Training Samples

Idealized sequence for selecting training samples:

Assemble external information from maps, photos, etc.

Conduct field studies to gain firsthand knowledge of study area. Collect field observations with the aid of maps, record approximate areas, condition of features, etc.

Preliminary exam of image to familiarize and assess quality.

Identify training samples.

Inspect frequency histograms to assess usefulness and if necessary modify samples.

Implement classification.
Signature Separability Listing

File:
Distance measure: Euclidean Distance
Using bands: 1 2 3 4 5 6 7
Taken 7 at a time

   Class
1    Class 1
2    Class 2
3    Class 3
4    Class 4
5    Class 5
6    Class 6
7    Class 7
8    Class 8
9    Class 9
10   Class 10

Best Minimum Separability

<table>
<thead>
<tr>
<th>Bands</th>
<th>AVE</th>
<th>MIN</th>
<th>Class Pairs:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: 2 1: 3 1: 4 1: 5 1: 6 1: 7 1: 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: 9 1:10 2: 3 2: 4 2: 5 2: 6 2: 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2: 8 2: 9 2:10 3: 4 3: 5 3: 6 3: 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3: 8 3: 9 3:10 4: 5 4: 6 4: 7 4: 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4: 9 4:10 5: 6 5: 7 5: 8 5: 9 5:10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6: 7 6: 8 6: 9 6:10 7: 8 7: 9 7:10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8: 9 8:10 9:10</td>
</tr>
</tbody>
</table>

5 6 7
| 113 154 10 14 22 26 36 |
| 47 70 115 9 17 17 29 |
| 39 62 107 9 14 22 34 |
| 57 102 13 14 27 49 94 |
| 14 22 45 90 14 35 80 |
| 24 68 45 |
**Supervised Classification Algorithms**

**ISODATA (hybrid):** Based on minimum distance but with changing centroids. If training samples are particularly variable centroids are repeatedly altered until no more changes occur.

**Maximum Likelihood Classification:** Based on probability principles (rather than deterministic) that take into consideration the continuity of the natural environment. Training sample clusters are represented by probabilities of membership. If overlapping classes, pixels are assigned to greatest membership. Requires Gaussian distributions, and highly sensitive to training samples.

**FIGURE 11.19.** Maximum likelihood classification. These frequency distributions represent pixels from two training fields; the zone of overlap depicts pixel values common to both categories. The relations of the values within the overlap region to the overall frequency distribution for each class forms the basis for assigning pixels to classes.
Maximum likelihood classification

Digital number band 3

Digital number band 4

water
urban
forest
crop
soil
heather
Decision Tree/Layered Classification

Hierarchical techniques designed to “simplify” the classification process.

Classes are systematically separated in order to reduce variability. For example, water is classified and then removed from the image; next forest is classified and removed from the image, and so on, until the most variable classes remain.

Alternatively, spatial subsets cut the image into more manageable pieces.
Textural Classifiers

Measure of spectral variability: distinctive spectral and spatial relationships between neighboring pixels.

E.g. rougher textures represent lower-density residential rather than higher-density; rougher textures represent mixed vegetation rather than homogeneous species.

Problems with defining the scale of texture and number of neighboring pixels.
Spatial (contextual) Classifiers

Spectral information is seldom enough to produce accurate classifications. Information on the geometric configuration of neighboring pixels also provide important information.

Attempts to bring in human preconceptions of land cover spatial structure. For e.g. linear arrangement of roads, rivers, rail, rectangular patterns of fields and buildings; all at variable scale.

Idea of “photomorphic” regions being “objects” of homogeneity. Rise of object-based classification using “spatial metrics” (fractals, contagions, fragmentation, area/perimeter ratio, etc.).
Ancillary Data

Ancillary or collateral or subsidiary data are all “non-spectral”, i.e. from beyond remote sensing. Usually, topographic, GPS readings, DEMs, etc. However, most data (esp. environmental) have some RS derivation.

Aimed at improving or updating classifications. E.g. relief used to stratify images and determine correct tree species at various elevation/slopes; census data used for urban identification.

Problems of compatibility (formats, positional, class definitions, time/season, reliability, etc.).
Many pixels are **mixed pixels** ("mixels"). Especially when the IFOV is **larger** than the spatial variability and compositional heterogeneity of the land cover.
What constitutes a mixel is a function of **class semantics and class definitions** (i.e., application). Because the world is infinitely variable, class definition is based on human interpretation. There are no clear-cut boundaries in the world.
Soft Classification Linear mixture models and fuzzy sets theory determine a relationships between pixels and proportions of class membership. So that each pixel may belong to more than one class.

Mirror continuity of nature and address the mixel problem.

Problems include pragmatic interpretation (no single label) and not as popular as first imagined.
Accuracy Assessment

Errors in classification are numerous (bad training samples, inappropriate number of classes, bad labeling, mis-registration, too many mixels, etc.).

Accuracy defined as “correctness”. Precision as “exactness”. Both dependent on scale and application. E.g. general forests more accurate than detailed species of plants.

Accuracy assessment based on “more reliable” information than remote sensing. Preferably ground surveys. Determine how many of a sample of pixels are correctly classified using a confusion matrix. Numbers in the diagonals are correct, numbers in the off-diagonal are incorrect. The process is highly simplistic at the nominal scale of measurement. Errors of omission (producer’s accuracy) (e.g. sample of agriculture on ground omitted from agriculture class. Errors of commission (consumer’s accuracy) (e.g. sample of agriculture on ground included in forest class). Kappa coefficients more elaborate accuracy measure using above.
<table>
<thead>
<tr>
<th>Classification data</th>
<th>W</th>
<th>S</th>
<th>F</th>
<th>U</th>
<th>C</th>
<th>H</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>480</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>485</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>52</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>72</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>313</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>353</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>126</td>
<td>0</td>
<td>0</td>
<td>142</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>342</td>
<td>79</td>
<td>459</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>24</td>
<td>60</td>
<td>359</td>
<td>481</td>
</tr>
<tr>
<td>Column total</td>
<td>480</td>
<td>68</td>
<td>356</td>
<td>248</td>
<td>402</td>
<td>438</td>
<td>1992</td>
</tr>
</tbody>
</table>

**Producer's Accuracy**

- W = \( \frac{480}{480} = 100\% \)
- S = \( \frac{052}{068} = 76\% \)
- F = \( \frac{313}{356} = 88\% \)
- U = \( \frac{126}{248} = 51\% \)
- C = \( \frac{342}{402} = 85\% \)
- H = \( \frac{359}{438} = 82\% \)

**User's Accuracy**

- W = \( \frac{480}{485} = 99\% \)
- S = \( \frac{052}{072} = 72\% \)
- F = \( \frac{313}{353} = 87\% \)
- U = \( \frac{126}{142} = 89\% \)
- C = \( \frac{342}{459} = 74\% \)
- H = \( \frac{359}{481} = 75\% \)

**Overall accuracy** = \( \frac{480 + 52 + 313 + 126 + 342 + 359}{1992} = 84\% \)

\(^a\text{W, water; S, sand; F, forest; U, urban; C, corn; H, hay.}\)