Workshop: Regionalization of Forest Stand Variables
-TCP /IND/3505 -

Lab 03 Generate a training dataset

Paul Magdon

Forest Survey India, August,2017, Dehradun, India

Selection of Sample Sites

Selection of sample sites is a (statistical) sampling problem:
"The sampling design is a set of rules for selecting which pixel will be visited to obtain the reference area" Stehman (2004)

- **Requirements**:
- Reference data need to cover all classes
- Reference data should cover the spectral variability of the classes within the image
- *Parametric classifiers:* reference data should contain only 'pure pixels'
- *Non-Parametric classifiers:* reference data should also include mixed pixels

Common approaches for sample site selection:

1. Non-Probability / Arbitrarily selection
2. Random sampling
3. Systematic sampling
4. Stratified sampling

Extraction of spectral information

- Once the sample site locations are determined a sample plot needs to be established
- As with forest inventory, different aspects of plot design have to be controlled:

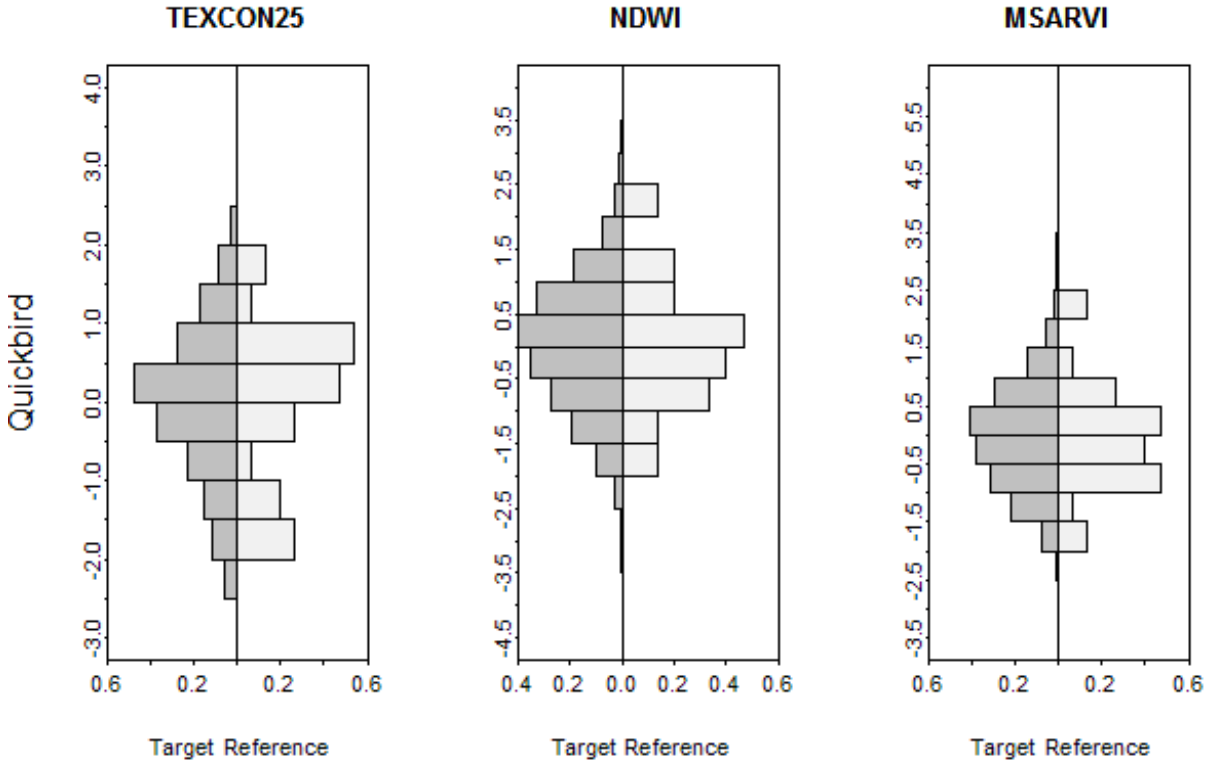
Shape	Size	Summary Statistic
• Point • Polygon • Rectangle • Circle	• Big enough to cover variability of spectral values therein • Small enough to avoid 'mixed pixels'	• Mean • Standard Deviation • Sum • Range

Evaluation of Reference Data

- Before the classification is started an intensive exploration and analysis of the reference data should be performed
  - Check the number of references per class
  - Check the distribution of spectral values for each class
  - Check the dissimilarity of the spectral distribution for each class and each feature
  - Delete noisy signatures (e.g. haze, sensor errors)
  - Merge classes / Split classes

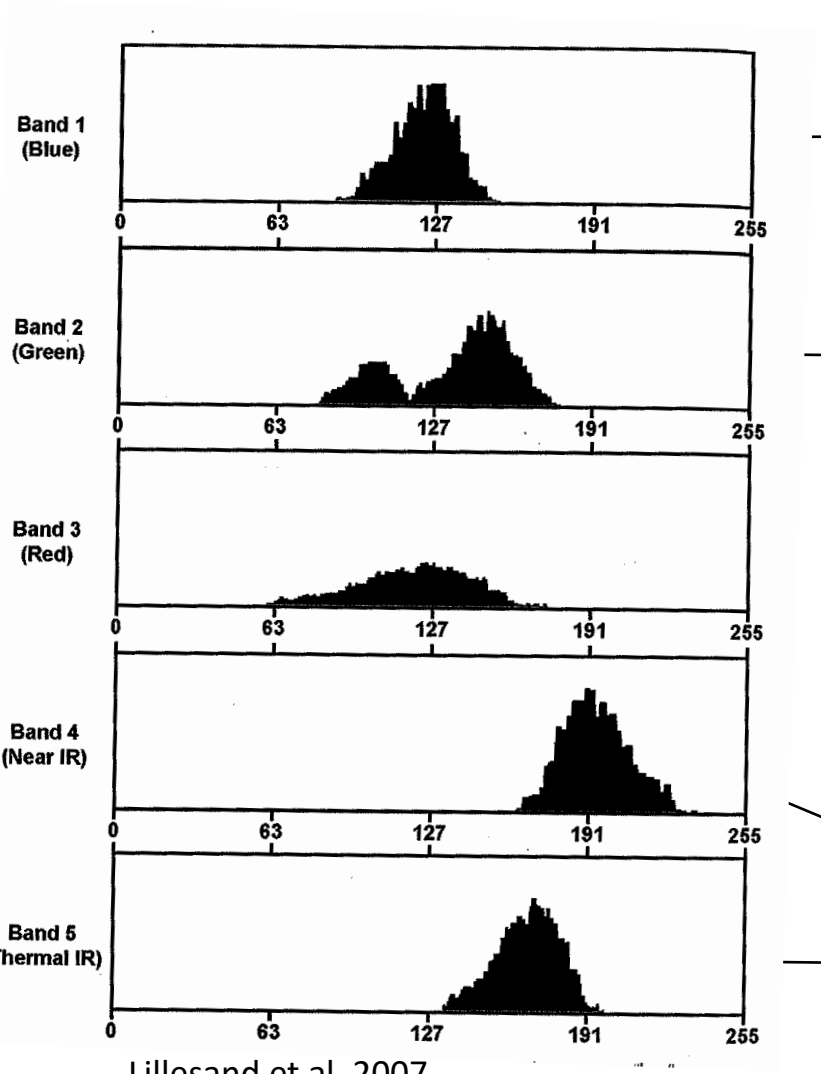
Graphical Evaluation: Feature Space

- As the classification is based on comparison of similarity, extrapolation beyond the range of the reference data will introduce bias



Fuchs et al. (2009)

Graphical Evaluation: Histogram



Lillesand et al. 2007

Normal Distribution

Bi-Modal Distribution

Skewed Distribution

Graphical Evaluation: Class Separability

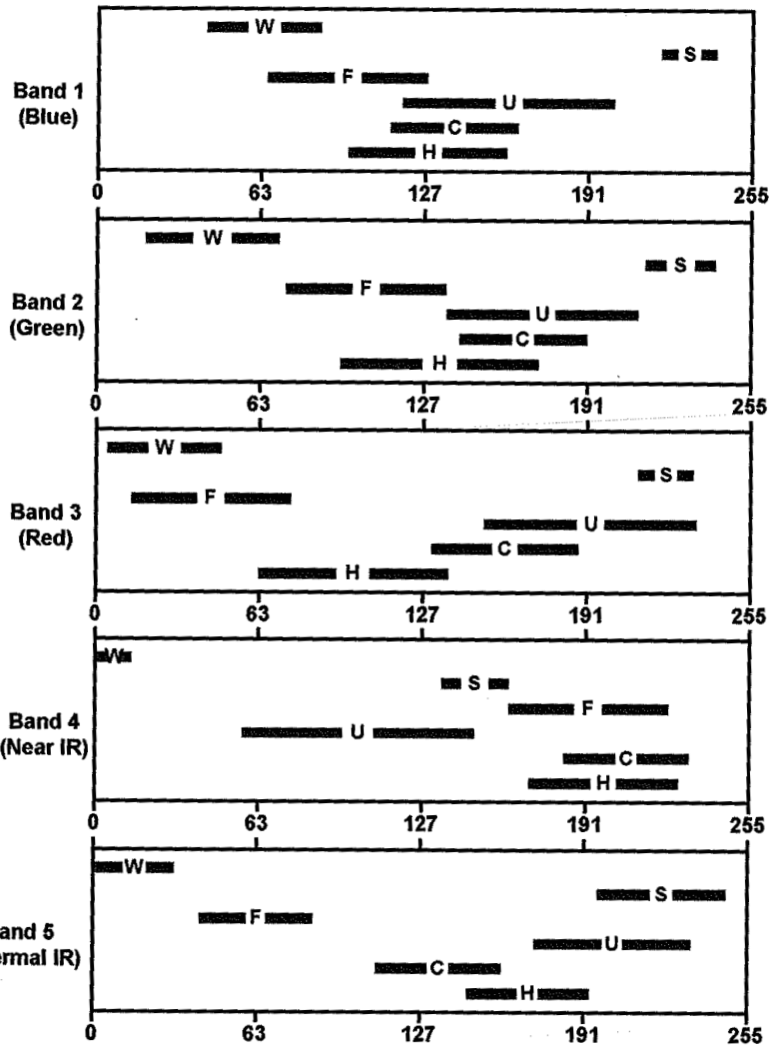
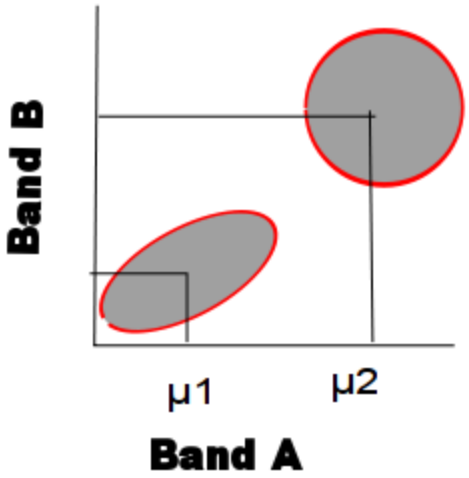
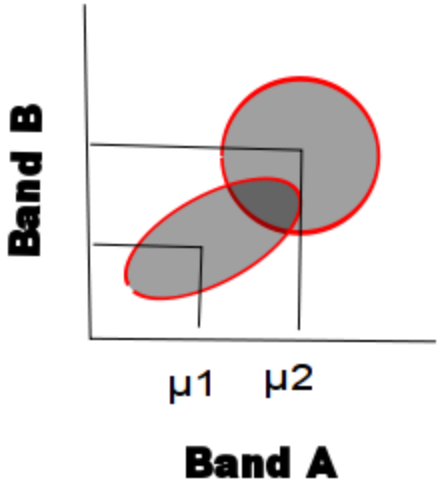


Figure 7.5 Lillesand et al. 2007

W=Water
H=Hay
U=Urban
C=Corn
F= Forest
S=Sand





Quantitative Evaluation: Class Separability

- Separability tests determine how similar the distributions of two groups of pixels are
- Measures looking at the distance between class means
  - Euclidean distance
  - Transformed Divergence
- Measure looking at both the differences between class means AND the distribution of the values around their means
  - Jeffries-Matusita Distance (JM-Distance)
  - Bhattacharyya Distance

References

- Fuchs H., Magdon P., Kleinn C., Flessa H. *2009 Estimating above ground carbon in a catchment of Siberian forest tundra: Combining satellite imagery and field inventory*, Remote Sensing of Environment 113(3)

- Lillesand T.M., Kiefer R.W., Chipman J.W. 2007 *Remote Sensing and Image Interpretation* 6th ed. , John Wiley & Sons

- Stemann S.V. 2004 *Sampling Design for Accuracy Assessment of Large-Area, Land-Cover Maps: Challenges and Future Directions* in *Remote Sensing an GIS Accuracy Assessment* Lunetta, R.S & Lyon, J.G (Editors) CRC Press

-