Hard vs Soft Classification

Hard- versus soft-classifiers

Why use soft-classifiers?
- Sub-pixel classification
- Uncertainty of classification/scheme
- Incorporating ancillary data (hardeners)
Soft Classification Scheme

Conventional Classification

- True
  - Water
  - Forested Wetland
  - Upland Forest

- False

Hypothetical Near-infrared Brightness Value

Fuzzy Classification

- Membership
  - Water
  - Forested Wetland
  - Upland Forest

Hypothetical Near-infrared Brightness Value
(Soft or Fuzzy) Signatures

- Training sites (homogeneous vs. fuzzy)

<table>
<thead>
<tr>
<th></th>
<th>Water</th>
<th>Forested Wetland</th>
<th>Upland Forest</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site#1</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Site#2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td>Site#3</td>
<td>0.5</td>
<td>0.1</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>


• Hard Classification - a pixel can only have one and only one category.
• In urban regions, a pixel in reality may have more than one category because of the heterogeneity of the land cover composing that pixel. We call this a mixed pixel.
• Soft or fuzzy classifiers - a pixel does not belong fully to one class but it has different degrees of membership in several classes. The mixed pixel problem is more pronounced in lower resolution data. In fuzzy classification, or pixel unmixing, the proportion of the land cover classes from a mixed pixel is calculated.
Fuzzy logic was developed by Lotfi A. Zadeh in the 1960s in order to provide mathematical rules and functions which permitted natural language queries. Fuzzy logic provides a means of calculating intermediate values between absolute true and absolute false with resulting values ranging between 0.0 and 1.0. With fuzzy logic, it is possible to calculate the degree to which an item is a member. For example, if a person is .83 of tallness, they are "rather tall." Fuzzy logic calculates the shades of gray between black/white and true/false.

Fuzzy logic is a super set of conventional (or Boolean) logic and contains similarities and differences with Boolean logic. Fuzzy logic is similar to Boolean logic, in that Boolean logic results are returned by fuzzy logic operations when all fuzzy memberships are restricted to 0 and 1. Fuzzy logic differs from Boolean logic in that it is permissive of natural language queries and is more like human thinking; it is based on degrees of truth.
FUZZY

BOOLEAN

Graph 1:

Graph 2:
Fuzzy logic may appear similar to probability and statistics as well. Although, fuzzy logic is different then probability even though the results appear similar. The probability statement, "There is a 70% chance that Bill is tall" supposes that Bill is either tall or he is not. There is a 70% chance that we know which set Bill belongs. The fuzzy logic statement, "Bill's degree of membership in the set of tall people is .80" supposes that Bill is rather tall. The fuzzy logic answer determines not only the set which Bill belongs, but also to what degree he is a member. There are no probability statements that pertain to fuzzy logic. Fuzzy logic deals with the degree of membership.

Fuzzy logic has been applied in many areas; it is used in a variety of ways. Household appliances such as dishwashers and washing machines use fuzzy logic to determine the optimal amount of soap and the correct water pressure for dishes and clothes. Fuzzy logic is even used in self-focusing cameras. Expert systems, such as decision-support and meteorological systems, use fuzzy logic. Fuzzy logic has many varied applications.
Fuzzy Sets and Traditional Sets

A fuzzy set is a set whose elements have degrees of membership. That is, a member of a set can be full member (100% membership status) or a partial member (e.g., less than 100% membership and greater than 0% membership). To fully understand fuzzy sets, one must first understand traditional sets.

A traditional or crisp set can formally be defined as the following:

• A subset U of a set S is a mapping from the elements of S to the elements of the set \{0,1\}. This is represented by the notation:
  
  U: S → \{0,1\}

• The mapping is represented by one ordered pair for each element S where the first element is from the set S and the second element is from the set \{0,1\}. The value zero represents non-membership, while the value one represents membership.

Essentially this says that an element of the set S is either a member or a non-member of the subset U. There are no partial members in traditional sets.
Here is an example of a traditional set:
Consider a set $X$ that contains all the real numbers between 0 and 10 and a subset $A$ of the set $X$ that contains all the real numbers between 5 and 8. Subset $A$ is represented in the figure below.

In the figure, the interval on the x-axis between 5 and 8 has y-value of one. This indicates that any number in this interval is a member of the subset $A$. Any number that has a y-value of zero is considered to be a non-member of the subset $A$. 
Again a **fuzzy set** is a set whose elements have degrees of membership. These can formally be defined as the following:

A fuzzy subset $F$ of a set $S$ can be defined as a set of ordered pairs. The first element of the ordered pair is from the set $S$, and the second element from the ordered pair is from the interval $[0,1]$. The value **zero** is used to represent non-membership; the value **one** is used to represent complete membership, and the values **in between** are used to represent degrees of membership.
EXAMPLE 1
Here is an example describing a set of young people using fuzzy sets. In general, young people range from the age of 0 to 20. But, if we use this strict interval to define young people, then a person on his 20th birthday is still young (still a member of the set). But on the day after his 20th birthday, this person is now old (not a member of the young set).

*How can one remedy this?*

By RELAXING the boundary between the strict separation of young and old. This separation can easily be relaxed by considering the boundary between young and old as "fuzzy". The figure below graphically illustrates a fuzzy set of young and old people.

![Graph](image)

Notice in the figure that people whose ages are >= zero and <= 20 are complete members of the young set (that is, they have a membership value of one). Also note that people whose ages are > 20 and < 30 are partial members of the young set. For example, a person who is 25 would be young to the degree of 0.5. Finally people whose ages are >= 30 are non-members of the young set.
**Membership Functions**

A **membership function** is a mathematical function which defines the degree of an element's membership in a fuzzy set. The best way to illustrate this concept is with an example. This example describes a fuzzy set for tallness. Below is the membership function for tallness.

```
tall(x) = { 0, if height(x) < 5ft, (height(x)-5ft)/2, if 5ft <= height(x) <= 7ft, 1, if height(x) > 7ft }
```

Essentially this function calculates the membership value of a certain height. For example, if a person is less 4'9", then this person has a membership value of 0.0 and thus is not a member of the set tall. If a person is 7'6", then this person has a membership value of 1.0 and thus is a member of the set tall. Finally, if a person is 5'5", then this person has a membership value of 0.21 and is a partial member of the set tall.
Below is a graphical representation of the fuzzy set for tallness.
Logical Operations on Fuzzy Sets

Now that we understand what fuzzy sets and membership functions are, we can discuss three basic operation on sets: negation, intersection, and union of fuzzy sets. In L.A. Zadeh first paper, he formally defined these operations in the following manner:

**Negation**

\[
\text{membership\_value(\text{not } x)} = 1 - \text{membership\_value}(x)
\]

where \( x \) is the fuzzy set being negated

**Intersection**

\[
\text{membership\_value}(x \text{ and } y) = \min(\text{membership\_value}(x), \text{membership\_value}(y))
\]

where \( x \) and \( y \) are the fuzzy set being negated

**Union**

\[
\text{membership\_value}(x \text{ or } y) = \max(\text{membership\_value}(x), \text{membership\_value}(y))
\]

where \( x \) and \( y \) are the fuzzy set being negated
**Negation**

In this figure, the red line is a fuzzy set. To negate this fuzzy set, subtract the membership value in the fuzzy set from one. For example, the membership value at 5 is one. In the negation, the membership value at 5 would be zero (1-1=0). For example, if the membership value is 0.4. In the negation, the membership value would be 0.6 (1-0.4=0.6). Put the mouse over the image to see the negation of the fuzzy set (blue curve).
**Intersection**

In this figure, the red and green lines are fuzzy sets. To find the intersection of these sets take the minimum of the two membership values at each point on the x-axis (see the formal definition above). For example, in the figure the red fuzzy set has a membership of ZERO when \( x = 4 \) and the green fuzzy set has a membership of ONE when \( x = 4 \). The intersection would have a membership value of ZERO when \( x = 4 \) because the minimum of zero and one is zero.
Intersection
**Union**

In this figure, the red and green lines are fuzzy sets. To find the union of these sets take the maximum of the two membership values at each point on the x-axis (see the formal definition above). For example, in the figure the red fuzzy set has a membership of ZERO when $x = 4$ and the green fuzzy set has a membership of ONE when $x = 4$. The union would have a membership value of ONE when $x = 4$ because the maximum of zero and one is one.
Union
The Concept of Hedging

Much has been made about the relationship of Fuzzy Logic to the human thought process and the ability to handle imprecise conditions that may arise. One of the terms frequently seen in the Fuzzy Logic literature is the concept of **Hedging**. Hedging can be described as the modifiers to a certain set, much like the way adjectives and adverbs modify statements in the English language.

When referring to a fuzzy set, hedges are used to adjust the characteristics of that fuzzy set by either:

- **Approximating**
- **Complementing**
- **Diluting**
- **Intensifying**
Some specific words and their effect on the fuzzy set include:

<table>
<thead>
<tr>
<th>Key Word</th>
<th>Effect on set characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>• about • near • close to • approximately</td>
<td>Approximate the set</td>
</tr>
<tr>
<td>• not</td>
<td>Complement the set</td>
</tr>
<tr>
<td>• somewhat • rather • quite</td>
<td>Dilute the set</td>
</tr>
<tr>
<td>• very • extremely</td>
<td>Intensify the set</td>
</tr>
</tbody>
</table>

In general, when a hedge is used to dilute a set, the set is expanded. When a set is intensified with a hedge, the set is compressed.
excellent!

You must be taller than this line to be considered TALL.
sharp-edged membership function for TALL

tall ($\mu = 1.0$)

not tall ($\mu = 0.0$)

definitely a tall person ($\mu = 0.95$)

really not very tall at all ($\mu = 0.30$)
1. **Fuzzify inputs**

- If *service is excellent* or *food is delicious* then *tip = generous*

  - \( \mu(\text{service} = \text{excellent}) = 0.0 \)  
  - \( \mu(\text{food} = \text{delicious}) = 0.7 \)

2. **Apply OR operator (max)**

- If \((0.0 \text{ or } 0.7)\) then *tip = generous*

  - \( \max(0.0, 0.7) = 0.7 \)

3. **Apply implication operator (min)**

- If \((0.7)\) then *tip = generous*

  - \( \min(0.7, \text{generous}) \)

*tip (fuzzy)*
Fuzzy Inference Systems
The General Case

Input → Output

Rules

Input terms (interpret) → Output terms (assign)

A Specific Example

service → tip

if service is poor then tip is cheap
if service is good then tip is average
if service is excellent then tip is generous

service is interpreted as {poor, good, excellent}
tip is assigned to be {cheap, average, generous}
Overview of Fuzzy Inference Process

Dinner for Two
a 2 input, 1 output, 3 rule system

Input 1
Service (0-10)

Rule 1
If service is poor or food is rancid, then tip is cheap.

Rule 2
If service is good, then tip is average.

Rule 3
If service is excellent or food is delicious, then tip is generous.

Input 2
Food (0-10)

Output
Tip (5-25%)

The inputs are crisp (non-fuzzy) numbers limited to a specific range.
All rules are evaluated in parallel using fuzzy reasoning.
The results of the rules are combined and distilled (defuzzified).
The result is a crisp (non-fuzzy) number.
Step 1. Fuzzify Inputs

1. Fuzzify inputs.

food is delicious

food = 8

input

Result of fuzzification

0.7
Step 2. Apply Fuzzy Operator

1. Fuzzify inputs.

   - excellent
   - delicious

2. Apply OR operator (max).

   - service is excellent
   - food is delicious

   service = 3 (input 1)
   food = 8 (input 2)

   result of fuzzy operator = 0.7
Step 3. Apply Implication Method

Antecedent

1. Fuzzify inputs.

Consequent

2. Apply OR operator (max).

3. Apply Implication operator (min).

result of implication

If service is excellent or food is delicious, then tip = generous

service = 3  food = 8

input 1  input 2
Step 4. Aggregate All Outputs

1. Fuzzify inputs.
   - service is poor
   - food is rancid
   - tip = cheap

2. Apply fuzzy operation (OR = max).
   - rule 2 has no dependency on input 2
   - service is good
   - tip = average

3. Apply implication method (min).
   - service is excellent
   - food is delicious
   - tip = generous

4. Apply aggregation method (max).
   - service = 3
   - food = 8
   - Result of aggregation
Step 5. Defuzzify

5. Defuzzify the aggregate output (centroid).

Result of defuzzification

\[ \text{tip} = 16.7\% \]
The Fuzzy Inference Diagram

Interpreting the fuzzy inference diagram

1. if and then

2. if and then

input 1

input 2

output
1. Fuzzify inputs.

1. If service is poor or food is rancid then tip = cheap

2. Apply fuzzy operation (OR = max).

2. Apply implication method (min).

3. If service is good then tip = average

3. If service is excellent or food is delicious then tip = generous

4. Apply aggregation method (max).

5. Defuzzify (centroid).

service = 3

food = 8

input 1

input 2

output

tip = 16.7%
• Hard Classification - a pixel can only have one and only one category.
• In urban regions, a pixel in reality may have more than one category because of the heterogeneity of the land cover composing that pixel. We call this a mixed pixel.
• Soft or fuzzy classifiers - a pixel does not belong fully to one class but it has different degrees of membership in several classes. The mixed pixel problem is more pronounced in lower resolution data. In fuzzy classification, or pixel unmixing, the proportion of the land cover classes from a mixed pixel is calculated.
• Land cover is usually only poorly represented by discrete classes for two main reasons, (1) pixels are often mixed – representing areas on the ground which have multiple land cover types and (2) land cover classes often intergrade due to the continuous nature of vegetative land cover (Foody,  

• For example, the Oak Openings Region of NW Ohio and SE Michigan is an extreme example of the continuous nature of land cover. By its very definition, savanna is a class that falls within the continuum between forest and savanna. It is a plant community consisting of scattered open-grown trees with a primarily herbaceous understory. “No definition that clearly separates savanna from prairie and forest has been developed (Nuzzo, 1994).” This makes the use of a fuzzy or soft classification appropriate.
• Zadeh defined fuzzy logic as, “a set of mathematical principles for knowledge representation based on degrees of membership rather than on crisp memberships of classical binary logic (Zadeh, 1965).”

• Researchers within the remote sensing field have adapted and/or applied fuzzy set theory to the classification of satellite imagery.

• Fuzzy set theory has been applied to classification to produce three main fuzzy classification methods including, (1) specific fuzzy classifiers (ex. fuzzy c-means & fuzzy k-means), (2) ‘softening’ the results of a crisp classification (ex. Maximum Likelihood Classifier), and (3) the use of artificial neural networks (Woodcock et al., 2000; Zhang et al., 2001).
**Hard Versus Soft Classifiers**

- Traditional classifiers can be called *hard* classifiers since they yield a hard decision about the identity of each pixel. In contrast, *soft* classifiers express the degree to which a pixel belongs to each of the classes being considered. Thus, for example, rather than deciding that a pixel is either deciduous or coniferous forest, it might indicate that its membership grade in the deciduous class is 0.43 and coniferous is 0.57 (which a hard classifier would conclude is coniferous). One of the motivations for using a soft classifier is to determine the mixture of landcover classes present. If we could assume that these two classes were the only ones present, it might be reasonable to conclude that the pixel contains 43% deciduous cover and 57% coniferous. Such a conclusion is known as *sub-pixel classification*.

- A second motivation for the use of a soft classifier is to measure and report the strength of evidence in support of the best conclusion that can be made. IDRISI introduces special soft classifiers that allow us to determine, for example, that evidence for deciduous is present to a level of 0.26, for coniferous to 0.19 and some unknown type to 0.55. This would immediately suggest that while the pixel has some similarities to our training sites for these two classes, it really belongs to some type that we have not yet identified.
The FCM clustering method is a modification of the common hard c-means clustering approach. Instead of assigning each pixel to only one cluster, fuzzy membership values for each cluster are assigned to each image pixel. FCM makes no assumption of data distribution and fuzzy membership values are not based on a probability density function (Key et al., 1989). Fuzzy membership values are calculated based on the weighted distance, i.e., Euclidean distance, between an input sample and cluster center in the feature space (equation [3], [4]). The cluster centers and fuzzy memberships are updated iteratively from the initial random cluster center. The original purpose of FCM was data reduction for pattern recognition (Bezdek, 1981). In image analysis, FCM is widely used as an unsupervised classification approach to segment remotely sensed image (Trivedi and Bezdek, 1986; Cannon et al, 1986). FCM has also been applied in a supervised classification approach (Key et al, 1989; Zhang and Foody, 1998).

The FCM is derived by:

$$M = \{U : u_{ik} \in [0,1]: \sum_{k=1}^{n} u_{ik} > 0, I = 1...c; \sum_{i=1}^{c} u_{ik} = 1, k = 1...n\}$$

(1)

where \(U\) is a fuzzy c-partition of a sample of \(n\) observations and \(c\) clusters; \(u_{ik}\) is an element of \(U\) and represents the membership of an observation, \(x_k\), to the \(i\)th fuzzy group (Bezdek 1981). Each \(x_k\) is a vector of length \(p\) where \(p\) is the number of attributes used. This algorithm is based on an iterative minimization of the criterion function (Bezdek 1981; Zimmermann 1991; Bezdek and Pal 1992):

$$J(U, V) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m |X_k - V_i|^2$$

(2)

where \(X_k\) is data sample vectors; \(V_i\) is cluster centers; \(m\) is an exponent weight factor. The iterative process updates the membership value of the input data, \(u_{ik}\), and the cluster center, \(V_i\), until the criteria, \(\max |u_{ik}^0 - u_{ik}^{(i)}|\), is met, where \(\delta\) is the pre-specified threshold value. The update is based on these equations:

$$V_i = \frac{1}{\sum_{k=1}^{n} u_{ik}^m} \sum_{k=1}^{n} u_{ik}^m x_{ik} \quad i = 1,2, \ldots, c.$$  

(3)
\[ u_{ik} = \frac{\left[ \frac{1}{\left| x_k - v_i \right|^2} \right]^{1/(m-1)}}{\sum_{j=1}^{c} \left[ \frac{1}{\left| x_k - v_j \right|^2} \right]^{1/(m-1)}} \quad i = 1, 2, \ldots, c; \quad k = 1, 2, \ldots, n. \] (4)

The FCM-based sub-pixel analysis is an one-step calculation algorithm where a single pixel value of each land cover class (the pure spectra) is given as the cluster center of the FCM (Key et al. 1989; Foody and Cox 1994; and Bastin 1997). As the cluster center values are given as the pure pixel of each target land cover class, the iterative process of FCM to find cluster center, equation [3], is not necessary. With the pure spectral value of each land cover class as the cluster center, the fuzzy membership values of each pixel to each land cover class are estimated based on equation...
Idrisi Soft Classifiers
Soft Classifiers

Unlike hard classifiers, soft classifiers defer making a definitive judgment about the class membership of any pixel in favor of a group of statements about the degree of membership of that pixel in each of the possible classes. Like traditional supervised classification procedures, each uses training site information for the purpose of classifying each image pixel. However, unlike traditional hard classifiers, the output is not a single classified landcover map, but rather a set of images (one per class) that express for each pixel the degree of membership in the class in question. In fact, each expresses the degree to which each pixel belongs to the set identified by a signature according to one of the following set membership metrics:

- BAYCLASS based on Bayesian probability theory,
- BELCLASS based on Dempster-Shafer theory,
- MAHALCLASS based on Mahalanobis distance,
- FUZCLASS based on Fuzzy Set theory, and
- UNMIX based on the Linear Mixture model.

It is important to recognize that each of these falls into a general category of what are known as *Fuzzy Measures* (Dubois and Prade, 1982) of which Fuzzy Sets is only one instance. Fuzziness can arise for many reasons and not just because a set is itself fuzzy. For example, measurement error can lead to uncertainty about the class membership of a pixel even when the classes (sets) are crisply defined. It is for this reason that we have adopted the term soft—it simply recognizes that the class membership of a pixel is frequently uncertain for reasons that are varied in origin.
Classification Uncertainty

In addition to these set membership images, each of these soft classifiers outputs an image that expresses the degree of classification uncertainty it has about the class membership of any pixel. Classification uncertainty measures the degree to which no class clearly stands out above the others in the assessment of class membership of a pixel. In the case of BAYCLASS, BELCLASS and FUZCLASS, it is calculated as follows:

\[
\text{ClassificationUncertainty} = 1 - \frac{\max - \sum}{1 - \frac{1}{n}}
\]

where

\[
\begin{align*}
\max &= \text{the maximum set membership value for that pixel} \\
\sum &= \text{the sum of the set membership values for that pixel} \\
n &= \text{the number of classes (signatures) considered}
\end{align*}
\]

The logic of this measure is as follows:

- The numerator of the second term expresses the difference between the maximum set membership value and the total dispersion of the set membership values over all classes.

- The denominator of the second term expresses the extreme case of the difference between a maximum set membership value of 1 (and thus total commitment to a single class) and the total dispersion of that commitment over all classes.

- By taking the ratio of these two quantities, one develops a measure that expresses the degree of commitment to a specific class relative to the largest possible commitment that can be made. Classification uncertainty is thus the complement of this ratio.

In spirit, the measure of uncertainty developed here is similar to the entropy measure used in Information Theory. However, it differs in that it is concerned not only with the degree of dispersion of set membership values between classes, but also the total amount of commitment present. Following are some examples that can clarify this concept.
Examples

Assuming a case where three classes are being evaluated, consider those with the following allocations of set membership:

(0.0 0.0 0.0) Classification Uncertainty = 1.00
(0.0 0.0 0.1) Classification Uncertainty = 0.90
(0.3 0.3 0.3) Classification Uncertainty = 1.00
(0.6 0.3 0.0) Classification Uncertainty = 0.55
(0.6 0.3 0.1) Classification Uncertainty = 0.60
(0.9 0.1 0.0) Classification Uncertainty = 0.15
(0.9 0.05 0.05) Classification Uncertainty = 0.15
(1.0 0.0 0.0) Classification Uncertainty = 0.00

With UNMIX, however, classification uncertainty is measured as the residual error after calculation of the fractions of constituent members. This will be discussed further below.
BAYCLASS and Bayesian Probability Theory

BAYCLASS is a direct extension of the MAXLIKE module. It outputs a separate image to express the posterior probability of belonging to each considered class according to Bayes' Theorem:

\[ p(h|e) = \frac{p(e|h) \cdot p(h)}{\sum_i p(e|h_i) \cdot p(h_i)} \]

where:
- \( p(h|e) \) = the probability of the hypothesis being true given the evidence (posterior probability)
- \( p(e|h) \) = the probability of finding that evidence given the hypothesis being true
- \( p(h) \) = the probability of the hypothesis being true regardless of the evidence (prior probability)

In this context, the variance/covariance matrix derived from training site data is that which allows one to assess the multivariate conditional probability \( p(e|h) \). This quantity is then modified by the prior probability of the hypothesis being true and then normalized by the sum of such considerations over all classes. This latter step is important in that it makes the assumption that the classes considered are the only classes that are possible as interpretations for the pixel under consideration. Thus even weak support for a specific interpretation may appear to be strong if it is the strongest of the possible choices given.

This posterior probability \( p(h|e) \) is the same quantity that MAXLIKE evaluates to determine the most likely class, and indeed, if the output images of BAYCLASS were to be submitted directly to HARDEN, the result would be identical to that of MAXLIKE. In essence, BAYCLASS is a confident classifier. It assumes that the only possible interpretation of a pixel is one of those classes for which training site data have been provided. It therefore admits to no ignorance. As a result, lack of evidence for an alternative hypothesis constitutes support for the hypotheses that remain. In this context, a pixel for which reflectance data only very weakly support a particular class is treated as unequivocally belonging to that class (\( p = 1.0 \)) if no support exists for any other interpretation.

The prime motivation for the use of BAYCLASS is sub-pixel classification—i.e., to determine the extent to which mixed pixels exist in the image and their relative proportions. It is also of interest to observe the underlying basis of the MAXLIKE procedure. However, for In-Process Classification Assessment (IPCA), the BELCLASS procedure is generally preferred because of its explicit recognition that some degree of ignorance may surround the classification process.
In the context of mixture analysis, the probabilities of BAYCLASS are interpreted directly as statements of proportional representation. Thus if a pixel has posterior probabilities of belonging to deciduous and conifer of 0.68 and 0.32 respectively, this would be interpreted as evidence that the pixel contains 68% deciduous species and 32% conifers. Note, however, that this requires several important assumptions to be true. First, it requires that the classes for which training site data have been provided are exhaustive (i.e., that there are no other possible interpretations for that pixel). Second, it assumes that the conditional probability distributions $p(e|h)$ do not overlap in the case of pure pixels. In practice, these conditions may be difficult to meet.

In testing at Clark Labs, we have found that while BAYCLASS is effective in determining the constituent members of mixed pixels, it is often not so effective in determining the correct proportions. Rather, we have found that procedures based on the Linear Mixture model (UNMIX) perform considerably better in this respect. However, Linear Spectral Unmixing has its own special limitations. Thus, we favor a hybrid approach using the better qualities of Bayesian decomposition and Linear Spectral Unmixing, as will be discussed below.
BELCLASS and Dempster-Shafer Theory

- In normal use, Dempster-Shafer theory requires that the hypotheses (classes) under consideration be mutually exclusive and exhaustive. However, in developing BELCLASS, we felt that there was a strong case to be made for non-exhaustive categories—that the pixel may indeed belong to some unknown class, for which a training site has not been provided. In order to do this we need to add an additional category to every analysis called [other], and assign any incompleteness in one’s knowledge to the indistinguishable set of all possible classes (including this added one)—e.g., [conifer deciduous grass urban water other]. This yields a result which is consistent with Dempster-Shafer theory, but which recognizes the possibility that there may be classes present about which we have no knowledge.

  Defining ignorance as a commitment to the indistinguishable set of all classes suggests that it may have some relationship to the classification uncertainty image produced by BAYCLASS. The classification uncertainty image of BAYCLASS expresses the uncertainty the classifier has in assigning class membership to a pixel. Uncertainty is highest whenever there is no class that clearly stands out above the others in the assessment of class membership for a pixel. We have found that in the context of BELCLASS, this measure of uncertainty is almost identical to that of Dempster-Shafer ignorance. As a result, we have modified the output of the classification uncertainty image of BELCLASS slightly so that it outputs true Dempster-Shafer ignorance.\(^7\)

- The operation of BELCLASS is essentially identical to that of BAYCLASS and thus also MAXLIKE. Two choices of output are given: beliefs or plausibilities. In either case, a separate image of belief or plausibility is produced for each class. In addition, a classification uncertainty image is produced which can be interpreted in the same manner as the classification uncertainty image produced by all of the soft classifiers (but which is truly a measure of Dempster-Shafer ignorance).

  The prime motivation for the use of BELCLASS is to check for the quality of one’s training site data and the possible presence of unknown classes during In-Process Classification Assessment. In cases where one believes that one or more unknown classes exist (and thus that some portion of total ignorance arises because of this presence of an unknown class), the BELCLASS routine should be used. BELCLASS does this by implicitly adding an [other] class to the set of classes being considered. This is a theoretical concession to the mechanics of the BELCLASS process and will not be directly encountered by the user.

  Comparing the output of BELCLASS with BAYCLASS, you will notice a major difference. Looking at the images produced by BAYCLASS, it will appear as if your training sites are strong. BAYCLASS is a very confident classifier (perhaps overly confident) since it assumes no ignorance. BELCLASS, however, appears to be a very reserved classifier. Here we see a result in which all of the uncertainties in our information become apparent. It does not presume to have full information, but explicitly recognizes the possibility that one or more unknown classes may exist.
FUZCLASS and Fuzzy Set Theory

- The third soft classifier in IDRISI is FUZCLASS. As the name suggests, this classifier is based on the underlying logic of Fuzzy Sets. Just as BAYCLASS and BELCLASS are based on the fundamental logic of MAXLIKE, FUZCLASS is based on the underlying logic of MINDIST—i.e., fuzzy set membership is determined from the distance of pixels from signature means as determined by MAKESIG.

- There are two important parameters that need to be set when using FUZCLASS. The first is the z-score distance where fuzzy membership becomes zero. The logic of this is as follows.

- It is assumed that any pixel at the same location in band space as the class mean (as determined by running MAKESIG) has a membership grade of 1.0. Then as we move away from this position, the fuzzy set membership grade progressively decreases until it eventually reaches zero at the distance specified. This distance is specified as a standard score (z-score) to facilitate its interpretation. Thus, specifying a distance of 1.96 would force 5% of the data cells to have a fuzzy membership of 0, while 2.58 would force 1% to have a value of 0.

- The second required parameter setting is whether or not the membership values should be normalized. Normalization makes the assumption (like BAYCLASS) that the classes are exhaustive, and thus that the membership values for all classes for a single pixel must sum to 1.0. This is strictly required to generate true fuzzy set membership grades. However, as a counterpart to BELCLASS, the option is provided for the calculation of un-normalized values. As was suggested earlier, this is particularly important in the context of In-Process Classification Assessment for evaluation of a MINDIST-based classification.
UNMIX and the Linear Mixture Model

The Linear Mixture Model assumes that the mixture of materials within a pixel will lead to an aggregate signature that is an area-weighted average of the signatures of the constituent classes. Thus if two parent materials (called end members in the language of Linear Spectral Unmixing) had signatures of 24, 132, 86 and 56, 144, 98 on three bands, a 50/50 mixture of the two should yield a signature of 40, 138, 92. Using this simple model, it is possible to estimate the proportions of end member constituents within each pixel by solving a set of simultaneous equations. For example, if we were to encounter a pixel with the signature 32, 135, 89, and assumed that the pixel contained a mixture of the two end members mentioned we could set up the following set of equations to solve:

\[
\begin{align*}
   f_1(24) + f_2(56) &= 32 \\
   f_1(132) + f_2(144) &= 135 \\
   f_1(86) + f_2(98) &= 89 
\end{align*}
\]

where \( f_1 \) and \( f_2 \) represent the fractions (proportions) of the two end members. Such a system of simultaneous equations can be solved using matrix algebra to yield a best fit estimate of \( f_1 \) and \( f_2 \) (0.75 and 0.25 in this example). In addition, the sum of squared residuals between the fitted signature values and the actual values can be used as a measure of the uncertainty in the fit.

The primary limitation of this approach is that the number of end members cannot exceed the number of bands. This can be a severe limitation in the case of SPOT imagery, but of little consequence with hyperspectral imagery. IDRISI thus offers three approaches (in UNMIX) to Linear Spectral Unmixing:

1. The standard linear spectral unmixing approach (as indicated above) for cases where sufficient bands are available.

2. A probability guided option for cases where insufficient bands exist. Although the total number of possible end members may be large, the number that coexist within a single pixel is typically small (e.g., 2-3). This approach thus uses a first stage based on the BAYCLASS module to determine the most likely constituents (up to the number of bands), with a second stage linear spectral unmixing to determine their fractions. Experiments at Clark Labs have shown this to produce excellent results.

3. An exhaustive search option for cases where insufficient bands exist. In this instance, one specifies the number of constituents to consider (up to the total number of bands). It then tests all possible combinations of that many end members and reports the fractions of that combination with the lowest sum of squared residuals. This approach is considerably slower than the other options. In addition, experiments at Clark Labs have shown that this approach yields inferior results to the probability guided procedure in cases where the end member signatures are drawn from training sites. Pure end member signatures, created with ENDSIG, should be used.
**Hardeners**

Once a soft classifier has been applied to a multispectral image set, the soft results can be re-evaluated to produce a hard classification by using one of the following hardeners from the module HARDEN:

**BAYCLASS**

Using the results from BAYCLASS, this option determines the class possessing the maximum posterior probability for each cell, given a set of probability images. Up to four levels of abstraction can be produced. The first is the most likely class, just described. The second outputs the class of the second highest posterior probability, and so on, up to the fourth highest probability.

**BELCLASS**

Using the results from BELCLASS, this option is essentially identical to the BAYCLASS hardener, except that it is designed for use with Dempster-Shafer beliefs.

**FUZCLASS**

Using the results from FUZCLASS, this option is essentially identical to the BAYCLASS hardener, except that it is designed for use with Fuzzy Sets.
• Foody states about accuracy: The methods generally used, such as the percentage correct allocation or kappa coefficient of agreement, are not without problems. It is, for instance, typically assumed that the classes are unordered (van Deusen, 1996). In many instances, however, the classes may be ordered yet the conventional accuracy assessment techniques are used (e.g., Joria and Jorgenson, 1996). In such circumstances, the use of a conventional accuracy assessment measure, designed for application to nominal level data, will not make full use of the information content of the data and may not provide an appropriate index of classification quality.

• It may sometimes be preferable to use approaches that can accommodate the ordinal nature of the data set such as the **weighted kappa coefficient** (Cohen, 1968; Foody et al., 1996). The analyst may at times be constrained to use a conventional measure of accuracy, for example, if evaluating the results against previously published work to maintain consistency, but the possibility that the measure does not fully use the data should be recognized and its implications considered.
Gopal and Woodcock outline three primary limitations to the traditional methods of accuracy assessment:

1. It is assumed that each area in the map can be unambiguously assigned to a single map category. In assessing ground truth for map accuracy, the expert has to resolve this issue by selecting a single category for each ground location and matching this against the map value.

2. Information on the magnitude of errors is limited to noting the pattern of mismatches between categories in the map. Data concerning the magnitude or seriousness of these mismatches as indicated by the conditions of the ground site cannot be used.

3. Third, the user needs to be provided with more complete and interpretable information about the map than is currently practiced. Detailed information on errors will help the users to check if the map can be used for a particular purpose (Gopal et al., 1994).
Gopal and Woodcock suggest an alternative, fuzzy, approach that takes their points into consideration. This approach is called an “expert evaluation” and it based on a linguistic scale:

1. *Absolutely Wrong*: This answer is absolutely unacceptable. Very Wrong.
2. *Understandable but Wrong*: Not a good answer. There is something about the site that makes the answer understandable but there is clearly a better answer. This answer would pose a problem for users of the map. Not Right.
3. *Reasonable or Acceptable Answer*: Maybe not the best possible answer but it is acceptable; this answer does not pose a problem to the user if it is seen on the map. Right.
4. *Good Answer*: Would be happy to find this answer given on the map. Very Right.
5. *Absolutely Right*: No doubt about the match. Perfect (Gopal, 1994).

With this linguistic scale, the expert is asked to evaluate each class for each validation point assigning a number to each. Once these values are evaluated as compared to the mapped data, information about magnitude of errors, source of errors, and nature of errors can be identified. This approach has been used in a number of studies including, Woodcock et al., 2000; McMahan et al., n/a; Laba et al., 2002.
Binaghi et al. propose another method, based on fuzzy set theory, for classification accuracy assessment that is designed specifically for “soft” classifications (Binaghi et al., 1999). Their approach is the generalization of the traditional error matrix. This fuzzy error matrix is built based on the computation of the degree of membership in the fuzzy intersection set. “The fuzzy error matrix can be used as the starting point for descriptive techniques, in the same way as the conventional error matrix (Binaghi et al., 1999).”
• Overall accuracy is the most common measure of evaluating the quality of classifications, although it does not take location into account. The kappa coefficient factors in the effect of chance in the classification.

• For example, a kappa value of 78% indicates that the classification is 78% better than a classification that resulted from random assignment. Therefore, kappa is lower than the overall accuracy.

• The weighted equivalents of overall accuracy and kappa can be used with fuzzy classification for accuracy.

• The arrangement shown in the following Table is used to assess the weighted overall accuracy and weighted kappa coefficient.

• This is similar to an error matrix except that the cell values are not absolute observations and computed values but proportions.
Table 10.3: Proportion of pixels distributed into \( k \) classes

<table>
<thead>
<tr>
<th>Classification</th>
<th>Reference</th>
<th>1</th>
<th>2</th>
<th>.....</th>
<th>( k - 1 )</th>
<th>( k )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p_{11} )</td>
<td>( p_{12} )</td>
<td>.....</td>
<td>( p_{1,k-1} )</td>
<td>( p_{1k} )</td>
<td>( p_{1+} )</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( p_{21} )</td>
<td>( p_{22} )</td>
<td>.....</td>
<td>( p_{2,k-1} )</td>
<td>( p_{2k} )</td>
<td>( p_{2+} )</td>
<td></td>
</tr>
<tr>
<td>( k - 1 )</td>
<td>( p_{k-1,1} )</td>
<td>( p_{k-1,2} )</td>
<td>.....</td>
<td>( p_{k-1,k-1} )</td>
<td>( p_{k-1,k} )</td>
<td>( p_{k-1+} )</td>
<td></td>
</tr>
<tr>
<td>( k )</td>
<td>( p_{k1} )</td>
<td>( p_{k2} )</td>
<td>.....</td>
<td>( p_{k,k-1} )</td>
<td>( p_{kk} )</td>
<td>( p_{k+} )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( p_{+1} )</td>
<td>( p_{+2} )</td>
<td>.....</td>
<td>( p_{+k-1} )</td>
<td>( p_{+k} )</td>
<td>( 1 )</td>
<td></td>
</tr>
</tbody>
</table>

(from Naesset, 1996)

To make this table, we divide all the original values in the error matrix by the total number of test points. Let

\[
p_{i+} = \sum_{j=1}^{k} p_{ij}
\]  

(10.1)

be the proportion of pixels classified into class \( i \) in the classified image, and

\[
p_{+j} = \sum_{i=1}^{k} p_{ij}
\]  

(10.2)

be the proportion of pixels confirmed as class \( j \) in the reference image. Let \( w_{ij} \) be the weight associated with the \( i,j \)th cell in the error matrix. If

\[
p_{\sigma} = \sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} p_{ij}
\]  

(10.3)

is the weighted agreement or weighted overall accuracy, and

\[
p_{c} = \sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} p_{i+} p_{+j}
\]  

(10.4)

is the weighted chance agreement, Cohen (1968) defines weighted kappa as

\[
K_w = \frac{p_{\sigma} - p_{c}}{1 - p_{c}}
\]  

(10.5)
To calculate for the weighted overall accuracy and weighted kappa, we need to assign a weight for each cell in the error matrix to reflect the severity of the misclassification error. Fleiss et al. (1969) state that weights are limited to the interval $0 \leq w_{ij} \leq 1$ for $i \neq j$, and that the weight for perfect agreement is 1 (i.e., $w_{ii} = 1$). Naesset (1996) suggested that weights may reflect the loss of utility because of misclassification. If $U_{c,j}$ is the utility when a pixel is correctly classified into class $j$ and $U_{E,ij}$ is the utility when a pixel belonging to class $j$ is wrongly classified into class $i$, then the weight is

$$w_{ij} = \frac{U_{E,ij}}{U_{c,j}}$$

(10.6)
FUZZY CLASSIFICATION OF MEDITERRANEAN TYPE FOREST USING ENVISAT MERIS DATA

S. Berberoğlu, O. Satir,
Department of Landscape Architecture, Çukurova University 01330 Adana, Turkey - (suha, osatir)@cu.edu.tr

KEY WORDS: Fuzzy classification, Linear mixture modeling, Artificial neural networks, Envisat MERIS.

ABSTRACT

The aim of this study was to classify Envisat MERIS and Landsat ETM satellite sensor imagery using fuzzy classification techniques such as, linear mixture modelling and artificial neural networks. The images were classified successfully using these two techniques. The fuzzy results were more accurate than hard classification. Landsat ETM imagery was classified using maximum likelihood classifier and the output was resampled to 300 m to produce test data. Land cover classes comprised agriculture, bare ground, Turkish pine (*Pinus brutia*), Crimean pine (*Pinus nigra*), Lebanese cedar (*Cedrus libani*), Taurus fir (*Abies sp.*), Juniper (*Juniperus sp.*) and water. The classification accuracy was poor as there was insufficient number of training pixels available for these classes. As a result of this, overall accuracy was considered to evaluate the potential of these techniques. Overall results of soft classification from linear mixture modeling and artificial neural network and hard classification were 80%, 78% and 57% respectively. It can be concluded that soft classifiers particularly ANN for classifying Mediterranean type forest cover has a great potential. Additionally, there is no significant difference between soft classification outputs for certain land cover classes from linear mixture modeling and artificial neural networks, however artificial neural networks tackled pixels with high degree of mixing more accurately than LMM.
FUZZY CLASSIFICATION OF HETEROGENEOUS VEGETATION IN A COMPLEX ARID ECOSYSTEM

Ben McMahan, Keith T. Weber, and Joel Sauder
GIS Training and Research Center
Idaho State University
Pocatello, ID 83209-8130
(http://giscenter.isu.edu, e-mail: giscenter@isu.edu)

ABSTRACT

Traditional methods of remote sensing classification may not accurately portray the complexity of ecosystems where vegetation type and structure is diverse and variable. We used fuzzy classification to better characterize the complexity and heterogeneity of sagebrush-steppe vegetation on the Upper Snake River Plain in southeast Idaho. Unlike supervised classification where pixels are classified into discrete categories, fuzzy systems classify each pixel into multiple categories based on estimated membership in each class. Field data (n = 370) collected in summer 2002 were used as training sites for supervised classification and fuzzy classification. We compared the results of supervised and fuzzy classification to determine which produced a more accurate depiction of land cover. Our results show fuzzy classification produces more accurate predictions of sagebrush and grass cover compared to supervised classification. Additionally, there is an apparent relationship between fuzzy membership and percent cover of sagebrush, which provides the user vegetation structure information as well.
IMAGE CLASSIFICATION BASED ON FUZZY LOGIC

I. Nedeljkovic

MapSoft Ltd, Zabjenska 26 11000 Belgrade, Serbia and Montenegro
igor.n@sezampro.yu

Commission VI, WG VI/1-3

KEY WORDS: fuzzy logic, classification, if-then rules, digital, imagery, remote sensing, land cover

ABSTRACT:

Fuzzy logic is relatively young theory. Major advantage of this theory is that it allows the natural description, in linguistic terms, of problems that should be solved rather than in terms of relationships between precise numerical values. This advantage, dealing with the complicated systems in simple way, is the main reason why fuzzy logic theory is widely applied in technique. It is also possible to classify the remotely sensed image (as well as any other digital imagery), in such a way that certain land cover classes are clearly represented in the resulting image. If that’s so, can we use fuzzy logic technique to diminish the influence of person dealing with supervised classification? Can we eliminate the prejudice? These questions were the light motive for this paper. In this paper, a priori knowledge about spectral information for certain land cover classes is used in order to classify SPOT image in fuzzy logic classification procedure. Basic idea was to perform the classification procedure first in the supervised and then in fuzzy logic manner. The later was done with Matlab’s Fuzzy Logic Toolbox. Some information, needed for membership function definition, was taken from supervised maximum likelihood classification. Also, the idea for result comparison came from PCI’s ImageWorks used for supervised procedure. Results of two procedures, both based on pixel-by-pixel technique, were compared and certain encouraging conclusion remarks come out.
2.1 Input data

The procedure of supervised image classification was conducted with PCI ImageWorks software. As the source for classification procedure, SPOT Image recorded in "XS" multispectral mode was used. This image contains three channels recorded in following bands:

- **band B1** covering 0.50 to 0.59 μm (green),
- **band B2** covering 0.61 to 0.68 μm (red) and
- **band B3** covering 0.79 to 0.89 μm (near infrared).

In order to use them further in different software (PCI ImageWorks, Matlab), SPOT image channels (named 701, 702, 703) are first converted from original SPOT format into tif, and then exported from tif into pix format in Geomatica Focus module (Figure 1.). The images were taken over the city of Cologne. The size of images is 3593x2990 pixels.

![Figure 1. SPOT image converted into three separated images](image-url)
As it was later used for fuzzy logic classification, the process of supervised image classification will be given in brief. Selected land cover classes are: deciduous trees, coniferous trees, urban area, water, crop1 and crop2. For these classes, training areas were pointed on the image (Figure 2.)

![Image](image.png)

**Figure 2.** Training areas shown in display window

Since the *signature separability* showed that deciduous trees and coniferous trees are very poorly separated (low values of *Transformed Divergence* and *Bhattacharyya Distance*; big overlap between the signatures of two classes).
A typical rule in a Sugeno fuzzy model has the following form: 
If Input 1 = x and Input 2 = y, then Output is z = ax + by + c 
For a zero-order Sugeno model, the output level z is a constant (a=b =0).

3.2.1 Membership function
Membership function is the mathematical function which defines the degree of an element's membership in a fuzzy set. The Fuzzy Logic Toolbox includes 11 built-in membership function types. These functions are built from several basic functions:
-piecewise linear functions,
-the Gaussian distribution function,
-the sigmoid curve and
-quadratic and cubic polynomial curve.

Two membership functions are built on the Gaussian distribution curve: a simple Gaussian curve and a two-sided composite of two different Gaussian curves.
Two membership functions are built on the *Gaussian* distribution curve: a simple Gaussian curve and a two-sided composite of two different Gaussian curves (Figure 3.)

![Gaussian membership functions](image)

**Figure 3.** Membership functions built on the Gaussian distribution curve
RESULTS:
at first sight, time necessary for fuzzy classification is longer comparing to maximum likelihood procedure, which takes several seconds to classify an image. But, if in ML procedure possible image transfer to recognizable format for certain software, formulation of the training areas, analysis concerning signature separability take place, than situation is quite different: *fuzzy logic takes advantage of already created simple rules and image classification (started from the scratch in both procedures) equal or even less time consuming. Of course, different conditions during image capture must be taken into account.*

-considering the level of classification accuracy, fuzzy logic can be satisfactory used for image classification.