Davis, CA

Land Cover

Methodology & Accuracy Assessment (2022)

Photo credit: Leland Wong

Davis CALIFORNIA

DAVEY Resource Group

Methodology and Accuracy Assessment

Davey Resource Group Classification Methodology

Davey Resource Group, Inc. utilized an object-based image analysis (OBIA) semi-automated feature extraction method to process and analyze current high-resolution color infrared (CIR) aerial imagery and remotely-sensed data to identify tree canopy cover and land cover classifications. The use of imagery analysis is cost-effective and provides a highly accurate approach to assessing your community's existing tree canopy coverage. This supports responsible tree management, facilitates community forestry goal-setting, and improves urban resource planning for healthier and more sustainable urban environments.

Advanced image analysis methods were used to classify, or separate, the land cover layers from the overall imagery. The semi-automated extraction process was completed using Feature Analyst, an extension of ArcGIS®. Feature Analyst uses an object-oriented approach to cluster together objects with similar spectral (i.e., color) and spatial/contextual (e.g., texture, size, shape, pattern, and spatial association) characteristics. The land cover results of the extraction process was post-processed and clipped to each project boundary prior to the manual editing process in order to create smaller, manageable, and more efficient file sizes. Secondary source data, high-resolution aerial imagery provided by each UTC city, and custom ArcGIS® tools were used to aid in the final manual editing, quality checking, and quality assurance processes (QA/QC). The manual QA/QC process was implemented to identify, define, and correct any misclassifications or omission errors in the final land cover layer.

Classification Workflow

1) Prepare imagery for feature extraction (resampling, rectification, etc.), if needed.
2) Gather training set data for all desired land cover classes (canopy, impervious, grass, bare soil, shadows). Water samples are not always needed since hydrologic data are available for most areas. Training data for impervious features were not collected because the City maintained a completed impervious layer.
3) Extract canopy layer only; this decreases the amount of shadow removal from large tree canopy shadows. Fill small holes and smooth to remove rigid edges.
4) Edit and finalize canopy layer at 1:2000 scale. A point file is created to digitize-in small individual trees that will be missed during the extraction. These points are buffered to represent the tree canopy. This process is done to speed up editing time and improve accuracy by including smaller individual trees.
5) Extract remaining land cover classes using the canopy layer as a mask; this keeps canopy shadows that occur within groups of canopy while decreasing the amount of shadow along edges.
6) Edit the impervious layer to reflect actual impervious features, such as roads, buildings, parking lots, etc. to update features.
7) Using canopy and actual impervious surfaces as a mask; input the bare soils training data and extract them from the imagery. Quickly edit the layer to remove or add any features. Davey Resource Group tries to delete dry vegetation areas that are associated with lawns, grass/meadows, and agricultural fields.

8) Assemble any hydrological datasets, if provided. Add or remove any water features to create the hydrology class. Perform a feature extraction if no water feature datasets exist.

9) Use geoprocessing tools to clean, repair, and clip all edited land cover layers to remove any self-intersections or topology errors that sometimes occur during editing.

10) Input canopy, impervious, bare soil, and hydrology layers into Davey Resource Group's Five-Class Land Cover Model to complete the classification. This model generates the pervious (grass/low-lying vegetation) class by taking all other areas not previously classified and combining them.

11) Thoroughly inspect final land cover dataset for any classification errors and correct as needed.

12) Perform accuracy assessment. Repeat Step 11, if needed.

Automated Feature Extraction Files

The automated feature extraction (AFE) files allow other users to run the extraction process by replicating the methodology. Since Feature Analyst does not contain all geoprocessing operations that Davey Resource Group utilizes, the AFE only accounts for part of the extraction process. Using Feature Analyst, Davey Resource Group created the training set data, ran the extraction, and then smoothed the features to alleviate the blocky appearance. To complete the actual extraction process, Davey Resource Group uses additional geoprocessing tools within ArcGIS®. From the AFE file results, the following steps are taken to prepare the extracted data for manual editing.

1) Davey Resource Group fills all holes in the canopy that are less than 30 square meters. This eliminates small gaps that were created during the extraction process while still allowing for natural canopy gaps.

2) Davey Resource Group deletes all features that are less than 9 square meters for canopy (50 square meters for impervious surfaces). This process reduces the amount of small features that could result in incorrect classifications and also helps computer performance.

3) The Repair Geometry, Dissolve, and Multipart to Singlepart (in that order) geoprocessing tools are run to complete the extraction process.

4) The Multipart to Singlepart shapefile is given to GIS personnel for manual editing to add, remove, or reshape features.

Table 1. Land Cover Classification Code Values

Land Cover Classification	Code Value
Tree Canopy	1
Impervious	2
Pervious (Grass/Vegetation)	3
Bare Soil	4
Open Water	5

Accuracy Assessment Protocol

Determining the accuracy of spatial data is of high importance to Davey Resource Group and our clients. To achieve to best possible result, Davey Resource Group manually edits and conducts thorough QA/QC checks on all urban tree canopy and land cover layers. A QA/QC process will be completed using ArcGIS® to identify, clean, and correct any misclassification or topology errors in the final land cover dataset. The initial land cover layer extractions will be edited at a 1:2000 quality control scale in the urban areas and at a 1:2500 scale for rural areas utilizing the most current high-resolution aerial imagery to aid in the quality control process.

To test for accuracy, random plot locations are generated throughout the city area of interest and verified to ensure that the data meet the client standards. Each point will be compared with the most current NAIP high-resolution imagery (reference image) to determine the accuracy of the final land cover layer. Points will be classified as either correct or incorrect and recorded in a classification matrix. Accuracy will be assessed using four metrics: overall accuracy, kappa, quantity disagreement, and allocation disagreement. These metrics are calculated using a custom Excel® spreadsheet.

Land Cover Accuracy

The following describes Davey Resource Group's accuracy assessment techniques and outlines procedural steps used to conduct the assessment.

1. *Random Point Generation*—Using ArcGIS, 1,150 random assessment points are generated.

2. *Point Determination*—Each point is carefully assessed by the GIS analyst for likeness with the aerial photography. To record findings, two new fields, CODE and TRUTH, are added to the accuracy assessment point shapefile. CODE is a numeric value (1–5) assigned to each land cover class (Table 1) and TRUTH is the actual land cover class as identified according to the reference image. If CODE and TRUTH are the same, then the point is counted as a correct classification. Likewise, if the CODE and TRUTH are not the same, then the point is classified as incorrect. In most cases, distinguishing if a point is correct or incorrect is straightforward. Points will rarely be misclassified by an egregious classification or editing error. Often incorrect points occur where one feature stops and the other begins.

3. *Classification Matrix*—During the accuracy assessment, if a point is considered incorrect, it is given the correct classification in the TRUTH column. Points are first assessed on the NAIP imagery for their correctness using a "blind" assessment—meaning that the analyst does not know the actual classification (the GIS analyst is strictly going off the NAIP imagery to determine cover class). Any incorrect classifications found during the "blind" assessment are scrutinized further using sub-meter imagery provided by the client to determine if the point was incorrectly classified due to the fuzziness of the NAIP imagery or an actual misclassification. After all random points are assessed and recorded; a classification (or confusion) matrix is created. The classification matrix for this project is presented in Table 2. The table allows for assessment of user's/producer's accuracy, overall accuracy,

omission/commission errors, kappa statistics, allocation/quantity disagreement, and confidence intervals (Figure 1 and Table 3).

Table 2. Classification Matrix

	Classes	Tree Canopy	Impervious Surfaces	Grass & Low-Lying Veg.	Bare Soils	Open Water	Row Total	Producer's Accuracy	Errors of Omission
Reference Data	Tree Canopy	272	4	5	0	0	281	**96.80%**	**3.20%**
	Impervious	5	393	10	2	0	410	**95.85%**	**4.15%**
	Grass/Low Veg.	7	7	164	1	0	179	**91.62%**	**8.38%**
	Bare Soils	3	8	4	111	0	126	**88.10%**	**11.90%**
	Water	0	0	0	0	4	4	**100.00%**	**0.00%**
	Column Total	287	412	183	114	4	1000		
	User's Accuracy	**94.77%**	**95.39%**	**89.62%**	**97.37%**	**100.00%**		**Overall Accuracy**	**94.40%**
	Errors of Commission	**5.23%**	**4.61%**	**10.38%**	**2.63%**	**0.00%**		**Kappa Coefficient**	**0.9204**

4. Following are descriptions of each statistic as well as the results from some of the accuracy assessment tests.

   Overall Accuracy – Percentage of correctly classified pixels; for example, the sum of the diagonals divided by the total points ((272+393+164+111+4)/1,000 = 94.4%).

   User's Accuracy – Probability that a pixel classified on the map actually represents that category on the ground (correct land cover classifications divided by the column total [272/287 = 94.77%]).

   Producer's Accuracy – Probability of a reference pixel being correctly classified (correct land cover classifications divided by the row total [272/281 = 96.8%]).

   Kappa Coefficient – A statistical metric used to assess the accuracy of classification data. It has been generally accepted as a better determinant of accuracy partly because it accounts for random chance agreement. A value of 0.80 or greater is regarded as "very good" agreement between the land cover classification and reference image.

   Errors of Commission – A pixel reports the presence of a feature (such as trees) that, in reality, is absent (no trees are actually present). This is termed as a false positive. In the matrix below, we can determine that 5.23% of the area classified as canopy is most likely not canopy.

Errors of Omission – A pixel reports the absence of a feature (such as trees) when, in reality, they are actually there. In the matrix below, we can conclude that 3.2% of all canopy classified is actually classified as another land cover class.

Allocation Disagreement – The amount of difference between the reference image and the classified land cover map that is due to less than optimal match in the spatial allocation (or position) of the classes.

Quantity Disagreement – The amount of difference between the reference image and the classified land cover map that is due to less than perfect match in the proportions (or area) of the classes.

Confidence Intervals – A confidence interval is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate. Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter based on the observed probability of successes and failures. Since all assessments have innate error, defining a lower and upper bound estimate is essential.

Table 3. Confidence Intervals

Confidence Intervals

Class	Acreage	Percentage	Lower Bound	Upper Bound
Tree Canopy	1,681.2	26.2%	25.6%	26.7%
Impervious Surfaces	2,823.5	43.9%	43.3%	44.6%
Grass & Low-Lying Vegetation	1,343.2	20.9%	20.4%	21.4%
Bare Soils	538.2	8.4%	8.0%	8.7%
Open Water	40.5	0.6%	0.5%	0.7%
Total	**6,426.5**	**100.00%**		

Statistical Metrics Summary	
Overall Accuracy =	94.4%
Kappa Coefficient =	0.9204
Allocation Disagreement =	4%
Quantity Disagreement =	1%

Accuracy Assessment

Class	User's Accuracy	Lower Bound	Upper Bound	Producer's Accuracy	Lower Bound	Upper Bound
Tree Canopy	94.8%	93.5%	96.1%	96.8%	95.7%	97.8%
Impervious Surfaces	95.4%	94.4%	96.4%	95.9%	94.9%	96.8%
Grass & Low-Lying Vegetation	89.6%	87.4%	91.9%	91.6%	89.5%	93.7%
Bare Soils	97.4%	95.9%	98.9%	88.1%	85.2%	91.0%
Open Water	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Rights-of-Way

The canopy rights-of-way (ROW) analysis is used to determine the tree canopy cover along roadways. Potential tree planting areas can be identified by areas with lower tree canopy. Planting in these areas will increase the total canopy along the street and help to reduce stormwater runoff and urban heat islands.

To perform this analysis, a 60 foot buffer was created using the street segment centerlines layer (USDA) and the ROW (city of Davis). An intersection tool was used to identify the intersecting areas between Davis's ROWs and the street centerline segments, which defines segments of the ROW with known acreages. The acreage of canopy within the boundary is divided by the total acreage within each segment to calculate the percentage of canopy.