Airborne Imaging Spectroscopy Workshop 7 October 2005 – Bruges, Belgium

APPLICATION OF MACHINE LEARNING TECHNIQUES FOR ECOTOPE CLASSIFICATION BASED ON HYPERSPECTRAL IMAGES (ECOMALT)

Jonathan C-W Chan[1]
Desiré Planekx[2]
Toon Van Daele[2]
Luc de Bruyn[2]
Hichem Sahli[1]
Jo VanValckenborgh[3]

1. Electronics and Informatics Department, Vrije Universiteit Brussel (ETRO-VUB)
2. Institute of Natural Conservation (Instituut voor Natuurbehoud)
3. The Flemish Land Agency (Ondersteunend Centrum GIS-Vlaanderen, Vlaamse Landmaatschappij VLM)

Vrije Universiteit Brussel

ETRO
Department of Electronics and Informatics

Contents

The Biological Valuation Map

Characteristics

- Classification in ecological relevant categories (ecotopes)
- Uniform field-driven survey
- High detail (scale of 1/10.000)
- Biological valuation
- Covers the entire Flemish region

Use

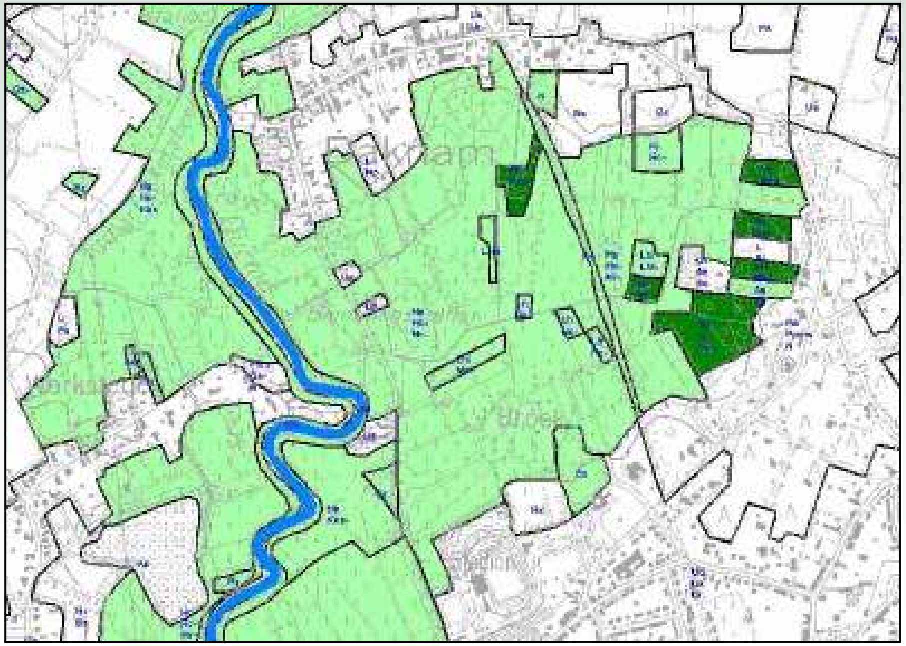
- Reference map (vulnerability assesment, nature development planning, …)
- Legal status

The Biological Valuation Map

History

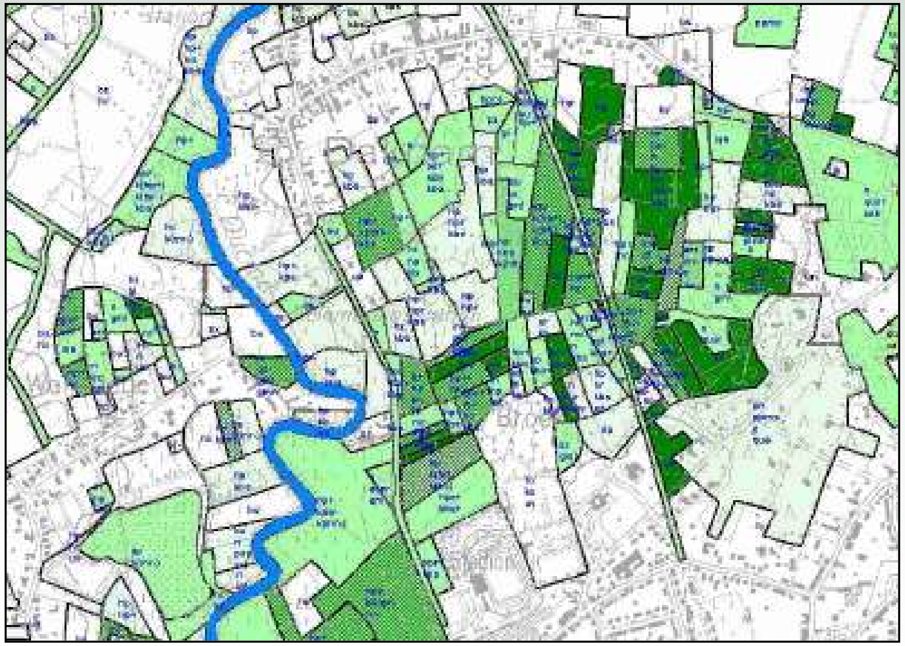
Version 1: **1978 – 1996**

global location of the biologically important landscapes



Version 2: **1997 – 2007**

more accurate & more detailed classes



The Biological Valuation Map

Shortcomings

- Time & labour intensive (17 full-time biologists – 10 years)
- changes within mapping period
- no monitoring of evolution in time

Use of remote sensing & automatic classification

- Snapshot of one particular moment
- regular updates & fast processing
- standardised & repeatable method

The Biological Valuation Map

Classification

- Full detail: 126 classes & variants → > 1000 combinations
- Level 3: 46 classes
- Level 2: 26 classes

In the framework of this study: ‘modified level 2’

- canopy ↔ undergrowth
- grassland types regrouped → 5 classes

The Biological Valuation Map

species poor
intensively used

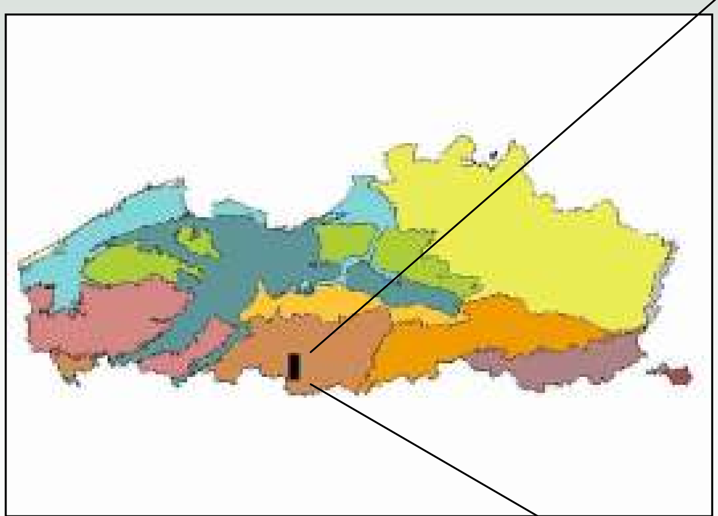
species rich
extensively used

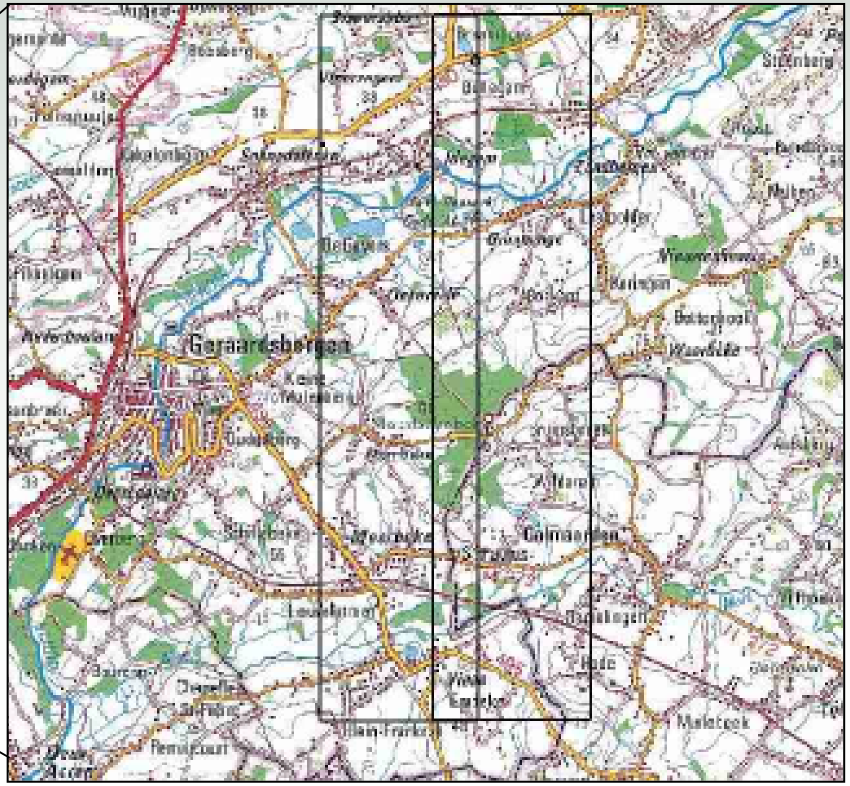
semi natural
grassland

The study area

Criteria

- Recent & detailed survey for BVM
- Fully processed & available in GIS
- Wide range of grasslands and forest types within a small area
- One field researcher





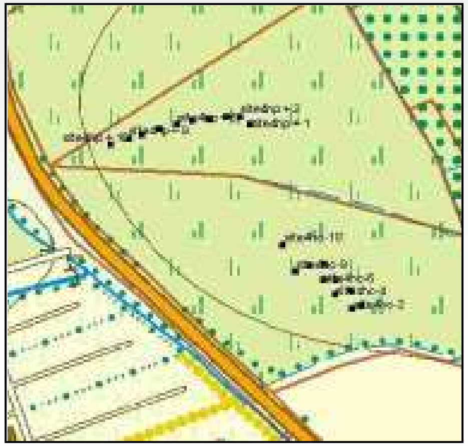
Ground measurement

Extra ground truth

- reference data
- preliminary investigation building a spectra library for BVM

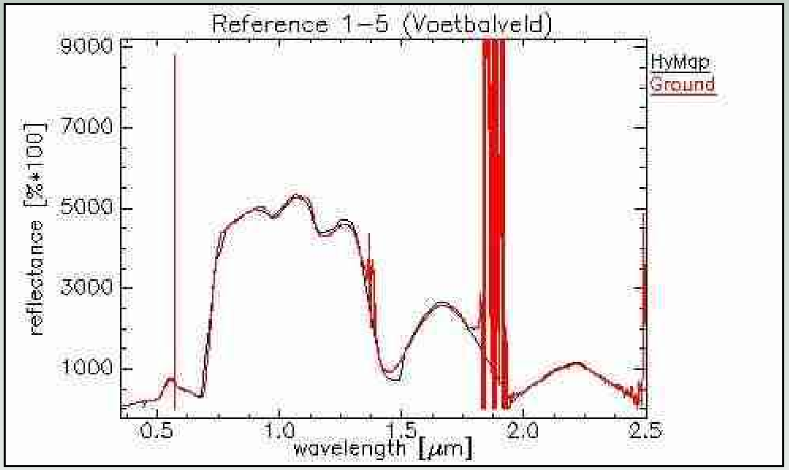
June 8 (overflight) & June 9, 2004

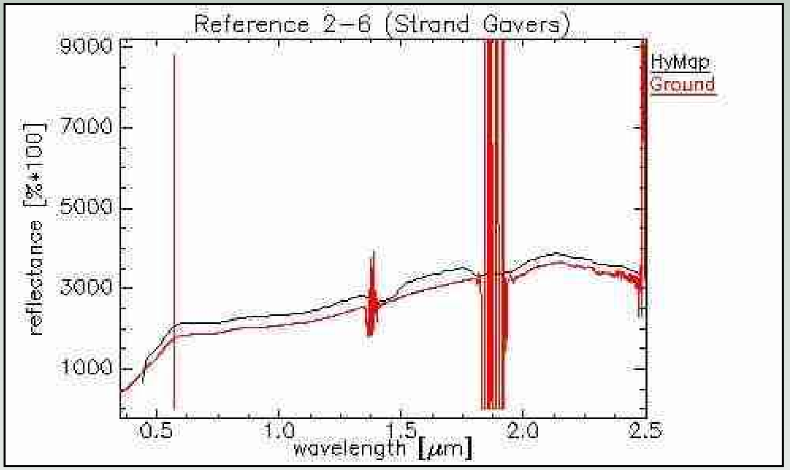
- Field spectroradiometer (VITO)
- 3 references sites - 11 grassland sites
- 10 locations / site



Ground measurement

Comparison of ground measurements and HyMap data





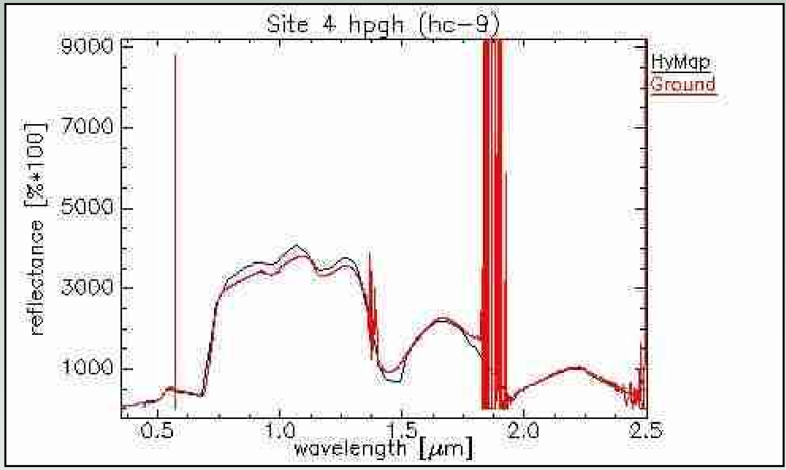
Reference 1 Soccer Field

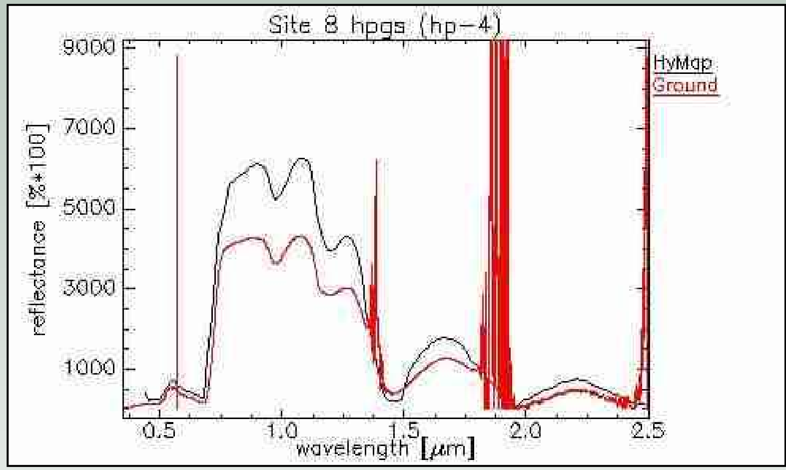
Reference 2 Sand Beach

Ground measurement

good matching

not matched





Site 4 hpgh mowed

Site 8 hpgs

Proposed Machine learning algorithms

- Decision Tree Classifiers
- Voting Classifications
- Feature Subset Selection
- Diffusion-based Post-classification Filtering

Decision Trees

- DT has been proven as an alternative learning model for land cover classification of remotely sensed data that produced results comparable to standard methods.
- Advantages include robust in training and classification, does not assume any model on data, high repeatability, high degree of automation, high interpretability, and good accuracy.
- DT predict class membership by recursively partitioning a data set into more homogenous subsets.
- In univariate decision trees, each node is formed from a binary split of one variable.
- One important component is the method used to estimate splits at each internal node of the tree.
- C5.0 uses the 'information gain ratio' (Quinlan 1993).
- The information gain measures reduction in entropy in the data produced by a split.
- Base on this information, the node to split is selected by maximizing the reduction in entropy of the descendant node.

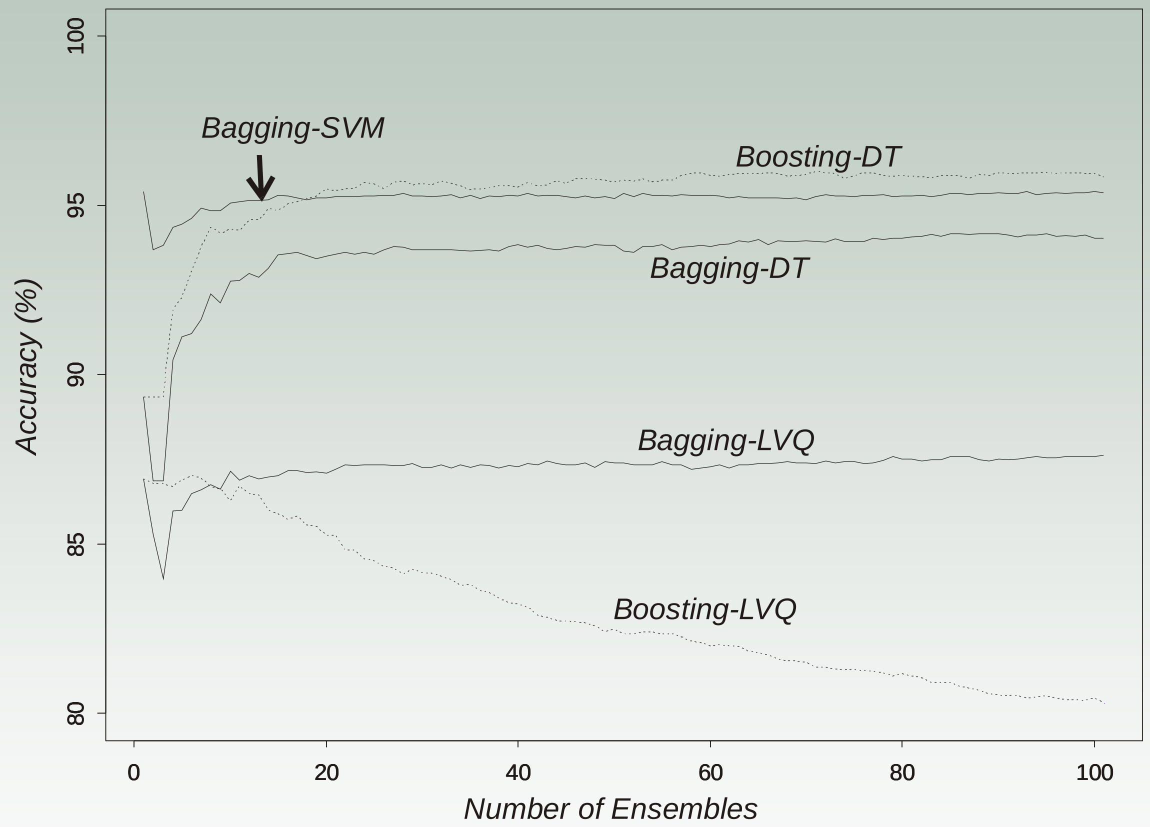
Quinlan, J.R. (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Voting Classification - Boosting

- Adaboost, the Boosting method proposed by Freund and Schapire (1996), is one of the most studied methods in voting classification.
- Adaboost attaches a higher weight to cases that are incorrectly classified in a present trial so that they have a higher probability of being chosen in a new training set which will used to create a new classifier in the next trial.
- By so doing, it forces new classifiers to focus on ‘difficult’ cases.
- After the first trial, Adaboost changes the probability of a misclassified case by the factor of $b_t = (1 - a_t)/a_t$ where a_t is the sum of the misclassified cases probabilities of a current classifier C_t at trial t.
- The total sum of the probability is then normalized to 1. If the performance is worse than a random guess (i.e. a_t is greater than 0.5), the trials will terminate and trial T becomes t-1. If $a_t = 0$, then trial T becomes t.
- Finally, the classifiers C_1, …, C_t are combined with weighted voting by $log(b_t)$.

Freund, Y. and Schapire R.E. (1996), Experiments with a new boosting algorithm, in Machine Learning: Proceedings of the Thirteenth International Conference, pages 148-156, 1996.

Figure 1



Chan, J.C-W, Huang, C. and R.S. DeFries (2001), "Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting", IEEE Transactions of Geoscience and Remote Sensing, 39 (3): 693-695.

Behavior of bagging and boosting as related to the number of ensemble classifiers

Chan, J.C-W, R.S. DeFries and J.R.G. Townshend (2001), "Improved recognition of spectrally mixed land cover classes using spatial textures and voting classifications", in Computer Analysis of Images and Patterns, 9th International Conference, CAIP'01, Warsaw, Poland September 5-7, 2001 Proceedings. Edited by Wladyslaw Skarbek, Lecture Notes in Computer Science 2124, pp. 217-227, Springer-Verlag.

Feature Subset Selection – a wrapper approach

- A wrapper approach of feature selection search for the optimal feature subset using the induction algorithm as part of the evaluation function.
- Previous studies with DT as inducer show that this approach select a smaller feature subset (Kohavi and John, 1997).
- Even when accuracy does not improve significantly, this approach generate smaller decision trees which is preferable because that eases the interpretation of the classification process.
- For a data set described by *n* features, the entire state space is of size 2^n. Unless n is small, to search the entire state space is prohibitive and impractical. A *best-first forward search* is adopted in this study to avoid searching the entire search base. The idea is to jump to the most promising node (the node with the highest estimated accuracy) generated so far that has not been expanded.
- The search is stopped if an improved node has not been found in the last *k* (=5 in this study) expansions.
- An improved node is defined as a node with an accuracy estimation of at least φ (=0.001% in this study) percent higher than the best node found so far.
- The feature subset selection in Machine Learning Library in C++ (Kohavi et al., 1996) is used for our implementation.

Kohavi, R., Sommerfield, D., and Dougherty, J. (1996), Data mining using MLC++: a machine learning library in C++, Tools with Artificial Intelligence, pp.234-245, IEEE Computer Society Press.

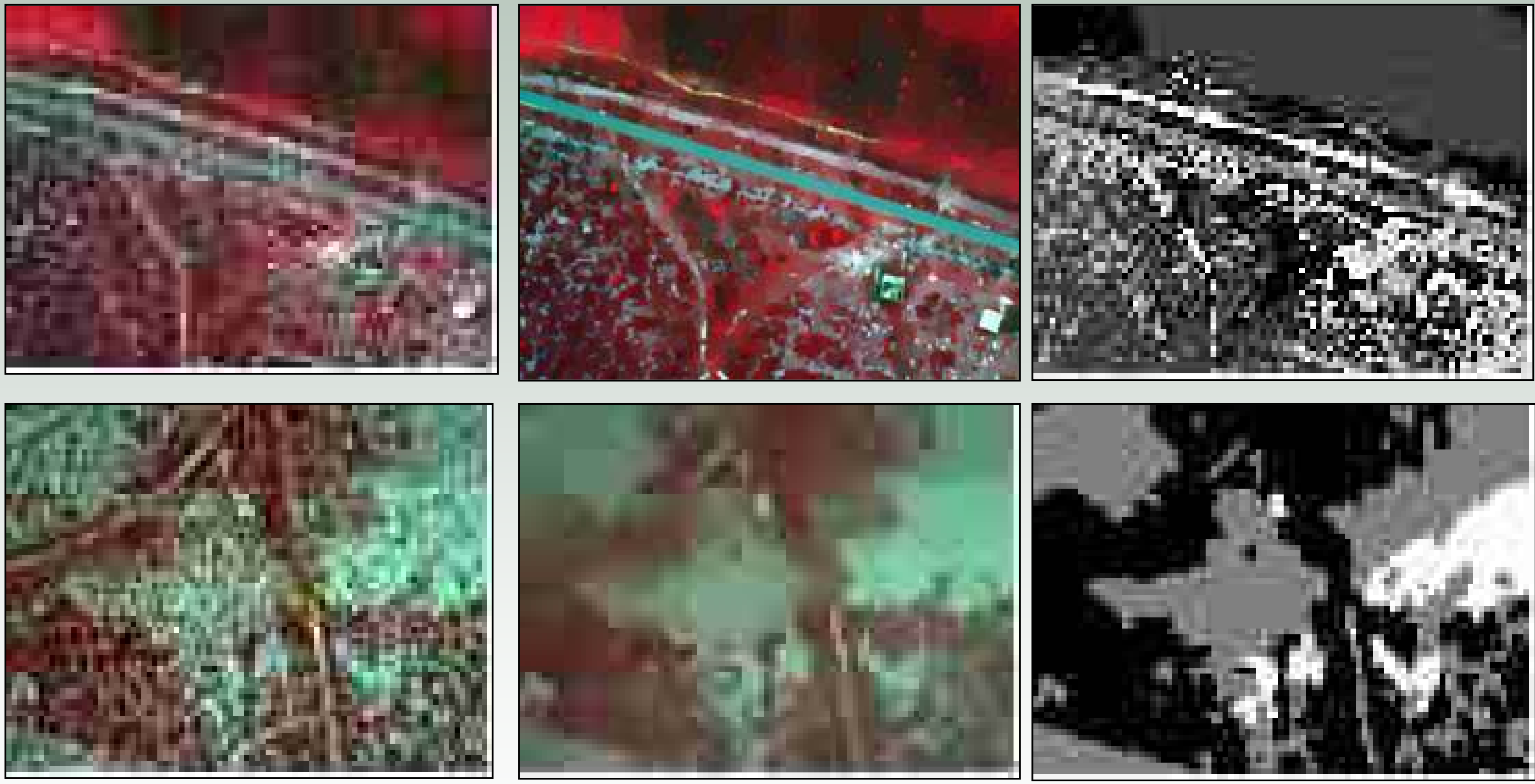
Diffusion-based post-classification processing

- Multi-scale diffusion has been investigated intensively in machine vision and image processing for many years
- Recently, multi-scale diffusion/segmentation has attracted popularity in remote sensing communities because of the availability of much finer data. Diffusion can be used to diffuse out noise and un-used information. The final products are closer to human perception.
- The concept of multi-scale diffusion is also simple: if the neighboring pixel is similar, make it more similar; if it is very different, make it more different.
- We have adopted a vector-valued diffusion scheme based upon Perona and Malik scheme which is an inhomogeneous process that reduces the amount of diffusion at locations that are more likely to be an edge

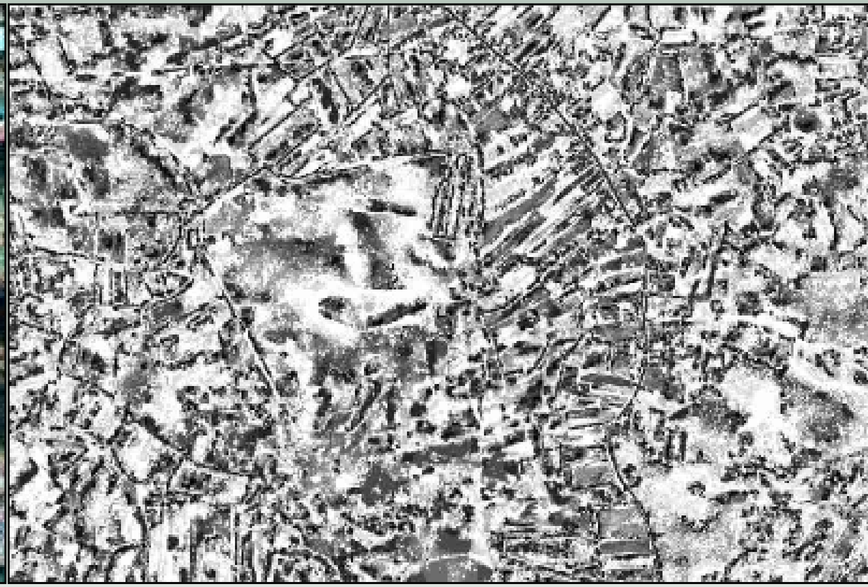
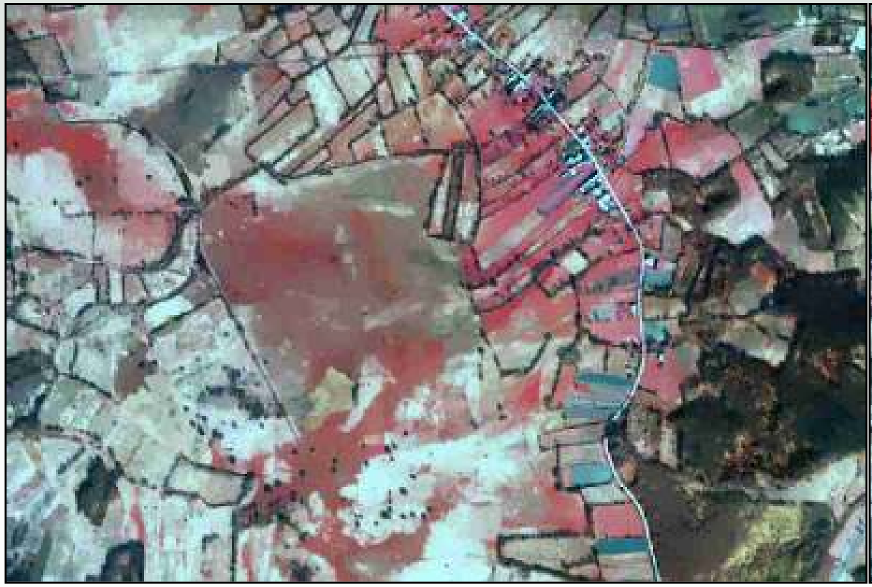
Scale Selection

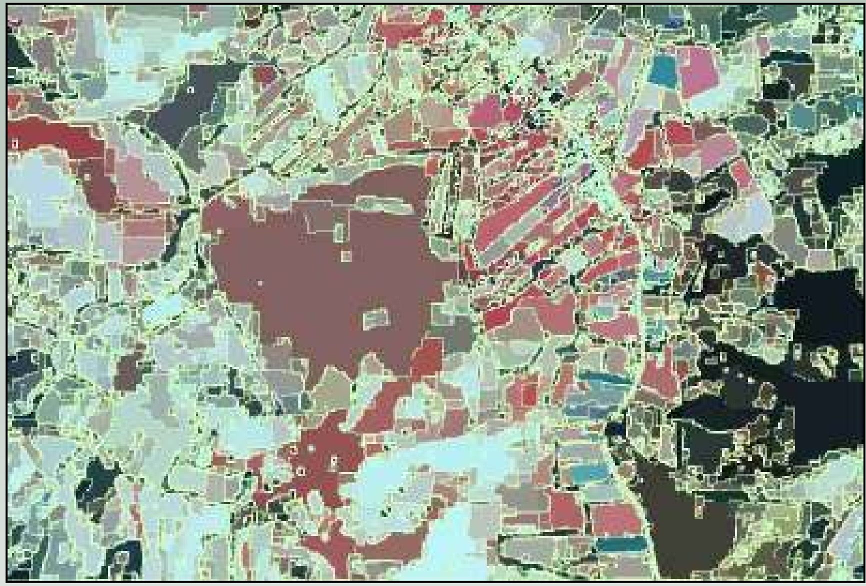
- How to choose an optimal *t* ?
- G.G'omez et al (2000) proposed a probabilistic measures for estimation of optimal local-scale for each pixel.
- The measure is based on two components: (1) a *likelihood estimator* of a diffused image to the original image, and (2) *granularity measure*, i.e. the smoothness of a diffused image.
- As an image continues to diffuse, it becomes smoother and smoother, but its likelihood to the original image declines. The rationale is to balance the smoothness with the likelihood of a scaled image.

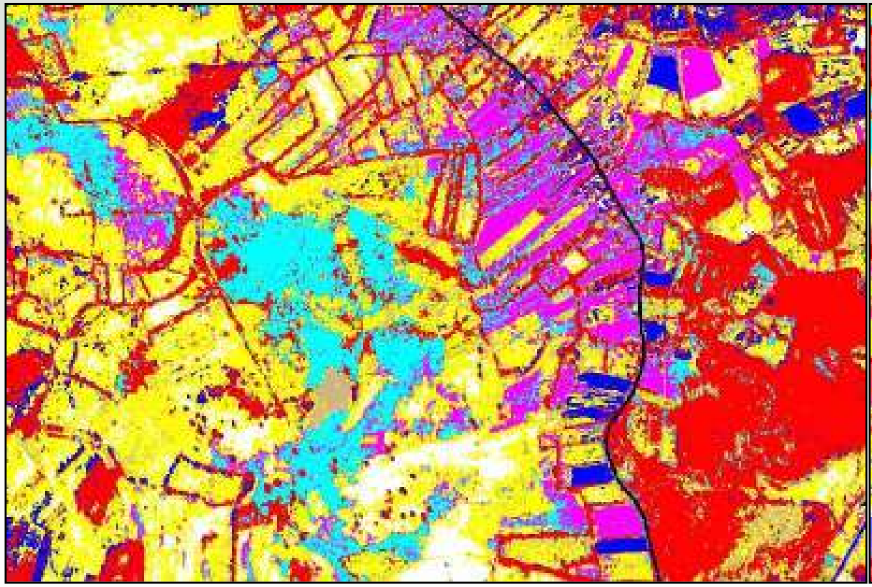
Gomez, G., Sucar, L.E. and Marroquin, J.L. (2000) Probabilistic estimation of local scale. Proceedings of the International Conference on Pattern Recognition (ICPR'00), vol 3, 798-801, Sept. 2000, Barcelona, Spain.

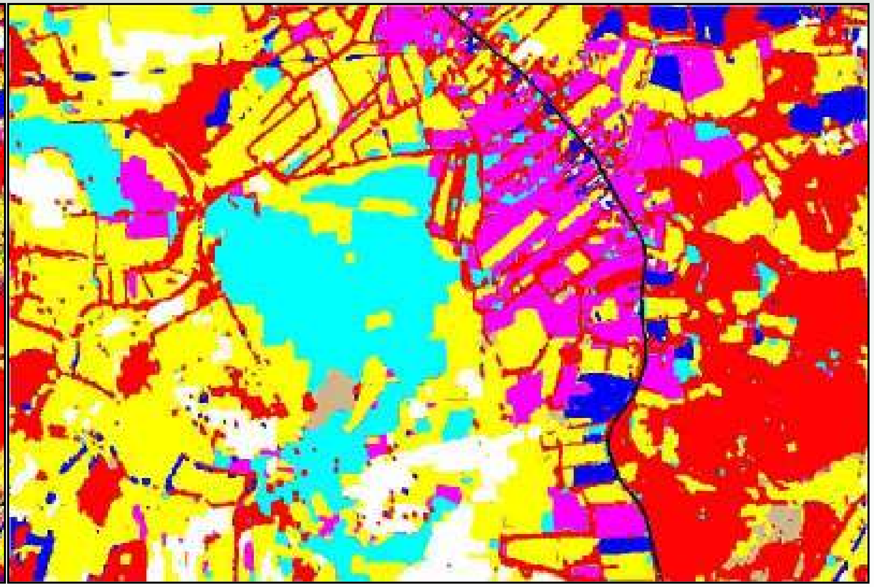


Chan, J. C-W, Vanhamel, I. and Suliga, M. (2003) Scale selection of multiscale image diffusion using probabilistic methods, Proceedings IGARSS (International Geoscience and Remote Sensing Symposium) 2003, Vol. III, pp.1811-1813, 21 to 25 July, Toulouse, France



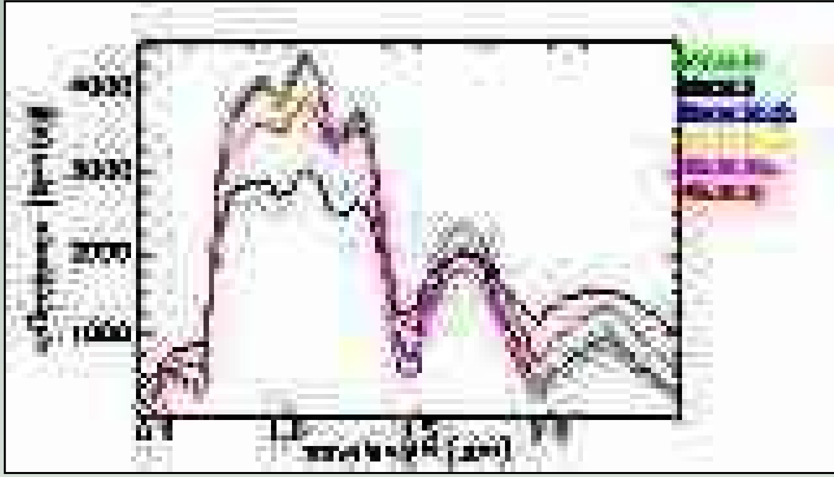


Grassland categories and Arable land

LEVEL 2			
ID	**Hierarchy**	**Class Code**	**Definitions**
2	Grassland	b	arable land
5	Grassland	hp	species poor improved grassland which are normally more homogenous for the whole parcel
6	Grassland	hpgh	semi natural grasslands
7	Grassland	hpgs	species rich improved grasslands (between hpgh & hp)
8	Grassland	hpv	grasslands with partly hp and partly scattered nature values
9	Grassland	hx	monocultures of one or sometimes some species equal to arable land sown with grasses of one or more years

Table 4.1 Classes belonging to grassland category and their definitions

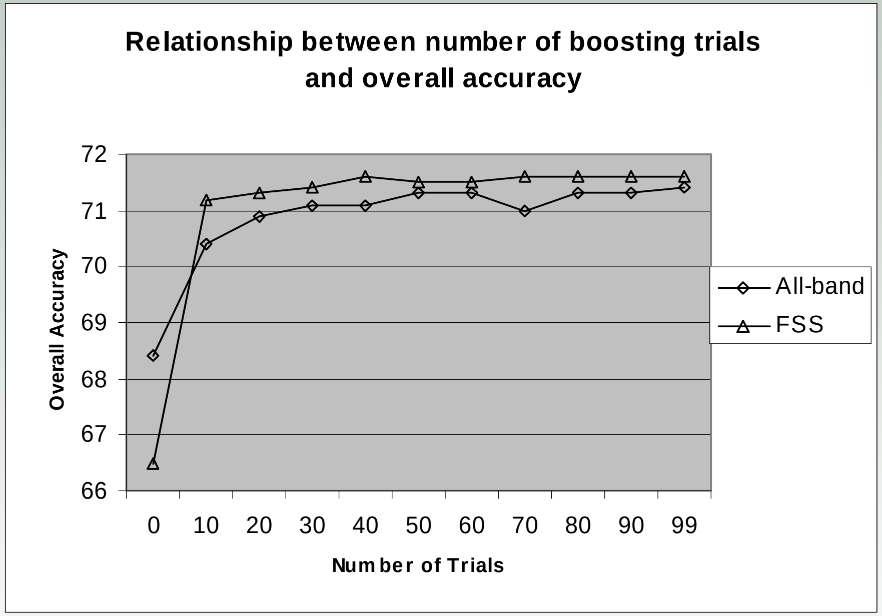
Final Training Samples for grasslands and arable land

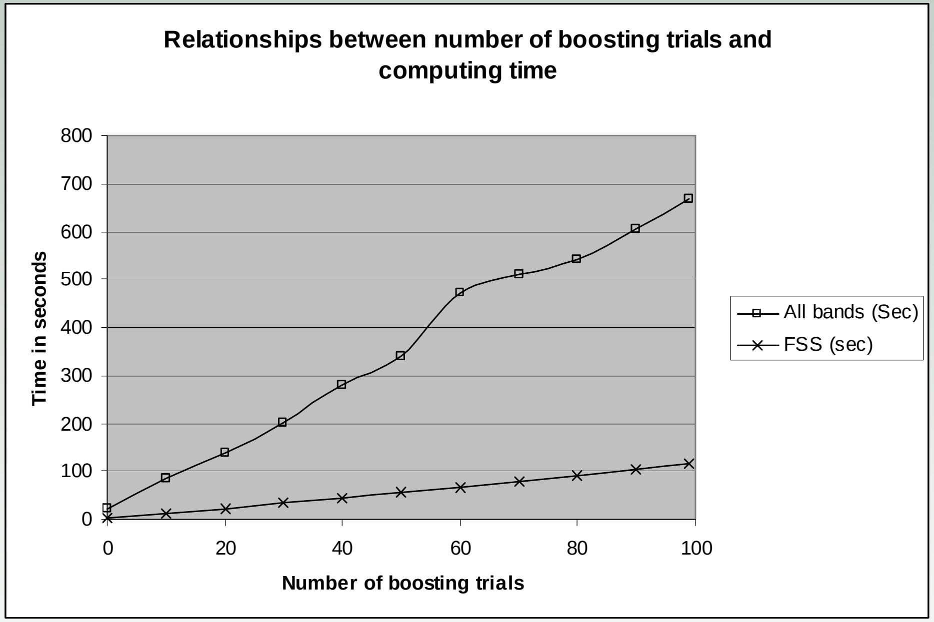


Class name	Number of cases for training	Number of cases for testing
b	5240	48028
hp	1692	10145
hpgh	292	1732
hpgs	11963	65721
hpv	573	15323
hx	475	334
Total	20235	141283

Figure 4.4. The clean ground truth area of the arable land and grassland categories that will be used for experiment. The grassland classes are filtered from recently harvested or ploughed area, as well as the edge pixels that might contain mixed pixels.

Comparison between All-band and FSS bands
Grasslands and Arable Land





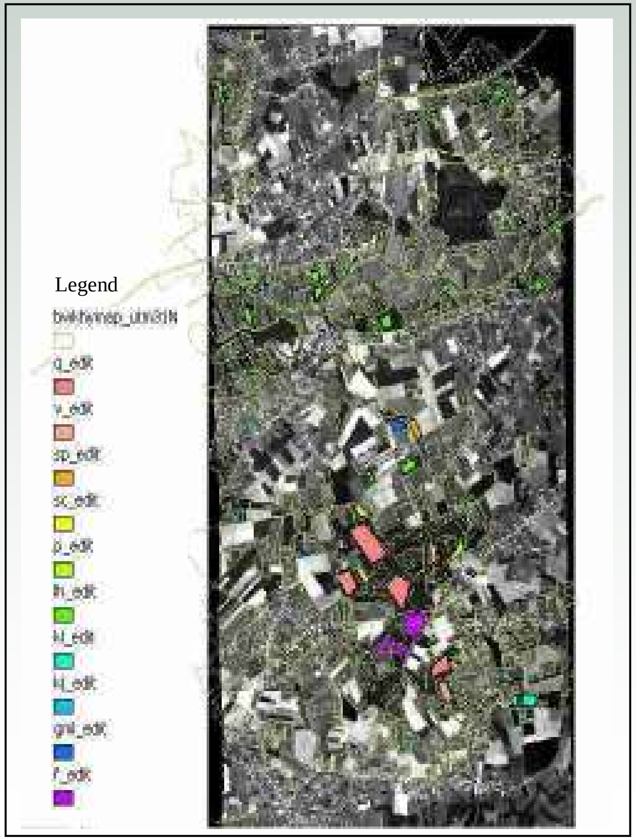
Tree Categories: classes and definitions

LEVEL 2		
Hierarchy	**Class Code**	**Definitions**
tree/tall_veg	f	deciduous forests dominated by Fagus
tree/tall_veg	gml	plantations of deciduous tree species other than Fagus, Quercus, Alnus and Poplar
tree/tall_veg	kj	tall tree orchard
tree/tall_veg	kl	low tree orchard
tree/tall_veg	lh	poplar plantations on wet soils
tree/tall_veg	p	conifer plantation
tree/tall_veg	q	deciduous forest dominated by Quercus
tree/tall_veg	sc	se (clearings) + sz (scrubs on abandoned land)
	sp	thorn ticket
tree/tall_veg	v	woodland of alluvial soil, fens and bogs

Final Training Samples for Tree Categories

	f	***gml***	***kj***	***kl***	***lh***	***p***	***q***	***sc***	***sp***	***v***	***Total***
Training	1663	852	369	607	2802	320	4907	675	361	1259	13815
Testing	1662	851	369	607	2802	319	4906	675	360	1259	13810
Total	3325	1703	738	1214	5604	639	9813	1350	721	2518	27625

Table 5.2 Number of cases for training and testing



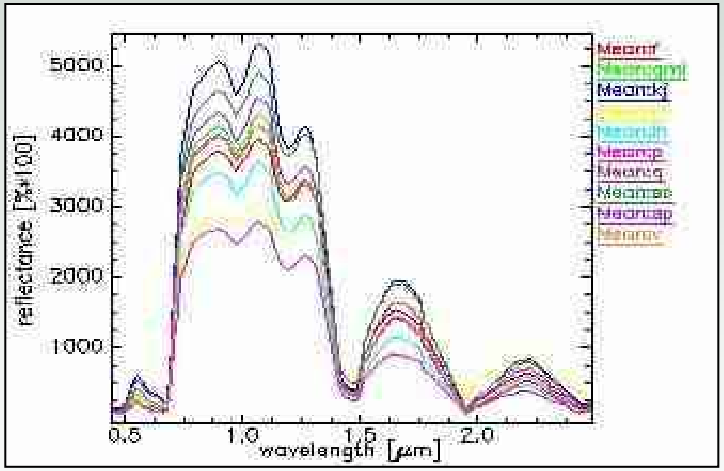
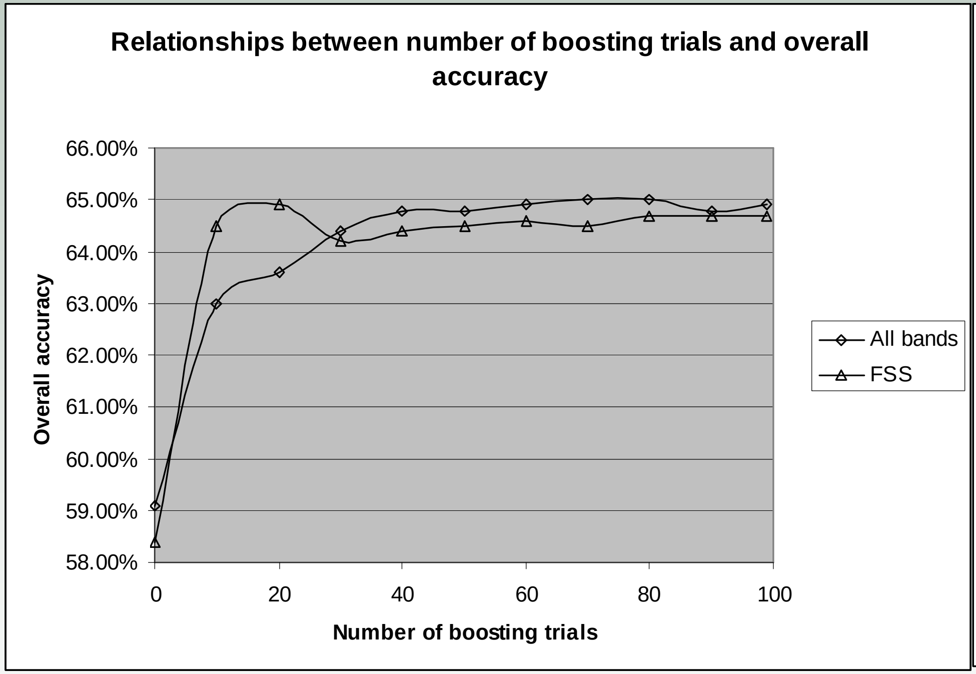
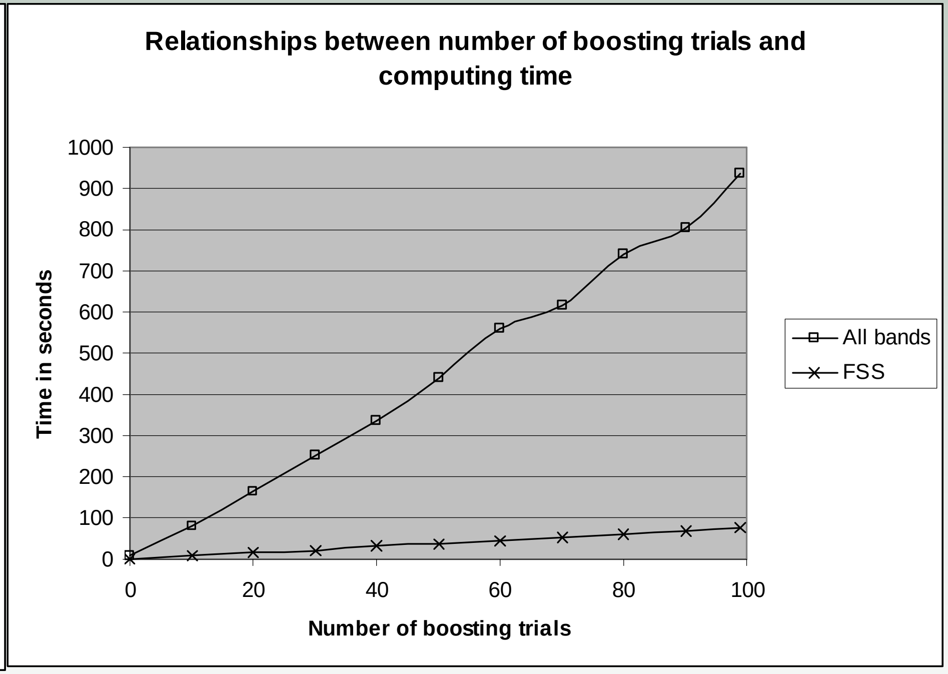


Figure 5.2 Mean spectral profiles of the ten tree categories generated from the training areas

Comparison between All-band and FSS bands
Tree classes





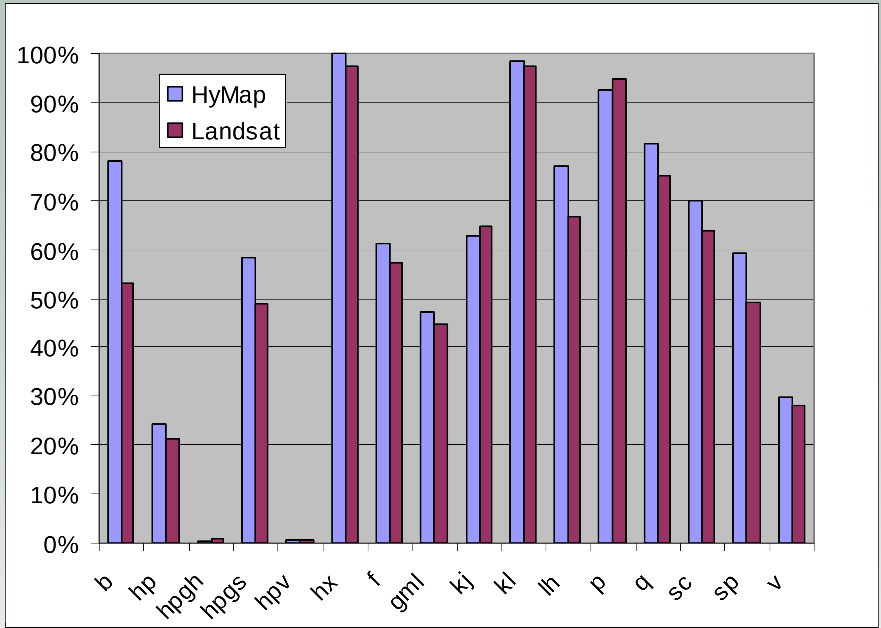
Training and testing cases for combined grassland and tree categories

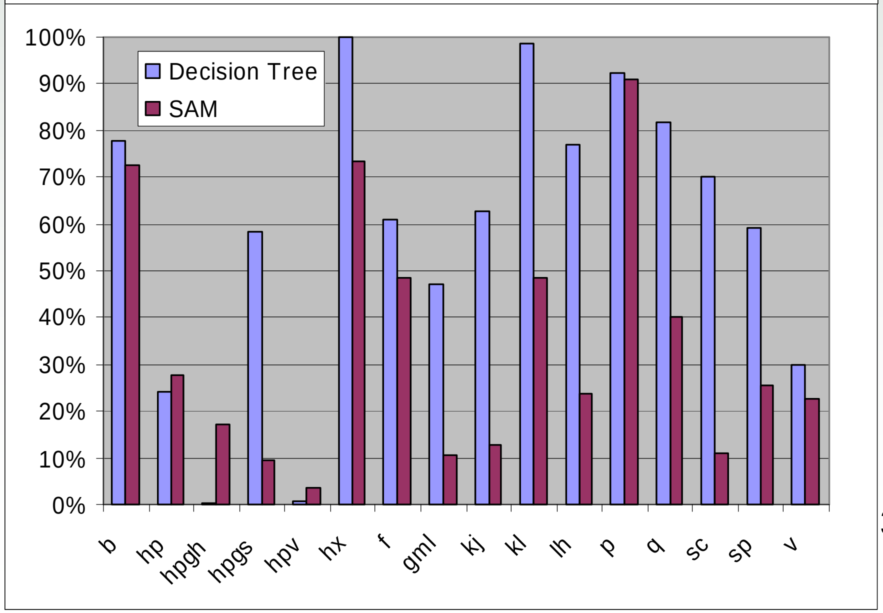
ID	Classes	Cases for training	Cases for testing
1	*b*	2620	4803
2	*hp*	846	1015
3	*hpgh*	146	1732
4	*hpgs*	5981	6573
5	*hpv*	286	1533
6	*hx*	237	334
7	*f*	1663	1662
8	*gml*	852	851
9	*kj*	369	369
10	*kl*	607	607
11	*lh*	2802	2802
12	*p*	320	319
13	*q*	4907	4906
14	*sc*	675	675
15	*sp*	361	360
16	*v*	1259	1259
	Total	**23931**	**29800**

Table 6.1 Number of cases in training and testing

Comparison between HyMap data and Similated Landsat data
Comparison between Decision Tree and Spectral Angular mapper

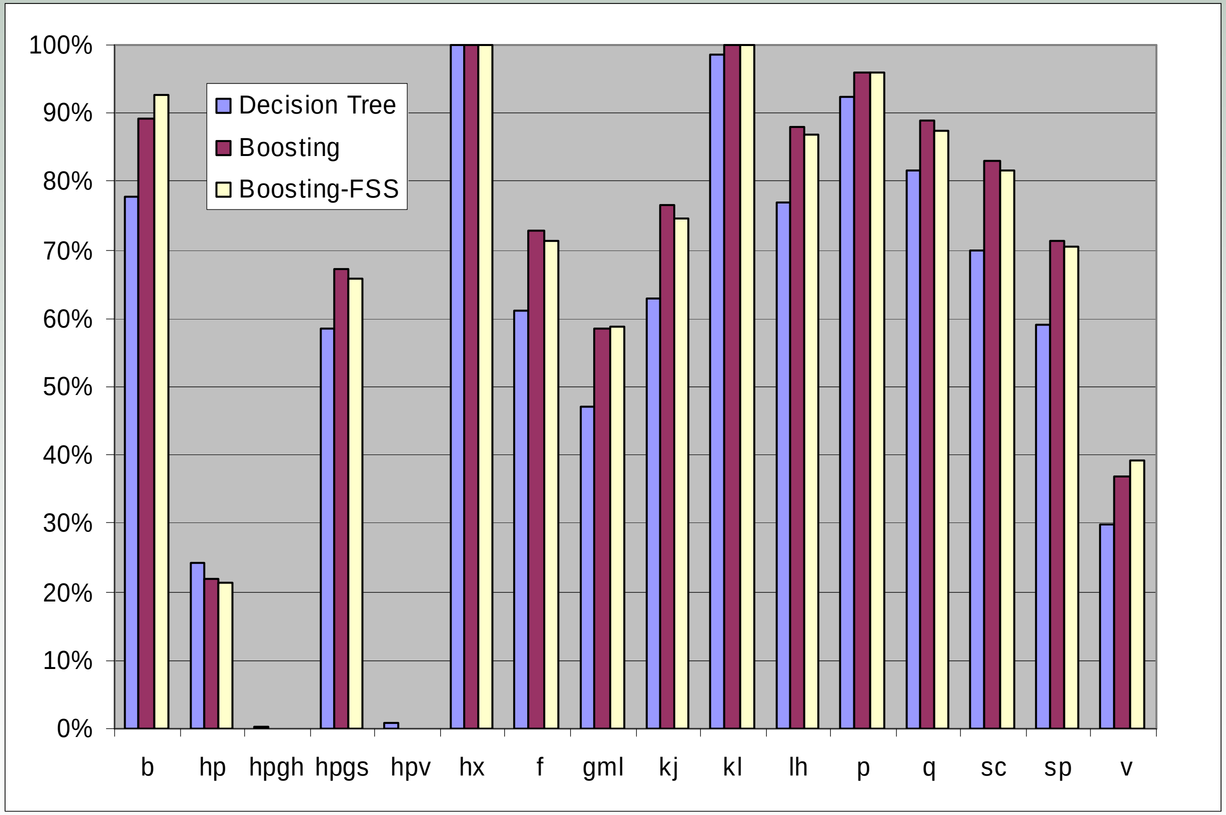
	HyMap	Landsat	SAM
b	78%	53%	73%
hp	24%	22%	28%
hpgh	0%	1%	17%
hpgs	58%	49%	10%
hpv	1%	1%	4%
hx	100%	97%	74%
GRASS	**51%**	**40%**	**26%**
f	61%	57%	49%
gml	47%	45%	11%
kj	63%	65%	13%
kl	98%	98%	49%
lh	77%	67%	24%
p	93%	95%	91%
q	82%	75%	40%
sc	70%	64%	11%
sp	59%	49%	26%
v	30%	28%	23%
TREE	**71%**	**65%**	**37%**
Overall	**60%**	**49%**	**29%**

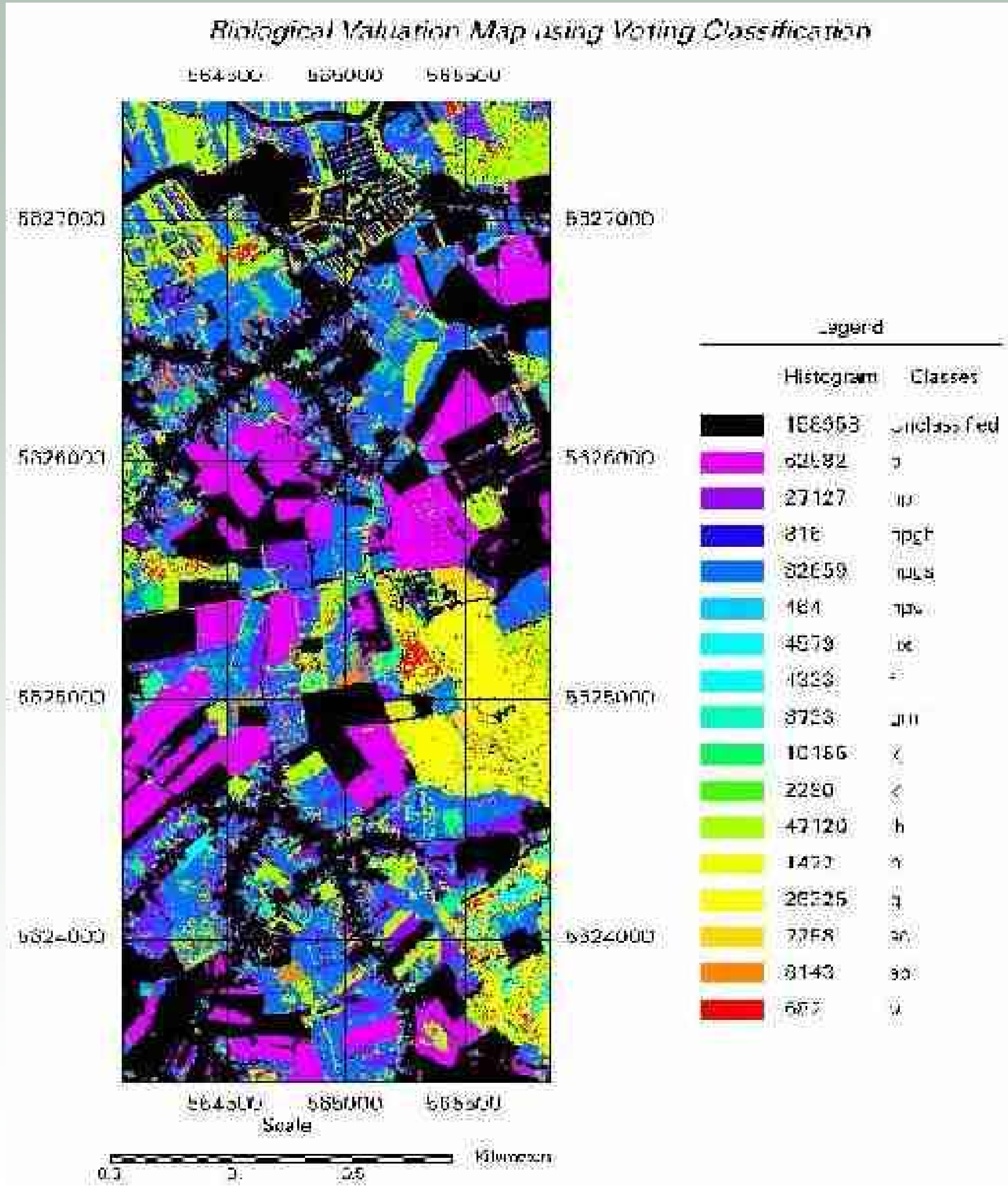




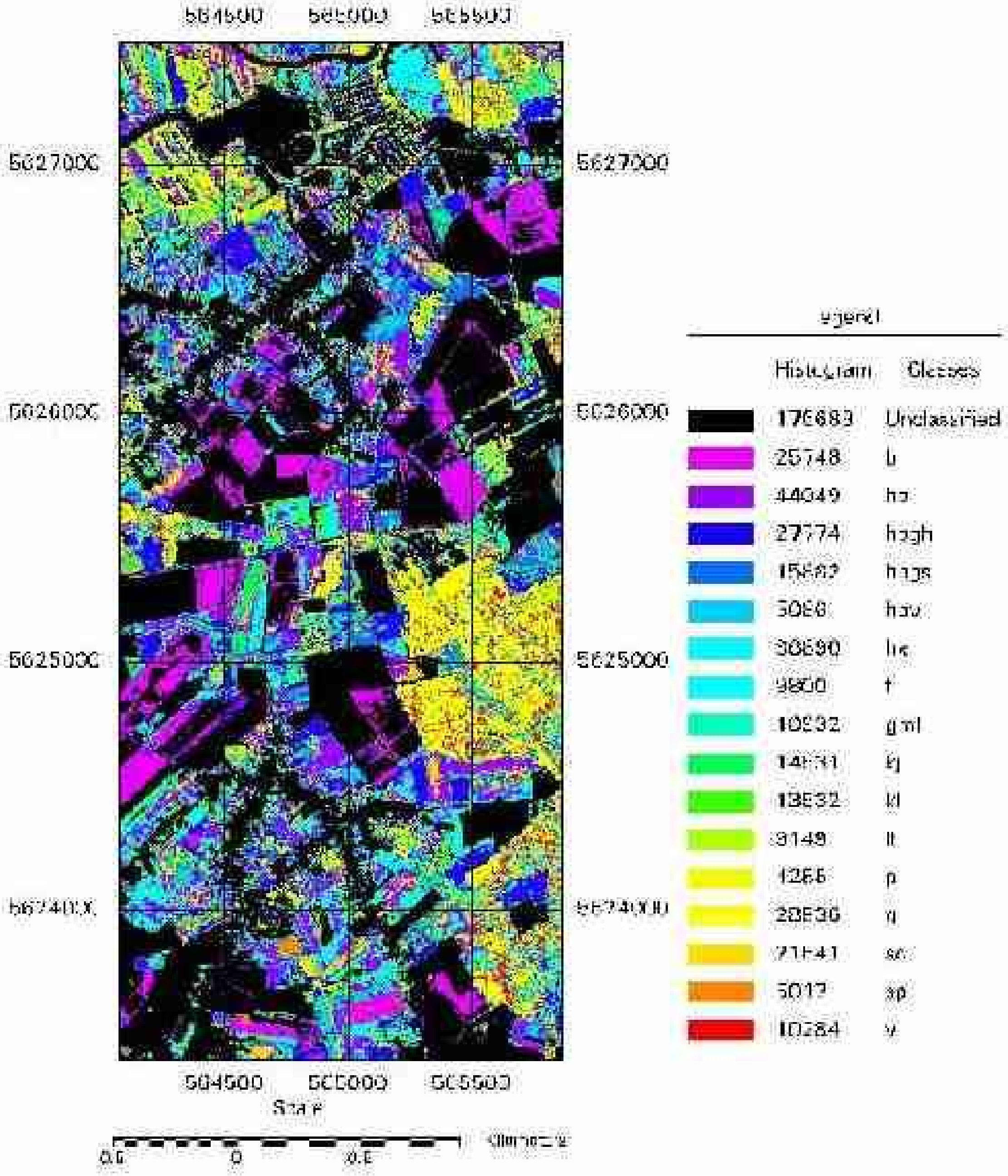
Performance of Voting Classification and Feature Subset Selection

	DT	Boost	FSS-Boost
b	**78%**	**89%**	**93%**
hp	**24%**	**22%**	**21%**
hpgh	**0%**	**0%**	**0%**
hpgs	**58%**	**67%**	**66%**
hpv	**1%**	**0%**	**0%**
hx	**100%**	**100%**	**100%**
GRASS	**51%**	**58%**	**58%**
f	**61%**	**73%**	**71%**
gml	**47%**	**59%**	**59%**
kj	**63%**	**77%**	**75%**
kl	**98%**	**100%**	**100%**
lh	**77%**	**88%**	**87%**
p	**93%**	**96%**	**96%**
q	**82%**	**89%**	**87%**
sc	**70%**	**83%**	**82%**
sp	**59%**	**71%**	**71%**
v	**30%**	**37%**	**39%**
TREE	**71%**	**80%**	**79%**
Overall	**60%**	**68%**	**68%**




Biological Valuation Map using Voting Classification
564500
565000
565500
5827000
5826000
5825000
5824000
Legend
Histogram
Classes
27127
47120
Scale
Kilometers

Biological Valuation Map using Spectral Angle Mapper



Post-classification Filter generated by Multi-scale Diffusion

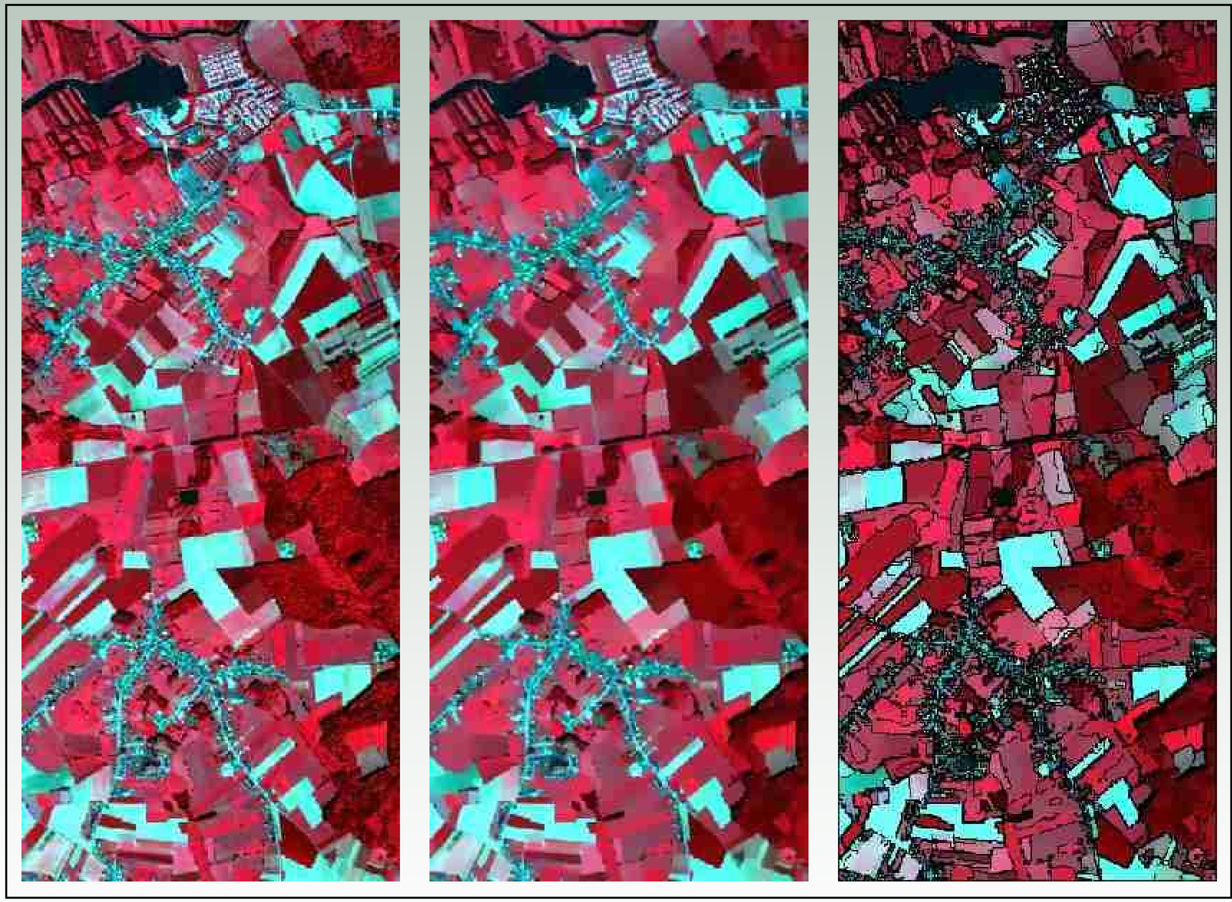
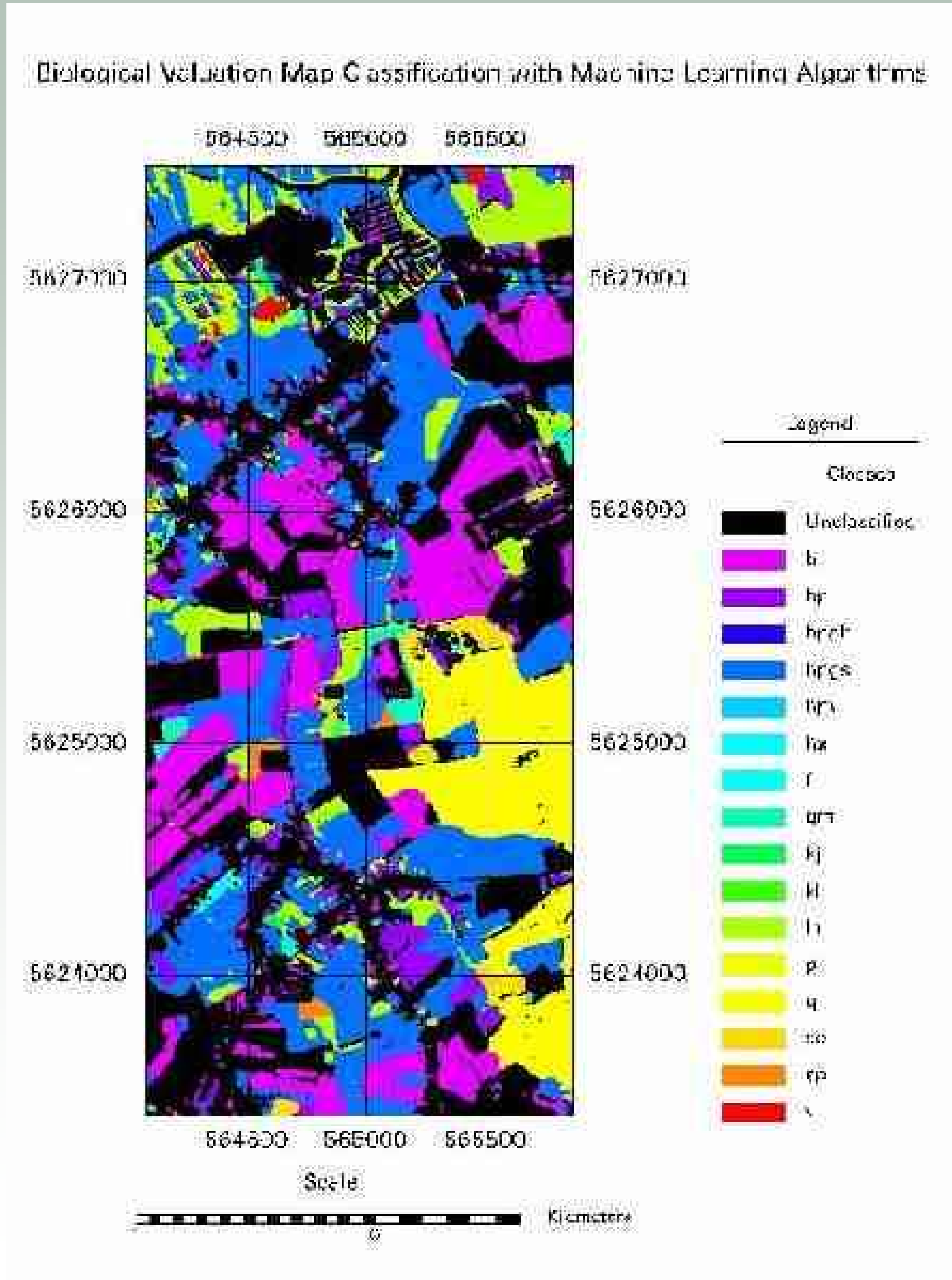


Figure 6.3 A multiscale-based filter is generated using statistical selection of an optimal diffused scale. Left: Original image. Middle: diffused image at the selected scale. Right: segments generated from the diffused image.


Biological Valuation Map Classification with Machine Learning Algorithms
564500
565000
565500
5627000
5626000
5625000
5624000
Legend
Classes
Unclassified
Scale
Kilometers
0

Analysis of FSS wavebands

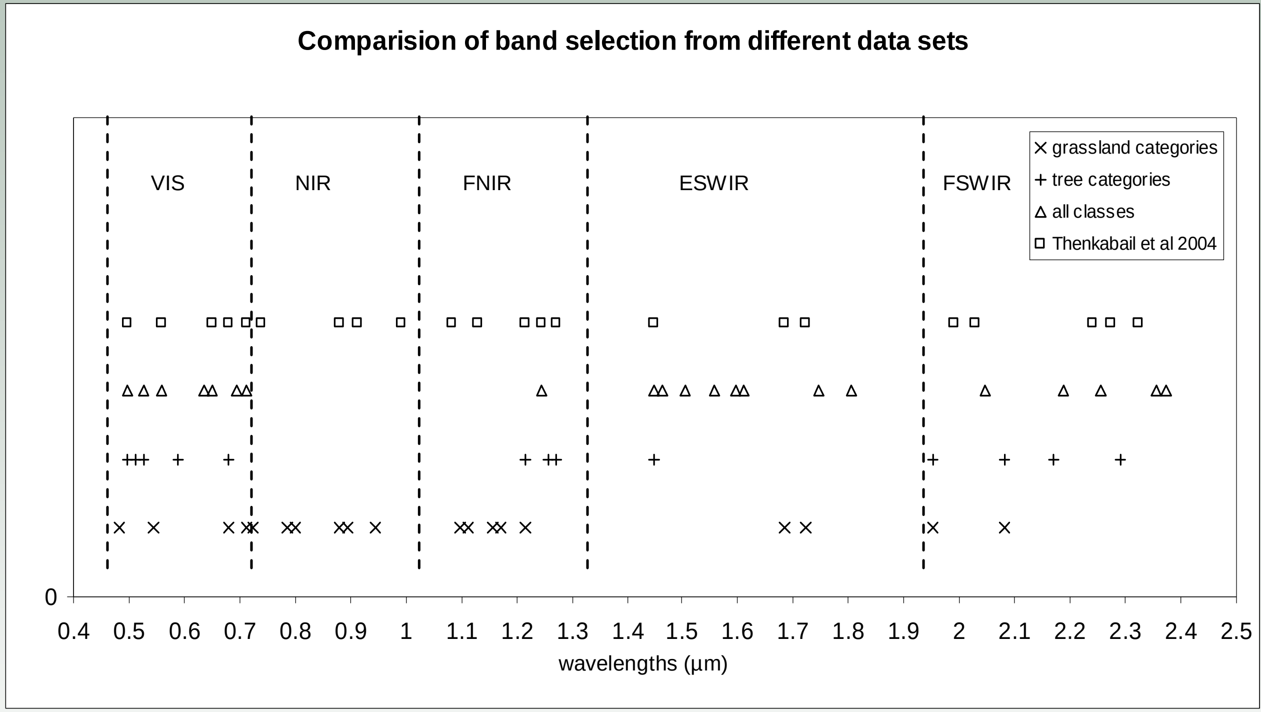
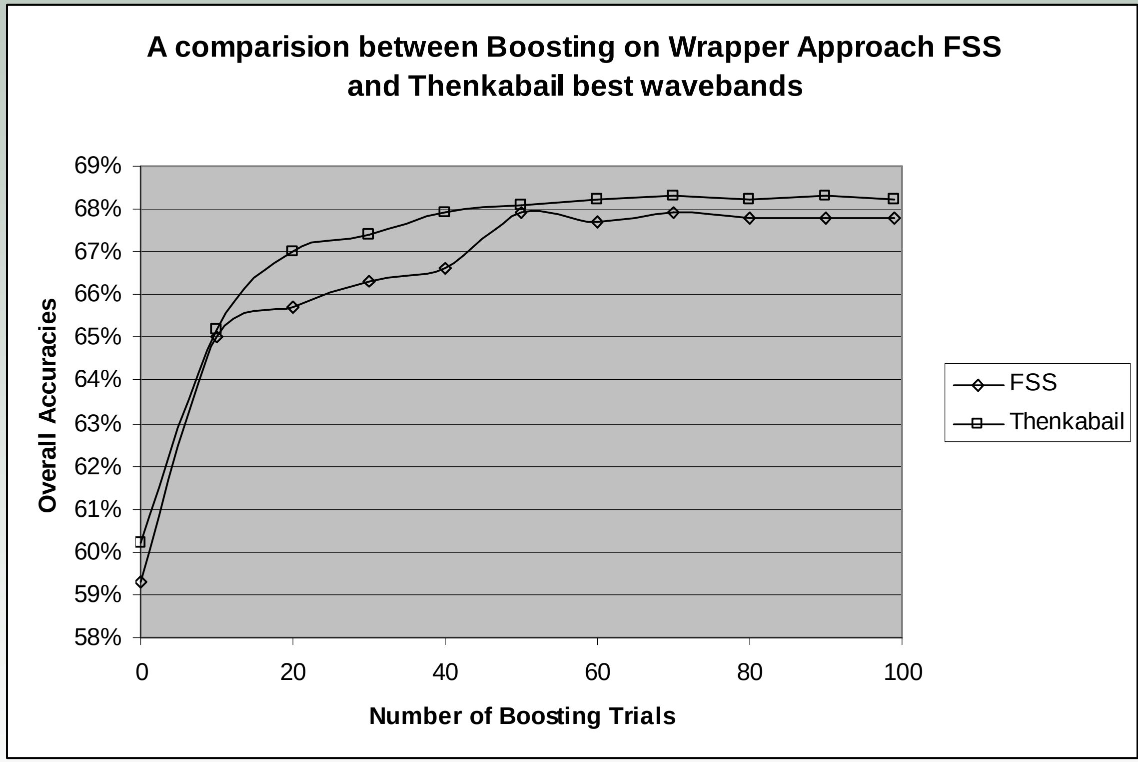


Figure 6.5 A comparison of band selection performed on different data sets (VIS=visible, NIR=near infrared, FNIR=far near-infrared, ESWIR=early short wave infrared, FSWIR=far short wave infrared.)

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., and Van Der Meer, B. (2004), Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications, Remote Sensing of Environment, 91: 354-376.



A comparision between Boosting on Wrapper Approach FSS
and Thenkabail best wavebands
69%
68%
67%
66%
65%
64%
63%
62%
61%
60%
59%
58%
Overall Accuracies
0
20
40
60
80
100
Number of Boosting Trials
FSS
Thenkabail

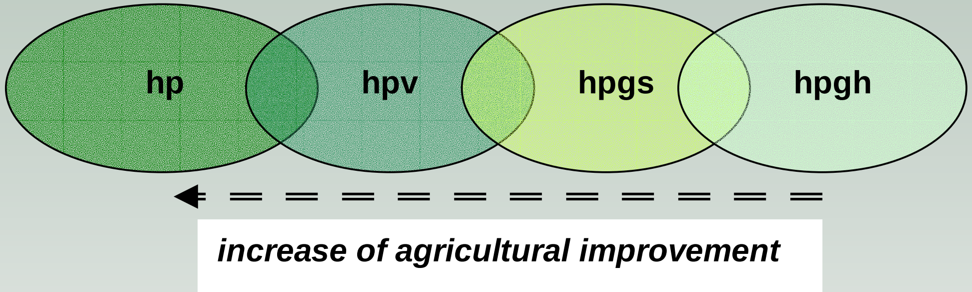
Discussions and Conclusions

Is the proposed methodology suitable for classification of ecotope, or Biological Valuation Map?

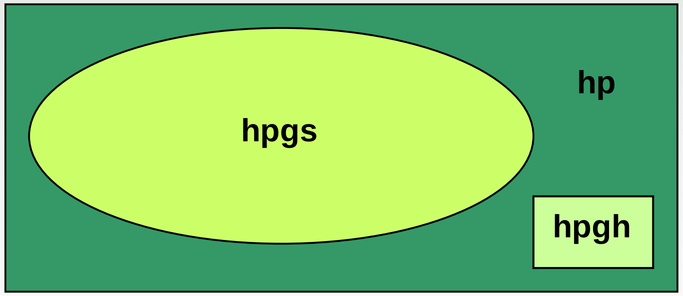
Technical Aspects:

- Hyperspectral data (60% accuracy) is superior than similated Landsat TM wavebands (48.6%)
- Decision Trees provide robust training and classification. In terms of accuracy Decision Tree approached 60%, outperformed SAM.
- Voting Classification (Boosting) increased accuracy by 5-7% on average, to 68%. The methodology seems to be robust with good accuracy.
- FSS is successful in selecting a feature subset (10-13% of the original set) without degrading accuracy. The contribution is mainly in dimension reduction and minimizing computing time (86% reduction at 99 boosting trials). The reduction in dimension with chosen input features simplified significantly data handling and processing.

Application Aspects:
In reality, the accuracy might be higher than 70% because:
1. It is common for overlapping classes within the grassland categories.



2. Possibility of heterogeneity within parcel.



Hierarchy	Class Code	Definitions
Grassland	*b*	arable land
Grassland	*hp*	species poor improved grassland which are normally more homogenous for the whole parcel
Grassland	*hpgh*	semi natural grasslands
Grassland	*hpgs*	species rich improved grasslands (between hpgh & hp)
Grassland	*hpv*	grasslands with partly hp and partly scattered nature values
Grassland	*hx*	monocultures of one or sometimes some species; equal to arable land sown with grasses of one or more years

Confusion Matrix of boosting 99 trials with overall accuracy at 68%.
Grassland Categories accuracy = 9262/15990*100 = 58%

	b	*hp*	*hpgh*	*hpgs*	*hpv*	*hx*	*f*	*gml*	*kj*	*kl*	*lh*	*p*	*q*	*sc*	*sp*	*v*	%
b	4279	33	0	72	0	17	0	84	2	5	235	0	0	17	59	0	**89.1**
hp	75	223	1	448	0	21	0	3	125	4	34	0	5	55	21	0	**22.0**
hpgh	1	49	1	1220	0	35	1	24	76	28	115	0	4	168	2	8	**0.1**
hpgs	130	749	23	4424	30	79	2	175	236	28	443	7	43	137	56	11	**67.3**
hpv	28	160	3	968	1	15	0	88	64	2	133	1	1	22	42	5	**0.1**
hx	0	0	0	0	0	334	0	0	0	0	0	0	0	0	0	0	**100.0**

‘*Arable land*’ (***b***) and ‘*temporary species poor improved grasslands*’ (***hx***) can be distinguished with a high degree of accuracy and that both differ clearly from the ‘*permanent grasslands group*’ (***hp – hpv – hpgs – hpgh***). If machine learning techniques described could be used to produce maps with these three categories, it is important for environmental policies because there are some legislation rules concerning permanent grasslands (e.g. see Wils & Paelinckx 2004). A large reduction on field works.

Hierarchy	Class Code	Definitions
tree/tall_veg	***f***	deciduous forests, dominated by Fagus
tree/tall_veg	***gml***	plantations of deciduous tree species other than Fagus, Quercus, Alnus and Poplar
tree/tall_veg	***kj***	tall tree orchard
tree/tall_veg	***kl***	low tree orchard
tree/tall_veg	***lh***	poplar plantations on wet soils
tree/tall_veg	***p***	conifer plantation
tree/tall_veg	***q***	deciduous forests, dominated by Quercus
tree/tall_veg	***sc = se +sz***	scrubs of clearings + scrubs on abandoned land
tree/tall_veg	***sp***	thorn ticket
tree/tall_veg	***v***	woodland of alluvial soil, fens and bogs

Table 6.4 Confusion Matrix of boosting 99 trials with overall accuracy at 68%.
Tree Categories accuracy = 11012/13810*100 = 80%

	b	*hp*	*hpgh*	*hpgs*	*hpv*	*hx*	*f*	*gml*	*kj*	*kl*	*lh*	*p*	*q*	*sc*	*sp*	*v*	%
f	0	0	0	4	0	0	1209	3	0	1	50	0	361	3	0	31	**72.7**
gml	7	2	1	23	0	0	11	498	0	0	164	0	114	11	0	20	**58.5**
kj	3	2	0	71	0	0	0	1	283	0	1	0	0	3	5	0	**76.7**
kl	0	0	0	0	0	0	0	0	0	607	0	0	0	0	0	0	**100.0**
lh	7	4	0	14	0	0	39	24	2	6	2468	12	122	13	5	86	**88.1**
p	0	0	0	0	0	0	1	0	0	0	9	306	2	0	0	1	**95.9**
q	0	0	0	4	0	0	252	29	4	1	139	0	4361	16	1	99	**88.9**
sc	0	0	0	32	0	1	1	7	0	5	15	0	28	561	9	16	**83.1**
sp	1	2	3	18	0	6	0	3	4	0	54	0	3	5	257	4	**71.4**
v	0	2	0	15	0	0	63	22	1	0	226	4	433	30	1	462	**36.7**

Most classes has quite good accuracies. The results indicate that for forests and plantations stands of different dominant tree species can be distinguished. In all cases it is possible that the degree of accuracy is even higher since the dominated species is used to label the whole parcel. Remain to be explained is the low accuracy of ‘***v***’.

With the above observations, we concluded the proposed methodology in this study is promising for ecotope mapping.

Future Researches

- Comparison of the map derived from the HyMap classification and a field driven map in a quantitative way and research the resemblances and differences with feed back to the field
- What are the influence of the amount of standing biomass for grassland typology (e.g. classification before and after mowing)?
- Expanding this research to more localities and/or more types of land covers, with special attention to the Natura 2000 habitat types
- Special attention to, and extending to a larger set of special chosen testing sites for more difficult habitat types (e.g. the “v”-class in this research)
- Building libraries of spectra for different types of land cover, vegetation types, habitats, ...
- The possibilities for detecting change including researches on repeatability of using the methodology for monitoring