Mapping dominant vegetation communities in Meili Snow Mountain, Yunnan Province, China using satellite imagery and plant community data

Zhiming Zhang
Institute of Ecology and Geobotany, Yunnan University, China
ming@biogroup.net

Eva De Clercq
Laboratory of Forest Management and Spatial Information techniques, Ghent University, Belgium
eva.declercq@UGent.be

Xiaokun Ou
Institute of Ecology and Geobotany, Yunnan University, China
xkou@ynu.edu.cn

Robert De Wulf
Laboratory of Forest Management and Spatial Information techniques, Ghent University, Belgium
Robert.DeWulf@UGent.be

Lieven Verbeke
Laboratory of Forest Management and Spatial Information techniques, Ghent University, Belgium
Lieven.Verbeke@UGent.be

Abstract: Mapping dominant vegetation communities is important work for vegetation scientists. It is very difficult to map dominant vegetation communities using only multispectral remote sensing data, especially in mountain areas. However the plant community data contains useful information about the relationships between plant communities and their environment. In this paper plant community data are linked with remote sensing to map vegetation communities. The Bayesian soft classifier was used to produce posterior probabilities images for each class. Then, these images were used to calculate the prior probabilities. One hundred eighty plant plots in Meili Snow Mountain, Yunnan Province, China were used to characterize the vegetation distribution for each class along altitude gradients. Then, the frequencies were used to modify the prior probabilities of each class. After stratification in a vegetation part and a non-vegetation part, a maximum-likelihood classification with equal prior probabilities was conducted, yielding an overall accuracy of 82.1% and a kappa accuracy of 0.797. Maximum-likelihood classification with modified prior probabilities in the vegetation part, conducted with a conventional maximum-likelihood classification for the non-vegetation part, yielded an overall accuracy of 87.7%, and a kappa accuracy of 0.861.
Keywords: dominant vegetation communities, prior probabilities, Meili Snow Mountain China, Maximum likelihood classification

1. Introduction:

Vegetation mapping and vegetation ecological research form an important contribution to management strategies and plans, and environmental conservation. For many years vegetation has been mapped by using information from the external aspects of the landscape, such as changes in elevation or soil type and by manually drawing boundaries between dissimilar units. Initially, this procedure was carried out in the field. With the establishment and progress of aerial photography and satellite remote sensing, boundaries could be drawn by visual interpretation of images and updated by field visits. However, identifying vegetation associations and locating their boundaries according to visual interpretation in the landscape is rather subjective. Moreover, conventional mapping neglects short-range variation[1]. Satellite images are a possible alternative for the direct retrieval of certain spatially distributed vegetation characteristics, but with current technologies, a clear identification of individual plant species other than trees is hardly possible[2].

For dominant community vegetation mapping, aerial photographs or high resolution satellite images might be sufficient for projects at a regional scale. However, when considering projects at a local scale, and considering the cost of the imagery and ancillary data, it is necessary to find other methods to support the quick and cheap acquisition of a vegetation map with the right balance between accuracy and costs.

Mountain ecosystems exhibit some of the steepest environmental gradients on Earth. In mountain areas, water and temperature determine the spatial patterns of the vegetation[3], [4]. Other factors like solar radiation and wind also influence vegetation distribution[5]. Solar radiation is the primary atmospheric control over soil moisture status between precipitation events in vegetation not receiving meltwater and appears to influence the local adaptation of vegetation. Relationship between slope aspect and solar radiation determine the range of vegetation types along the altitudinal gradient.

Several methods have been proposed to incorporate ancillary data, such as a digital elevation model (DEM), to improve the image classification in mountain areas [5], [6], [7], [8], [9], [10], [11], [12]. One means of incorporating ancillary information in the classification is by modifying prior probabilities of class membership. This can improve classification results by resolving confusion among classes that are poorly separable spectrally, and by reducing bias when the training sample is not representative of the population being classified[13]. The principle of how to modify prior probabilities to improve the accuracy of the classification results is discussed by Strahler[14]. Many authors have demonstrated that this methodology has significantly improved the performance of a maximum-likelihood classifier[13], [15], [16], [17], [18], [19].

However, most of the research work has focused on integrating DEM or multi-temporal images to improve image classification[5], [6], [7], [8], [9], [10], [11], [12], [20], [21], [22], [25]. Rarely, the focus has been on plant community data in mountain regions[1], [23]. As mentioned above, in mountain areas vegetation distribution is affected by environmental factors[4], [24]. Numerous authors have shown the relationships between ecosystem structure and composition and topographic features such as elevation, slope angle, aspect and indices of relative moisture based on potential solar radiation and topographic redistribution of precipitation[5], [25], [26]. Combining the plant community data and vegetation distribution for image classification has proven to be effective, especially for the mapping of dominant plant vegetation communities[12], [23]. Currently, there exists a need to establish and clarify the link between theory and practice of classification by ecologists on one side, and remote sensing scientists on the other side[23].

In this study, plant community data has been used to calculate the prior probabilities of each vegetation type in function of altitude. Combining this data with digital elevation (DEM) and Landsat Thematic Mapper (TM) data, a dominant community vegetation map was derived by means of the maximum likelihood algorithm.

2. Background:

1) Vegetation zones in Meili Snow Mountain

The spatially distributed vegetation characteristics in Meili Snow Mountain have been categorized along an altitudinal gradient into six zones, based on vegetation dominance[27]:

- Dry and warm valley shrub zone (below 2500 m)

The dominant species are *Sophora viciifolia*, *Platycladus orientalis*, *Spiraea* sp, and *Elsholtzia capituligera*. In this zone, most plants are drought-resistant because the environmental conditions are arid.

- Warm coniferous forest (between 2500 m and 3000 m)

The dominant species are *Pinus densata, Pinus armandii, Corylus chinensis.* As the altitude gradient changes, the vegetation types also change. In this zone, the dominant vegetation communities belong to pine forest.

- Coniferous and broad leave mixed forest (between 3000 m and 3500 m)

The dominant species are *Pinus armandii, Acer* sp., *Populus* sp., *Betula* sp., *Tsuga* sp., *Quercus panosa, Pseudotsuga forrestii, Abies* sp., *Picea* sp., *Pinus densata.* Compared with other zones, the biodiversity is very high in this zone.

- Cold coniferous forest (between 3500 m and 4300 m)

This dominant vegetation community is the typical subalpine forest type. The dominant species are *Abies georgei, Abies georgei var. smithii, Abies forrestii, Picea likiangensis, Picea brachytyla var. camplantata*. This type of forest occurs below the tree line, which is situated at about 4300 m.

- Alpine shrub and meadow (between 4300 m and 4500 m)

This zone is above the treeline and below the glacier and alpine talus. The dominant species are *Berberis* sp., *Rhododendron* sp., *Cassiope* sp., *Salix myrtillacea, Sambucus adnata, Fragaria orientalis, Taraxacum eriopodum.*

- Glacier and alpine talus (above 4500 m)

Above 4500m, the temperature and soil moisture is very low, and winds are usually strong. Plant growth is very difficult in this area. The main types of land cover are glacier and rocks or so called alpine talus.

These six zones illustrate the characteristics of the vegetation distribution as the elevation changes. However, the six zones are a qualitative description of the vegetation distribution along the altitudinal gradient. How to accurately delineate the boundaries between the vegetation zones is subject of further research. The 500 m interval is merely an approximative value, as noted above. A dominant community vegetation map is important for delineating the boundaries between the vegetation zones. The units of this vegetation map within the vegetation zones should be delineated on the basis of relationship among plant species, and on relationships between plants and their topographic setting[5]. Plant community mapping units have a relatively uniform structure and floristic composition while vegetation type mapping units have similar habitat and physiognomic characteristics. Mapping plant communities in mountain environments requires a method to recognize units of plant species that occur together. Therefore, plant community mapping units are generally compiled from field observations. Vegetation type mapping units are more suitable for mapping with remote sensing[8]. Identification of vegetation type mapping units can be attempted through a synthesis of spectral, spatial, textural, and associative characteristics [6].

2) Factors influencing vegetation distribution

Water, temperature and solar irradiation are the three driving forces to control the spatial pattern in subalpine forests at the landscape scale[28], [29]. However, much of the environmental variability controlling the spatial patterns of land cover might be attributed to topographical factors and their influence on soil-water distribution[24], [30], [31].

Research has demonstrated abundantly the influence of topography on the spatial patterns of vegetation[24], [32], [33]. Moreover, the influence of landforms on microclimate and land cover patterns is predominant in topographically complex landscapes.

In Meili Snow Mountain, the climate is controlled by a cold temperate zone on the plateau and the monsoon in the mountains. The climate is also influenced by the topographical factors of this mountain area. This results in a complex spatial pattern. In Meili Snow Mountain, the typical characteristics of the climate are:

- High solar irradiation
- Clear dry and wet seasons
- Climate changes along altitudinal gradients
- Most rainfall and snowfall occurs above 3000 m

Moreover, the distribution of the vegetation is also affected by other factors, such as stochastic, dispersal and historical processes[4].

3. Study area

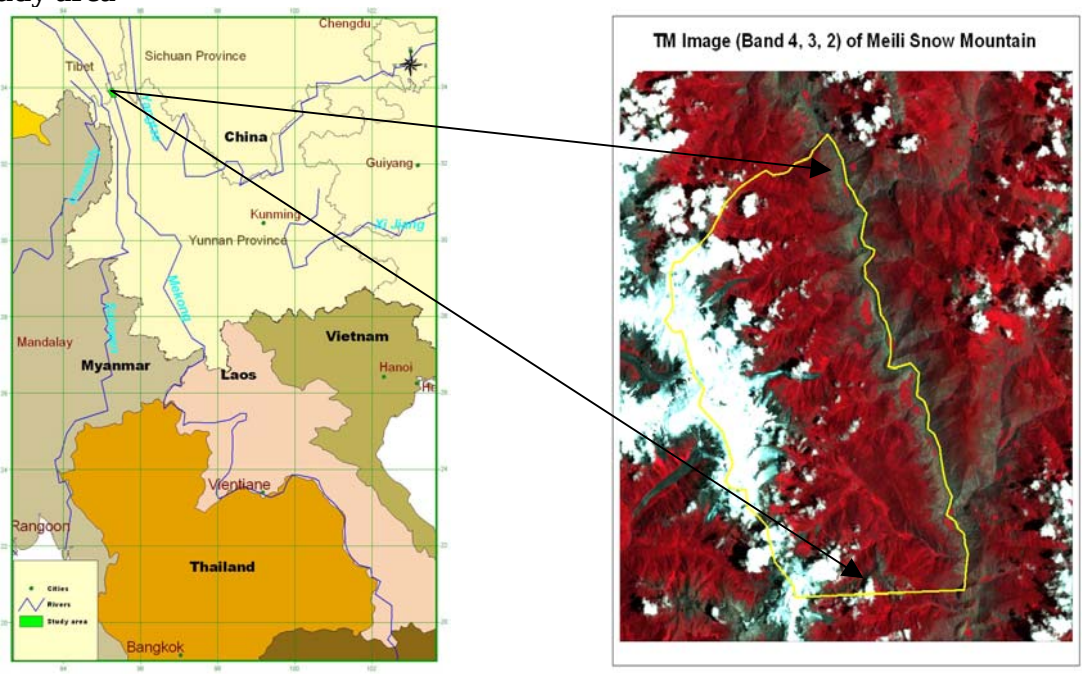

Fig. 1 The location of the Meili Snow Mountain, Yunnan Province, China

The study area is located near the border of Yunnan Province and Xizang Autonomous Region (Tibet) (28°17´—28°35´N 98°36´—98°52´E) (Fig. 1). It is a part of the Hengduan Mountain. The peak of Meili Snow Mountain (Kawagebo) is 6740 m high. It is also the highest peak of Yunnan province. Meili Snow Mountain is one of the eight holy mountains in Tibetan Buddhism[34]. Every year, many Tibetans, some of whom are from Sichuan, Gansu, Qinghai as well as from Xizang, come to Meili Snow Mountain to pray. The forest is protected very well by the local people. Meili Snow Mountain is located between Lancang River (Mekong) and Nujiang River (Salween). The mountain ridge forms the boundary of Yunnan and Xizang. The study area doesn't include the west side (Tibet). On the east side of Meili Snow Mountain, the differences in altitude range to 4700 m.

4. Material and Methods

1) Vegetation data

Three vegetation inventory field trips were undertaken in the period from May 2001 to May 2002. One hundred and eighty vegetation community sampling plots and eight hundred sixty eight GPS points were collected. The method of vegetation inventory was based on the Braun-Blanquet method[35].

The sampling plot size for forest is 10×10 m^2; for shrubs and grassland it is 5×5 m^2. The shrub layer and grass layer in tree plotting sites were surveyed in 5×5 m^2 quadrats at the corners of the plots.

In each plot, quantitative vegetation parameters (e.g. abundance, frequency and cover) were measured. Moreover, environmental characteristics of each plot were described (e.g. altitude, slope, aspect, position, soil types, rocks). Human impact, like previous harvesting, cutting, logging or grazing was also recorded. The physiognomic features were also recorded, e.g. community physiognomy (e.g. coniferous or broad-leaved, ever-green or deciduous, etc), physiognomic structures (e.g. different layers' height and cover), typical ecological phenomena and succession stages.
Analyzing the community data, the different classes of the vegetation map were defined through a synthesis of spectral, spatial, textural and associative characteristics (Huthchinson, 1982) (table 1).

2) Image data and ground truth data

- Image data

A Landsat TM image acquired on Oct. 27th, 1999, as well as the topographic map was used in this study. The topographic maps (1: 50,000) were digitized. The image, topographic map and GPS points were geocorrected and projected in the same projection system (Gauss-Kruger).

- The training data and validation data

Eight hundred sixty eight GPS points and one hundred and eighty vegetation community sampling plots were collected in the field investigation. Based on the GPS points and community sampling plots, ground truth data for each vegetation class was digitized. After ground truth digitizing, polygons were converted into a raster map with a cell size of 30m × 30m. Finally, this map was randomly divided into training data and validation data.

Table 1. The dominant vegetation classes and dominant species

Code	Classes	Dominant species
DWS	Dry and warm valley shrubs	*Sophora viciifolia*, *Elsholtzia capituligera*, *Incarvillea arguta*, *Bauhinia brachycapa*, *Vitex microphylla*
POF	*Platycladus orientalis* forest	*Platycladus orientalis*
CYS	*Corylus yunnensis* shrubs	*Corylus yunnanensis*
AG	Agriculture	*Juglans* sp., corn
OF	Oak forest	*Quercus arquifolioides*, *Quercus gilliana*
WCF	Warm coniferous forest	*Pinus densata*, *Pinus armandii* etc.
MF	Mixed forests	*Pseudotsuga forrestii*, *Pinus densata*,*Pinus armandii*, *Quercus arquifolioides*, *Quercus gilliana*, *Betula utilis.*,*Acer spp.*, *Picea likiangensis*, *Picea brachytyla*, *Sorbus* sp.,*Pentapanax* sp., *Abies georgei*, *Abies georgei var. smithii*, *Abies forrestii*
DF	Deciduous forest	*Betula utilis.*, *Acer cappadocicum.*, *Sorbus pteridophylla.*, *Pentapanax* sp.
CCF	Cold coniferous forest	*Picea likiangensis*, *Picea brachytyla*, *Abies georgei*, *Abies georgei var. smithii*, *Abies forrestii*
SM	Sub-alpine meadows	*Sambucus adnata*, *Rumex nepalensis*, *Anemone* sp.
AS	Alpine shrubs	*Rhododendron sanguineum*, *Rhododendron phaeochrysum*, *Salix annulifera*, *Salix hiriticaulis*, *Salix myrtillacea*
BL	Bare land	
RO	Rock	
CL	Cloud	
GL	Glacier	
RI	River	

The classes of the vegetation map include the five non-vegetation classes (cloud, glacier, rock, river and bare land) (table 1).

3) Topographic normalization

A band ratio was used to reduce the topographic effect. The use of ratio algorithms has been reported to partly resolve the problem of variable illumination, provided that the atmospheric path radiance term is eliminated [36]. Band ratioing is based on the principle that a certain component of reflectance in all bands is a result of angular effects. These effects are experienced as a uniform multiplier to reflectance values that are otherwise determined by real differences in earth materials. By dividing on band by another, the uniform angular component is divided away, leaving the variation which represents differences in earth materials[37]. However, variations resulting from atmospheric effects, scattering, and other anomalies may remain.
The bands chosen for this analysis were the Near Infrared and Red band. This is in part because for many earth materials such as vegetation, the variation between these bands is maximal.

In addition, many of the often significant additive effects to reflectance values (like atmospheric scattering) are experienced similarly in these two highly proximate bands. The ratio between them therefore does not produce artificial variance in the output image[37]. Other band ratios are possible. Ratios of the values of bands 5 and 4, for example, produce different ranges and variations[38], [39]. However, for the goal of controlling topographic effects (rather than highlighting various specific earth materials), infrared and red bands continue to be the most commonly employed.

4) Classification method

One of the most widely known supervised classifiers is the maximum likelihood (ML) algorithm. The ML decision rule is based on a normalized (Gaussian) estimate of the probability density function of each class. The class statistics are obtained from training data.

The ML decision rule can be modified to calculate the prior probabilities of each class[14].

5) Image stratification

In this step, the image was stratified into non-vegetation types, for example cloud, snow, rock, river and bare land, and vegetation classes. The spectral reflectances of these cover types are very different and can therefore be easily discriminated by supervised classification.

6) Creating the prior probability images

After the training sites were established, the Bayesian soft classifier was run. A set of images (one for each class) was produced expressing the posterior probabilities of belonging to each class based on the standard ML discrimination function.

The posterior probabilities were adjusted to account for error since the classification could not be assumed as perfectly accurate. As this preceeded the accuracy assessment, an accuracy level of 85% was assumed[40]. The formula used to accommodate this was

$$P(C_f) = \{[P(C_{f/x}) - 1/n]\times 0.85\} - 1/n \qquad (1)$$

where $P(C_f)$ = prior probability of a forest class

$P(C_{f/x})$ = posterior probability of belonging to a forest class

n = total number of classes

Since the posterior probabilities are images, this yields a set of spatially explicit prior probability images (i.e., the prior probability varies from one pixel to the next).

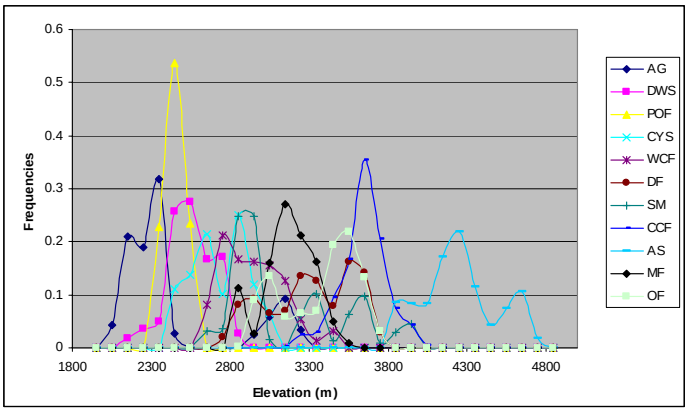

Fig. 2 The frequencies of the vegetation types along an elevation gradient (for an explanation of the symbols, see table 1)

To integrate the plant community data into the supervised classification, they were used to calculate the frequency distribution for each vegetation class. The frequencies indicate that the different vegetation classes distribution changes with elevation. It can be used to modify the prior probability of each vegetation class. As noted above, 180 community samples were recorded during the field surveys. These are used to calculate the frequency of each vegetation type in each elevation gradient (every 100m) using formula 2 (Fig. 2). Incorporating such information into the prior probability images of any class is fairly simple and requires the following steps:

$$f_{ij} = n_{ij}/180 \qquad (2)$$

Where f_{ij} = frequency of class i in altitude gradient j

n_{ij} = the number plots of class i in in altitude gradient j

The frequency of each class in each elevation gradient was used to modify the prior probability of each vegetation class.

- Assign the residual relative frequencies (residual probabilities) to the appropriate bins (elevation gradient bins).
- Add the result to the prior probability image for the class concerned.
- Divide the result by (n-1) and subtract it from the prior probability images for all other vegetation classes.

7) Accuracy assessment

A validation set was used to assess the accuracy of the classifications. One of the most common means of expressing classification accuracy is the classification error matrix (sometime called confusion matrix). Congalton and Green have described the principles and practices currently in use for assessing classification accuracy[41]. In this study, the accuracy of the classifications was assessed using confusion or error matrices and Kappa statistics (KHAT)[41].

5. Result

As noted above, the image of the study area was stratified into non-vegetation and vegetation parts (Fig. 3 and Fig. 4). The vegetation stratum was subjected to ML supervised classification with modified prior probabilities using 5 Landsat TM bands and one extra band (ratio band4/band3) (Fig.4). After the non-vegetation map (Fig. 3) was overlaid with the vegetation part (Fig. 4), the final dominant community vegetation map was obtained (Fig. 5). The accuracy assessment of this classification shows that the overall accuracy and Kappa value respectively are 87.7% and 0.861 (table 2).

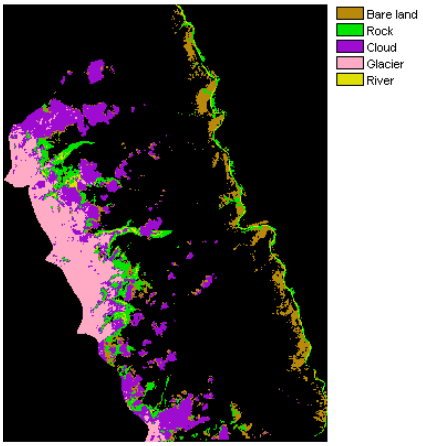

Fig. 3 Map of non-vegetation class

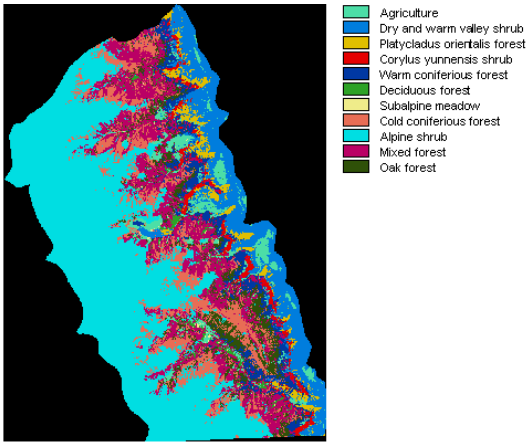

Fig. 4 Map of vegetation classes

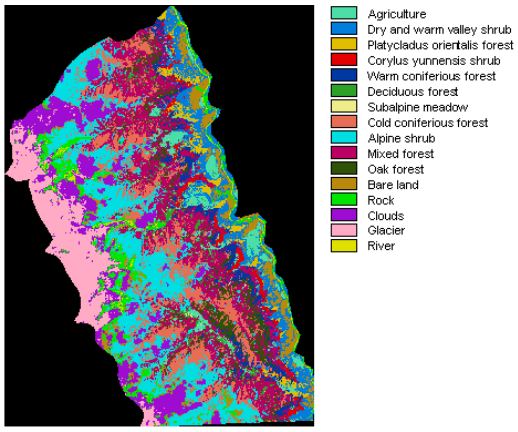

Fig. 5 The final vegetation map of maximum likelihood classification with prior probabilities

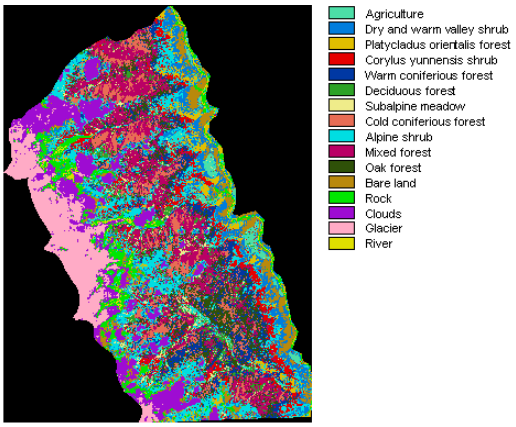

Fig. 6 The vegetation map of maximum likelihood classification with equal probabilities

When the ML classification with equal prior probabilities was performed, and when this result was overlaid with the non-vegetation part, another vegetation map was obtained (Fig. 6). However, compared with the ML classification with prior probabilities, the accuracies are lower. The overall accuracy is 82.1%. The Kappa value is 0.797 (table 3). The result of the Kappa value was improved about 8% by using ML classification with modified prior probabilities.

If the training data were used to perform ML supervised classification with equal prior probabilities and without stratification, the poorest accuracy vegetation map will be obtained. The overall accuracy is 79.3%. The Kappa value is 0.761. Even the clouds and glacier aren't classified accurately. The producer's accuracy of cloud is 22.2%. The user's accuracy of glacier is 75.9%.

Table 2. Error matrix resulting from the ML classification with prior probabilities

	DWS	WCF	DF	SM	CCF	AS	MF	OF	BL	RC	CL	GL	RI	Total	UA
DWS	548	0	0	0	0	0	0	0	0	0	0	0	0	564	97.2%
WCF	0	386	0	0	1	0	12	2	0	0	0	0	0	426	90.6%
DF	0	0	118	115	1	0	15	0	0	0	0	0	0	289	40.8%
SM	0	0	2	6	0	0	0	0	0	0	0	0	0	24	25.0%
CCF	0	12	0	0	657	5	9	34	0	0	0	0	0	717	91.6%
AS	0	0	58	46	0	523	0	0	0	0	0	0	0	627	83.4%
MF	0	37	7	16	15	2	319	37	0	0	0	0	0	519	61.5%
OF	0	42	0	0	57	0	32	126	0	0	0	0	0	257	49.0%
BL	31	0	0	0	0	0	0	0	290	0	0	0	0	321	90.3%
RC	0	0	0	0	0	0	0	0	0	289	0	0	5	294	98.3%
CL	0	0	0	0	0	0	0	0	0	0	901	131	0	1032	87.3%
GL	0	0	0	0	0	0	0	0	0	0	26	2148	0	2174	98.8%
RI	0	0	0	0	0	0	0	0	0	6	0	0	199	205	97.1%
Total	593	477	189	200	734	530	391	199	290	295	927	2279	204	8302	
PA	92.4%	80.9%	62.4%	3.0%	89.5%	98.7%	81.6%	63.3%	100.0%	98.0%	97.2%	94.3%	97.6%		

OA=87.7% Kappa=0.861 (for an explanation of the symbols, see table 1)
OA: overall accuracy; UA: user's accuracy; PA: producer's accuracy

Table 3. Error matrix resulting from the ML classification with equal probabilities

	WCF	CCF	AS	UA
WCF	376	139	3	66.6%
CCF	22	509	0	88.1%
AS	15	3	218	74.2%
PA	78.8%	69.4%	41.1%	

OA=82.1% Kappa=0.797 (for an explanation of the symbols, see table 1)
OA: overall accuracy; UA: user's accuracy; PA: producer's accuracy

6. Discussion and conclusion

The results show that the accuracy of the classification can be improved with about 8% using prior probabilities. The error matrix resulting from ML classification with prior probabilities shows that the accuracies of the five non-vegetation classes are quite good (table 2). However, some of the vegetation types have low accuracy, especially sub-alpine meadow and deciduous forest. The producer's and user's accuracies of sub-alpine meadow are very poor (3.0% and 25.0% respectively). The producer's and user's accuracies of deciduous forest respectively are 62.4% and 40.8%. From the error matrix, the confusion between these two classes is apparent.

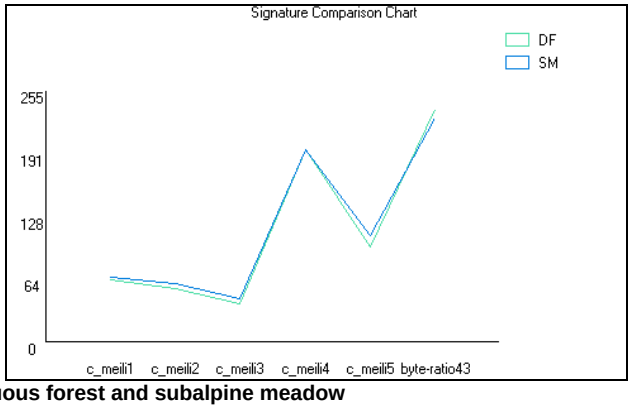

Fig. 7 The signatures of deciduous forest and subalpine meadow

The signature responses also show that the spectral characteristics of the two classes are very similar (Fig. 7). From the field survey it was noted that the natural sub-alpine meadow occupies a very small area; most of the meadow has been deforested for grazing purposes. The image of the study area was acquired in October. This is the season for deciduous trees to lose their leaves. Most of the deciduous forest occurs in meadow or shrub land, and is usually a secondary forest. Most of the deciduous forest is mixed with meadows or occurs around the meadows because these two kinds of vegetation are caused by human activities. This resulted in confusion between deciduous forest and sub-alpine meadow, and explains the poor accuracies.

From table 2, it can be concluded that except for two classes (sub-alpine meadow and deciduous forest), the accuracies of oak forest are poor. In this study area, the oak forest is often mixed with pine and fir trees. Additionally, the pure oak forest area is relatively small.

For the ML classification with equal probabilities, the producer's and user's accuracies of sub-alpine meadow also are poor: 30.0% and 23.1% respectively. The producer's and user's accuracies of deciduous forest are 45.0% and 22.9%. Comparing with the results of ML classification with prior probabilities, the producer's and user's accuracies of sub-alpine meadow are higher, but the accuracies of deciduous forest are lower. As noted above, in this area, the sub-alpine meadow serves as grazing land and the areas are very small. This implies that the distribution of the sub-alpine meadow doesn't show a strong relationship with altitude, and that the prior probability of the sub-alpine meadow is probably not correct. In contrast, the classification of vegetation classes whose spatial distributions on regional scale are mainly controlled by altitude, like fir forest (cold coniferous forest), pine forest (warm coniferous forest) and alpine shrub, improved notably.

Moreover, when the classification was performed with equal prior probabilities, the vegetation map resulted in unrealistic distributions of some of the vegetation classes. It is known that the pine forest (warm coniferous forest) and agriculture land can't be found above 3300 m. However the vegetation map displays occurrence above 4000 m (Fig. 8). Fir forest (cold coniferous forest) is absent below 3000 m, but it can be found even below 2500 m in the vegetation map (Fig. 8). The dry and warm valley shrubs and *Platycladus orientalis* forest dealt with also can be found above 3000 m (Fig. 8). In this study area, the mixed forest and oak forest can't be found in areas higher than 4000 m. However, Fig. 8 shows that they can be found above 4000 m, even above 4300 m. In Meili Snow Mountain, elevations above 4500 m, feature only alpine shrub, alpine talus and glacier. However, Fig. 8 shows occurrence of vegetation classes above 4500 m. ML classification with modified prior probabilities for each vegetation class resulted in much more realistic distributions (Fig. 9).

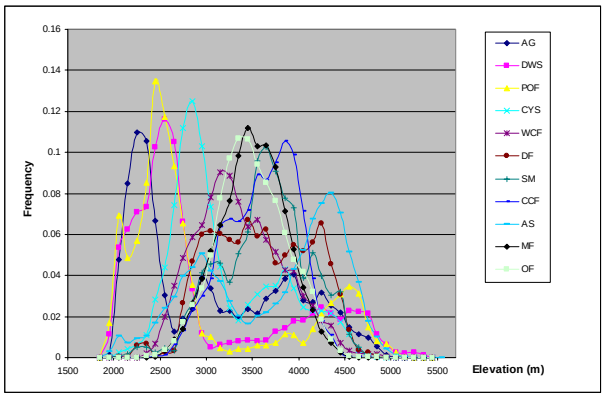

Fig. 8 The distribution of vegetation classes obtained by ML classifier with equal prior probabilities

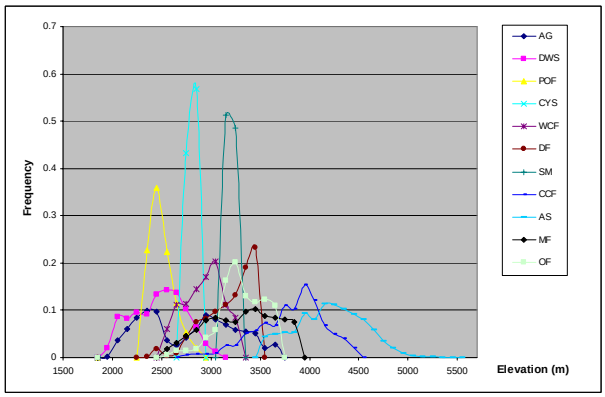

Fig. 9 The distribution of vegetation classes obtained by ML classifier with prior probabilities

Neverteless, in the vegetation map with prior probabilities, several problems are still apparent. Agriculture should not be found above 3500 m. The tall shrub (*Corylus yunnensis* shrub) should not be distributed between 2700 m and 2900 m. Fig. 9 shows that the sub-alpine meadow only can be found between 3100 m and 3300 m, but it should also occur above 3500 m. These errors were due to the small number of plant community samples for these vegetation types.

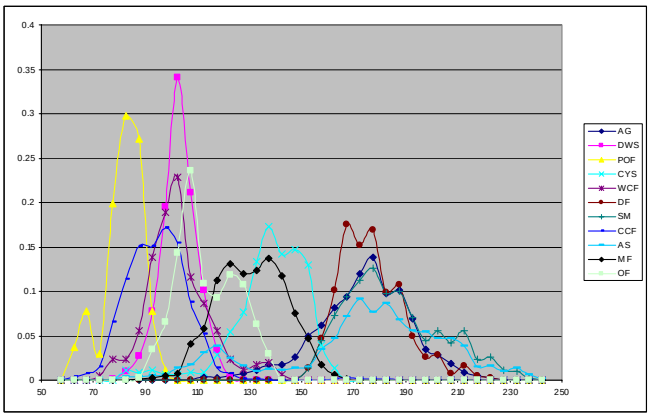

Fig. 10 The frequency distribution of training site in infrared band (band 4)

As mentioned above, the conventional ML classification depends on the assumption of multivariate Gaussian statistical distribution for the training class data to be classified[42]. However, the data in feature space may not follow the assumed model. Further, it is possible that a single class may be represented at multiple locations in feature space[43]. Consequently, statistical approaches may be regarded as restrictive because of the underlying assumption of the model. A further problem with statistical approaches is that they require non-singular (invertible) class-specific covariance matrices[44]. Fig. 10 shows the distributions of eleven vegetation classes from training data in the infrared band (band 4). It can be observed that only the distribution of the training data of dry and warm valley shrubs, warm coniferous forest and cold coniferous forest, resemble the Gaussian distribution. That is why the accuracies of the three classes are satisfactory (table 2). However, the distributions of deciduous forest, sub-alpine meadow, mixed forest and oak forest are different from the Gaussian distribution. This probably can explain why the accuracies of these classes are below expectation.

The advantages of the ML classifier have led to an enormous spread of its use for remote sensing application. This has also encouraged research towards possible improvements and extensions of the basic procedure. In this study, the results show that classification accuracy can be improved using prior probabilities. However, the estimates of probabilities lack precision because the 180 plant community samples are not sufficient. Stratifying the image and drawing out the non-vegetation types can also reduce the effect of incorrect prior probabilities and can improve the accuracy of the classification with 3.6%. Although the amount of sample plots is not sufficient and the probabilities are not entirely reliable, the probabilities still can reflect the spatial vegetation distribution patterns in this mountain area and can be used to improve the classification accuracy.

7. Conclusion

This article integrates dominant plant community data with altitude gradients into the ML classification by calculating and modifying the prior probabilities. It has been demonstrated that the classification accuracy can be improved. However, the aspect and slope also affect the vegetation distribution in mountain area. Further work will incorporate the aspect and slope effects into the image classification. The spatial distributions of several plant communities are markedly influenced by slope and aspect in the study area. In order to reduce the limitation of the number of community samples, ordination and cluster analysis are often used by plant ecologists to analyze the relationships between vegetation and environmental factors. Including these relationships into image classification can perhaps achieve interesting results. To improve the accuracies of the classes whose training data distributions do not follow the Gaussian distribution, an artificial neural networks classifier is likely to yield better results.

Acknowledgements

The authors are grateful to Dr. Wang Chongyun, Prof. Lu Shugang and Mr. Wu Yucheng for their help to collect plant community data. This work was supported by grants the auspices of the National Key Project for Basic Research on Ecosystem Changes in Longitudinal Range-Gorge Region and Transboundary Eco-security of Southwest China (2003CB415102), the Nature Conservancy (TNC) (1804357112-8010) and Vlaamse Interuniversitaire Raad (VLIR ZEIN2002PR264-886), Belgium. The staff in the TNC office in Deqin County, Yunnnan, China are gratefully acknowledged for their support for the field survey.

Reference

[1] Pfeffer K., Pebesma E. J., Burrough P. A., 2003. Mapping alpine vegetation using vegetation observation and topographic attributes. *Landscape Ecology*, 18, 759-776.

[2] Nagendra H., 2001. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22 (17), 3385-3405.

[3] Whittaker, R. H., Neiring, W.A., 1965. Vegetation of the Santa Catalina Mountains, Arizona: . A gradient analysis of the south slope, *Ecology* 46, 429-452.

[4] Dymond C.C. and Johnson E. A, 2002. Mapping vegetation spatial patterns from modeled water, temperature and solar radiation gradients. *ISPRS Journal of Photogrammetry & Remote Sensing*, 57, 69-85.

[5] Frank T. D., 1988. Mapping dominant vegetation communities in the Colorado Rocky Mountain Front Range with Landsat Thematic Mapper and digital terrain data. *Photogrammetic Engineering and Remote Sensing*, 54 (12), 1727-1734.

[6] Hutchinson C. F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetic Engineering and Remote Sensing*, 48 (1), 123-130.

[7] Schreier H., 1982. The use of digital multi-date Landsat image in terrain classification. *Photogrammetic Engineering and Remote Sensing*, 48 (1), 111-119.

[8] Frank T. D. and Isard S. A., 1986. Alpine vegetation classification using high resolution aerial imagery and topoclimatic index values. *Photogrammetric Engineering and Remote Sensing*, 52 (3), 381-388.

[9] Skidmore A. K., 1989 An expert system classifies Eucalypt forest types using Thematic Mapper data and a digital terrain model. *Photogrammetic Engineering and Remote Sensing*, 55 (10), 1449-1464.

[10] Franklin S. E. and Wilson B. A., 1992 A three-stage classifier for remote sensing of mountain environments, *Photogrammetic Engineering and Remote Sensing*, 58 (4), 449-454.

[11] Elummoh A. and Shrestha R. P., 2000. Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand. *Photogrammetic Engineering and Remote Sensing*, 66 (3), 297-304.

[12] Liu Q. J., Takamura T. and Takeuchi N., 2002. Mapping of boreal vegetation of a temperate mountain in China by multitemporal Landsat TM imagery. *International Journal of Remote Sensing*, 23 (17), 3385-3405.
[13] Mciver D. K. and Friedl M.A., 2002 Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment* 81:253-261.
[14] Strahler A. H., 1980. The use of prior probabilities in the maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment* 10:135-163.
[15] Mather P.M., 1985. A computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely-sensed data, *International Journal of Remote Sensing*, 6 (2), 369-376.
[16] Maselli F., Conese C., Petkon L. and Resti R., 1992. Inclusion of prior probabilities derived from a nonparametric process into the maximum-likelihood classifier. *Photogrammetic Engineering and Remote Sensing*, 58 (2), 201-207.
[17] Maselli F., Conese C., De Filippis T. and Romani M., 1995. Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50 (2), 2-11.
[18] Foody G. M., Campbell N. A., Trodd N. M. and Wood T. F., 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetic Engineering and Remote Sensing*, 58 (9), 1335-1341.
[19] Pedroni L., 2003. Improved classification of Landsat Thematic Mapper data using modified prior probabilities in large and complex landscapes. *International Journal of Remote Sensing*, 24 (1), 91-113.
[20] Paradella W. R., Silva DA M. F. F., Rosa DE A., Kushigbor C. A., 1994. A geobotanical approach to the tropical rain forest environment of the Carajás Mineral Province (Amazon Region, Brazil), based on digital TM-Landsat and DEM data. *Int. J. Remote Sensing*, 15 (8), 1633-1648.
[21] Fahsi A., Tsegaye T., Tadesse W. C. T., 2000. Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy, *Forest Ecology and Management*. 128, 57-64.
[22] Dorren L. K.A., Maier B., Sejimonsbergen A. C., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification, *Forest Ecology and Management*. 183, 31-46.
[23] Thomas V., Treitz P., Dennis J., Miller J., Lafleur P. and McCaughey H. J., 2002. Image classification of a northern peatland complex using spectral and plant community data, *Remote Sensing of Environment.* 84, 83-89.
[24] Cantón, Y., Barrio, G. Del., Solé-Benet, A. and Lázaro R., 2004. Topographic controls on the spatial distribution of ground cover in the Tabernas badlands of SE Spain, *Catena*, 55, 341-365.
[25] Parker, A. J., 1995. comparative gradient structure and forest cover types in Lassen Volcanic and Yosemite National Parks, California, *Bulletin of the Torrey Botanical Society* 122, 58-68.
[26] Pinder, J. E., Kroh, G. C., White, J. D., Basham May, A. M., 1997. The relationship between vegetation types and topography in Lassen Volcanic National Park, *Plant Ecology* 131, 17-29.
[27] Qiu Lianqing, 1958. The Vegetation Vertical Distribution Regularity in Simang Snow Mountains. Yunnan University, Institute of Ecology and Geobotany press (in Chinese).
[28] Wu Zhengyi, Jiang Hanqiao, Jin Zhenzhou et.al, 1982 Yunnan Vegetation, Yunnan Science Technology Pubish (in Chinese).
[29] Larcher, W., 1995. Physiological Plant Ecology, 3rd ed. Springer-Verlag, New York.
[30] Del Barrio, G., Alvera, B., Puigdefabregas, J., Diez, C., 1997. Response of high mountain landscape to topographic variables: Central Pyrenees, *Landscape Ecology* 12, 95-115.
[31] Forman R. T. T. and M. Godron, 1986. Landscape Ecology, John Wiley and Sons, New York, N. Y..
[32] Crave, E., Gascuel-Odoux, C., 1997. The influence of topography on time and space distribution of soil surface water content. *Hydrological Processes* 11, 203-210.
[33] Western, A.W., Grayson, R.B., Blöschl, G., Willgoose, G.R., McMahon, T.A., 1999. Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resources Research* 35, 797–810.
[34] Fang Zhengdong, (1997). Lauding Kawagebo, Yunnan Art press (in Chinese).
[35] Braun-Blanquet, J., 1932. Plant Sociology: the study of plant communities, (Translated by G. D. F. Fuller and H. S. Conrad), McGraw-Hill Book Co., Inc. New York.
[36] Kowalik, W. S., Lyon R. J. P. and Switzer P., 1983. The effects of additive radiance terms on ratios of Landsat data, *Photogrammetric Engineering & Remote Sensing*, 49 (5), 659-669.
[37] Holben B. N.,Justice C. O.,1980. The topographic normalization effect on spectral response from nadir-pointing sensors, *Photogrammetric Engineering & Remote Sensing*, 46 (9), 1191-1200.
[38] Ekstrand S., 1996. Landsat TM-Based forest damage assessment: correction for topographic effects, *Photogrammetric Engineering & Remote Sensing*, 62 (2), 151-161.
[39] Hale S. R. and Rock B. N., 2003. Impact of topographic normalization on land-cover classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 55 (9), 1303-1309.
[40] Eastman J. R. and Sangermano F., 2005. Application of Spatial Priors in the Maximum Likelihood Classification of Tropical Dry Forest Classes, http://www.clarklabs.org/Summer2005.
[41] Congalton R. G. and Green K, 1999. Assessing the Accuracy of Remotely Sensed Data: *Principles and Practices*. Lewis Publisher, Boca Raton, FL.
[42] Paola J. D. and Schowengerdt R. A., 1995. A detailed comparison of backpropagation neural network and maximum likelihood classifiers for urban land use classification, *IEEE Transactions on Geoscience and Remote Sensing*, 33 (4), 981-996.
[43] Atkinson P. M. and Tatnall A. R. L., 1997. neural networks in remote sensing, *International Journal of Remote Sensing*, 18 (4), 699-709.
[44] Benediktsson J. A., Swain P. H. Ersoy O. K., 1990. Neural network approaches versus statistical methods in classification of mutisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 28 (4), 540-551.