (38.36%)
Kappa statistic	0.4728
Mean absolute error	0.1529
Root mean squared error	0.3843

Reference class	Adult crop	Young crop	Unproductive	Bare soil	Tree-covered
Adult crop	17	3	1	0	13
Young crop	7	15	3	1	13
Unproductive	4	6	14	0	16
Bare soil	5	4	0	13	6
Tree-covered	2	0	0	0	76

> *The strongest source of confusion occurs between adult crops and tree-covered parcels. Both classes contain dense vegetation, so their spectral and structural descriptors can overlap when the parcel boundary includes trees or mixed cover.*

6. CHANGE DETECTION AND VALIDATION

The Weka predictions were joined back to the QGIS parcel layer and compared with cadastral land-use information. Because the cadastral categories and the modelling classes do not describe the landscape at the same thematic level, the comparison was treated as an indicator of possible change rather than a direct one-to-one replacement.

A validation sample of **100 random parcels** was created across the study area. Each parcel was inspected using the orthophoto and assigned to one of four outcomes: true positive, false positive, false negative, or true negative.

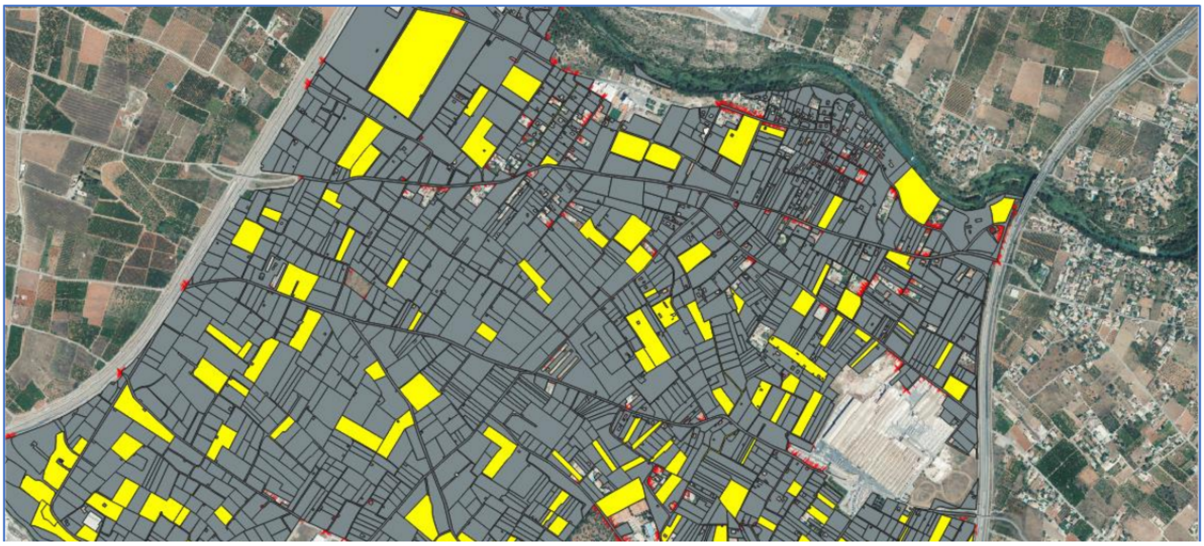


Figure 5. *Random validation parcels distributed across the study area. These parcels were used to compare change indications with orthophoto interpretation.*

Prediction / reference	Reference positive	Reference negative	Total
Predicted positive	36	24	60
Predicted negative	22	18	40
Total	58	42	100

Validation measure	Formula	Value
Sensitivity / true positive rate	TP / (TP + FN)	0.62
Specificity / true negative rate	TN / (TN + FP)	0.43
Precision	TP / (TP + FP)	0.60
Overall accuracy	(TP + TN) / Total	0.54
False positive rate	FP / (FP + TN)	0.57
False negative rate	FN / (FN + TP)	0.38

The model correctly identified 54% of the validation parcels. The 60% precision indicates that most positive change indications were plausible, but the low specificity shows that negative cases remain difficult to classify reliably.

7. METHODOLOGICAL ASSESSMENT AND IMPROVEMENT OPTIONS

The workflow provides a practical basis for land-use database updating because it combines parcel-level GIS operations, remote-sensing indices, feature extraction, machine-learning classification, and manual validation. Its main value lies in prioritising parcels that require review.

The main limitation is thematic ambiguity. Several parcels contain more than one visible cover type, and the parcel-level approach forces each geometry into a single class. This creates errors when one parcel contains both young and adult crop patterns, or when agricultural trees resemble natural tree-covered areas.

Figure 6. *Example of a parcel with mixed cover. A single parcel may contain both young and adult crop patterns, making one-label classification difficult.*

Improvement option	Expected benefit
Increase and balance training samples	Reduces bias and improves class separation.
Add specialised classifiers for buildings and water	Separates non-agricultural objects before crop classification.
Use sub-parcel segmentation where parcels are heterogeneous	Avoids forcing mixed parcels into one class.
Add Copernicus layers such as CORINE Land Cover	Provides thematic context to reduce confusion with tree-covered areas.
Report class-level accuracy and independent hold-out validation	Improves transparency and comparability.

8. CONCLUSIONS

The workflow demonstrates that land-use database updates can be supported without systematic field visits when reliable imagery, parcel geometry, auxiliary cadastral information, and good training samples are available. The process is suitable as a decision-support method for detecting parcels that may require review.

The classification result is moderate rather than definitive. It reaches 61.64% accuracy in cross-validation and 54% accuracy in the change-detection validation sample. The main problem is confusion between adult crops, tree-covered parcels, and some unproductive areas.

Despite these limitations, the remote-sensing classification can reveal inconsistencies in cadastral data and guide targeted inspection. The most robust operational use is not automatic replacement of the database, but prioritised updating supported by GIS, satellite-derived variables, and expert visual verification.

> *Remote sensing and GIS reduce the cost of land-use revision by identifying where the database is most likely outdated. Field verification should then focus on those high-priority parcels.*

9. REFERENCES AND FURTHER READING

QGIS Project. QGIS Geographic Information System. https://qgis.org

Google Earth Engine. Cloud-based geospatial analysis platform. https://earthengine.google.com

Copernicus Data Space Ecosystem. Sentinel data access. https://dataspace.copernicus.eu

ESA Sentinel-2 Mission Overview. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2

Weka Machine Learning Software, University of Waikato. https://www.cs.waikato.ac.nz/ml/weka/

Copernicus Land Monitoring Service. CORINE Land Cover. https://land.copernicus.eu/en/products/corine-land-cover

Keywords: GIS · Remote Sensing · Land-Use Updating · QGIS · FETEX · Google Earth Engine · Sentinel-2 · NDVI · Weka · Random Forest · Change Detection