INTEGRATION OF ANCILLARY INFORMATION INTO OBJECT-BASED CLASSIFICATION FOR DETECTION OF FOREST STRUCTURES AND HABITATS

M. Förster [a, *], B. Kleinschmit[a]

[a] Berlin University of Technology, Department of Landscape Architecture and Environmental Planning, Str. d. 17. Juni 145 (EB 5), 10623 Berlin, Germany; michael.foerster@tu-berlin.de, birgit.kleinschmit@tu-berlin.de

WG IV/4 and WG VIII/11

KEY WORDS: ancillary information, forest structure, forest habitats, QuickBird, object-based classification, fuzzy logic

ABSTRACT:

The detection of forest types and structure parameters is of major importance for the design of forest inventories, for the application of forest management practices as well as for the monitoring of biodiversity in the context of the NATURA 2000 network. For these purposes the use of additional information about the natural behaviour of forest structure within classification processes of satellite data is widely known. Especially parameters concerning potential natural forest locations, such as elevation, aspect, precipitation, wetness or soil acidity, were taken into account. Although natural site conditions strongly influence the forest types and structure, the presented results will additionally improve the classification results by integrating silvicultural knowledge into the classification process. The study was carried out using QuickBird data at test sites, which are located in the pre-alpine area in Bavaria (Southern Germany). Within the test sites, different semi-natural mixed forest types exist. First results of the presented approach show higher classification accuracy than can be reached without usage of additional data. It is recognisable that higher classification accuracy depends on the kind of ancillary data. While the effects of the local variability of total height and the aspect are very limited in pre-alpine areas, additional soil-data or information of the forestry site map in combination with fuzzy-based rules can significantly improve classification results. In contrast to improved results with ancillary soil data, silvicultural information tends to have less influence on the classification quality. Additionally, for habitats and species with very distinctly defined ecological niches (e. g. alluvial types of forest) a better definition and integration of rules is possible than for habitats with very broad ecological ranges.

1. INTRODUCTION

With the development of a standardised and pan-European available geodata-infrastructure (Craglia et al., 2005), remote sensing applications which integrate available GIS information will attain a higher importance. Therefore, various (mostly methodological) studies of integrating additional data and knowledge into classification processes (Maselli et al., 1995; Stolz, 1998) were undertaken. However, with the availability of very high spatial resolution (VHSR) satellites, such as QuickBird and IKONOS, the challenge of combining higher data amounts from remote sensing data and GIS data for the purpose of gaining the most valuable knowledge about the landscape is still given.

As a contribution to a better understanding of the integration of different data sets the presented approach uses a rule-based fuzzy logic classifier that combines spectral and textural information of a QuickBird scene with ancillary data-layers and a knowledge base for the identification of forest structures and habitats. This example is especially suitable to show chances and challenges of data-integration techniques, because a long-term information about silvicultural practices and ecological woodland development is available together with a good (geo)database.

1.1 Types of Additional Information

Additional information can be differentiated into two sections. On the one hand, geo-data in measured or computer-generated form is available. For forestry applications a broad range of techniques has been adopted, namely:

- simulation of data (Hagner and Olofson, 2004; Verbeke et al., 2005)
- usage of height information, especially with LIDAR techniques (Diedershagen et al., 2004)
- integration of silvicultural maps (Förster et al., 2005b) as well as soil and hydrology maps

On the other hand, knowledge about processes of the forested landscapes is abundantly available and recorded. Of high relevance is information about:

- land-use history
- silvicultural practices (Pretzsch, 2002)
- potential natural vegetation (Walentowski et al., 2004)

Therefore, the task is to integrate only ancillary data into the classification process, which is a decisive factor for the land use and has a spatially distinguishable component.

2. DATA AND METHODS

For the presented approach the satellite data were delineated in a multi-scale segmentation process (Burnett and Blaschke, 2003). This task was performed in an object-oriented approach using the software eCognition (Benz et al., 2004). The segmentation levels of different resolution were delineated and assigned to hierarchical organized groups of objects, such as

* Corresponding author

forest habitats, crown combinations and crown types of single tree species.

The segments were then classified with and without ancillary information and the results subsequently compared. Additional sources of information are combined using a fuzzy knowledge base (Stolz and Mauser, 1996). Since expert knowledge about the test area is available as verbal description, which often contains cognitive uncertainties and is imprecise, fuzzy logic represents a possibility to express these vague statements in a mathematical framework as a degree of membership to a fuzzy set (Zadeh, 1983).

The results with usage of additional information were applied to derive NATURA 2000 forest habitat types and qualities with conventions developed for implementing the habitats directive in Germany (Burkhardt et al., 2004).

2.1 Satellite Data

For the presented investigation QuickBird data were used. The QuickBird sensor is the first commercial satellite that provides submeter resolution. Its panchromatic band collects data with a 60 cm resolution at nadir while the multispectral (visible and near infrared) ground sampling distance is 2.4 m at nadir.

In summer 2005 data were acquired from the forested NATURA 2000 site “Angelberger Forst” in the pre-alpine area of Bavaria, Germany, which covers approximately 650 ha. Within this NATURA 2000 site, different semi-natural mixed forest types exist, including Beech forests (9110, 9130) and Alluvial forests with Alnus and Fraxinus (91E0). The scene was acquired at the 11.08.2005 and had a cloud coverage of 10 % and an off-nadir angle of 11.3 degree.

2.2 Additional Data

As in Germany commonly available geo-data, a digital terrain model (DTM 5 and DTM 25), a conceptual soil map (1 : 25.000) as well as a silvicultural site maps were used. The knowledge-base to built up rule-sets for potential forest types were available from a previous project in cooperation with the Bavarian State Institute of Forestry (Kleinschmit et al., 2006). These rules were complemented by silvicultural rules attained from local forest rangers and silvicultural literature (Walentowski et al., 2004).

2.3 Segmentation and Classification

After a geometric correction and the pan-sharpening of the original data to a merged resolution of 0.6 m (Zhang, 2002), the scenes were segmented at three landscape scales (see figure 1). These levels were named as single tree / small tree group level (Scale Parameter (SP) 15, shape factor 0.1, compactness 0.5), as tree group patch level (SP 40, shape factor 0.1, compactness 0.5) and as combined patch level structure (SP 150, shape factor 0.1, compactness 0.5). All derived segments depend on the sensor specification, pan-sharpening algorithm, tree types, and silvicultural practices. Therefore scale parameters for segmentation have to be adapted to scene specifications and desired results.

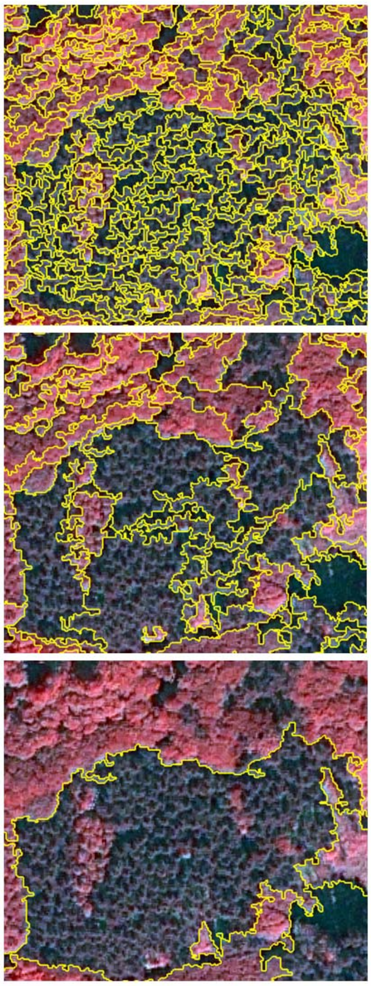

Figure 1. Example of scale parameters for different landscape levels of a forested: single tree / small tree group level of SP 15 (upper figure), tree group patch level of 40 (middle figure) and combined patch level of SP 150 (lower figure)

In advance of the forest classification non-forest land uses, such as agriculture or urban area were masked, based on thresholds for shape, texture and spectral mean value of these classes. The classification of forest types was performed on single tree / small tree group level as nearest neighbour classification of the mean spectral values of the segments. The training areas for the process were taken from field work, silvicultural maps and aerial photographs. The results of level 1 were aggregated to tree-group patch level, where a threshold of 70 per cent had to be achieved to be assigned to a single species. Mixed stands were assigned to a new introduced group “Mixed deciduous” and “Mixed”. The third level was used to improve the classifications of the sub-levels (see 2.4.2) and to derive potential NATURA 2000 habitat types (see 3.2). Shadowed areas were separately masked and classified, using the NDVI and relations to neighbour objects (“border to” and “distance to”).

2.4 Integration of Class Rules via Fuzzy Logic

The probability of assignment for each object (values from 0 to 1) is combined with a fuzzy knowledge base, which consists of silvicultural and natural information about the possibility of existence of tree species and habitats and geo-factors for a GIS-database.

A fuzzy set for each class concerning each geo-factor is defined, containing membership functions. For each parameter a set of possible verbal descriptions (linguistic terms) such as "very steep" or "flat" for the variable "slope" have to be defined and formalized by fuzzy membership functions. Furthermore, fuzzy rules need to be developed describing the relationship between each linguistic term of each linguistic variable and the degree of possibility of each class. As result of this process, defuzzicated membership function tables are derived for each geo-factor.

In combining the fuzzy sets and the hierarchical classification results the approach uses the minimum (AND-) rule, which specifies that the most unacceptable factor is the critical value for the forest type to occur. In a next step the minimum possibility of each possible class will be compared. The class with the highest membership will be assigned to the object (maximum – OR – rule, see figure 2).

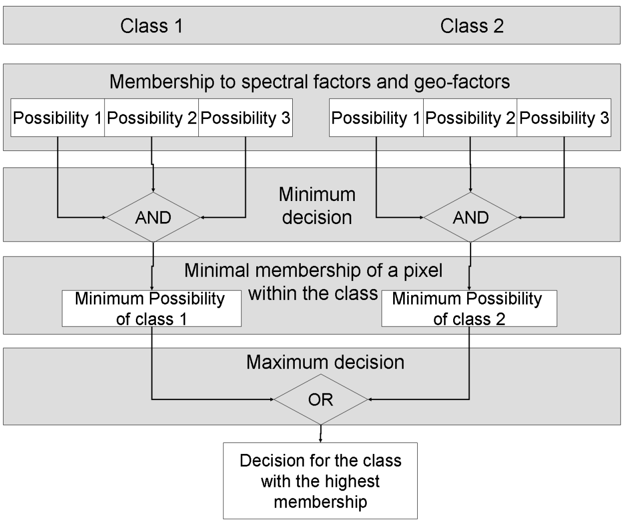

Figure 2. Schematic application of a fuzzy class decision with spectral classification and geo-factor possibilities

2.4.1 Natural Site-Conditions:

For habitat types which can possibly exist in this natural woodland composition a register of location factors developed, consisting of slope, aspect, curvature, and height of a medium resolution DEM, soil type from a conceptual soil map, and available water, soil substrate, and availability of nutrients from a forestry site map. Therefore, for all woodland species of this specific climatic region, an index of location factors was developed based on knowledge about Bavarian woodland types (Walentowski et al., 2004; Walentowski et al., 2005).

Especially for this kind of information the integration via fuzzy logic is useful, because there are no sharp thresholds. The expressions of local experts and literature sources are merely in linguistic terms, such as: "Sycamore can be found in higher pre-alpine regions, especially at very steep slopes". The integration of ancillary geo-data for natural site conditions is exemplarily shown in figure 3 for one object and the site-condition "mean slope equals 10.5 %".

As figure 3 indicates, the species with very distinctly defined ecological niches, which cannot be distinguished by spectral values are better recognisable with ancillary data.

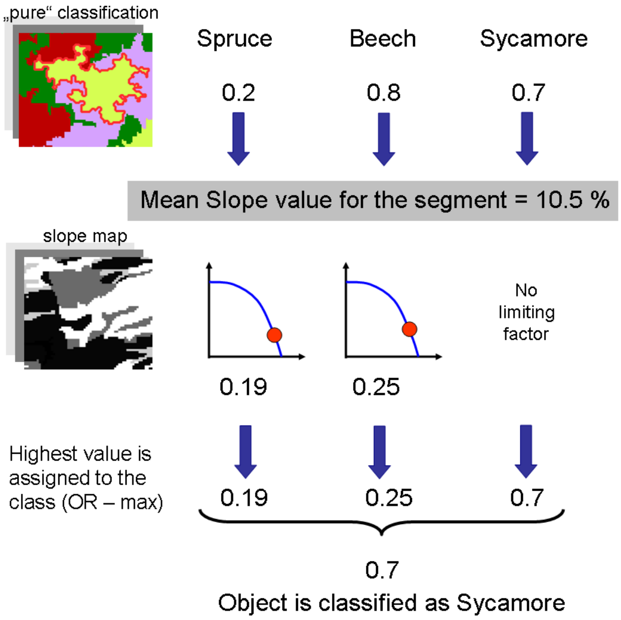

Figure 3. Example of fuzzy classification with natural site-condition geo-factors. Classification without ancillary data would be classified as "Beech" (possibility of 0.8), while with the classification with natural site-conditions (Mean Slope of the object is 10.5 %), the segment is assigned to "Sycamore".

2.4.2 Silvicultural Site-Conditions:

To include forestry practices, two approaches were used. Firstly, silvicultural preferred mixture types in Germany were taken from literature (Jansen et al., 2002) and extracted from silvicultural maps. Statistics of classified tree-type composition were taken at segmentation level 3 (combined patch level structure). If a classified mixture type was similar to a silvicultural preferred mixture type, the sub-level1 included these possibilities for the tree species of the mixture as one ancillary layer in a second classification loop.

Another approach was undertaken to improve the classification accuracy of elder spruce (from 120 years) stands. In classification level 1 small clearances were classified with the nearest neighbour approach. The existence of clearances was used in level 2. If a certain (fuzzy defined) amount of clearance and old spruce was detected in the sub-object, the possibility to assign the class to "old spruce". Vice versa, if "old spruce" is detected in level 2, the possibility of clearances in level 1 rises.

3. RESULTS

The results of the classification processes with and without ancillary data were compared to test samples, taken from silvicultural maps, field work and aerial photographs.

3.1 Classification of Tree-Types

The results of the accuracy assessment are seen in table 1. A significantly higher classification accuracy can better be reached with instead of without usage of additional data. Especially the detection of species with small ecological niches is improved. With pure classification a tree type such as Black Alder is not distinguishable spectrally from other deciduous forest while showing the highest classification accuracy with ancillary data. This is especially due to the influence of the natural site conditions, especially the geo-data and rules for the available water from the forestry site map and the curvature derived by the DEM. Other decisive factors can be the substrate (Larch) and the slope (Sycamore).

Forest Type	Ancillary Data		Pure Classification	
	Level 1	Level 2	Level 1	Level 2
Beech	0.81	0.68	0.75	0.76
Beech – young	0.32	0.14	0.15	0.14
Spruce	0.74	0.97	0.74	0.81
Spruce – old	0.42	0.65	0.32	0.61
Black Alder	0.98	0.89	0.17	0.69
Afforestation	0.97	0.85	0.95	0.82
Larch	0.96	0.68	0.59	0.68
Sycamore	0.88	0.72	0.68	0.70
Overall Accuracy	**0.77**	**0.75**	**0.64**	**0.70**

Table 1. Accuracy assessment for tree-type species (level 1 and level 2).

A further improvement could be made without differentiation of age levels (e.g. beech – young). Natural site-conditions are not useful for separation of the same species. Therefore, rule sets for silvicultural site-conditions and age classes, similar to the usage of clearances, could be useful. Between level 1 and level 2 no advance in classification accuracy is visible. This is probably due to the introduction of mixture classes. These classes were not assessed, because the process of taking samples was carried out in level 1. However, with the evaluation of these classes, a better classification result can be expected. Nevertheless, the overall accuracy alone is certainly not sufficient for a reliable tree-type classification. A further improvement is possible with careful analysis of the dependency of accuracy to geo-data types, integration of other additional data (such as LIDAR data), and a more efficient usage of silvicultural site-conditions.

3.2 Derivation of NATURA 2000 habitat types

For the classification levels 2 and 3 the attempt to obtain NATURA 2000 habitat types and their qualities was undertaken. At the moment, this information is manually mapped and combined to forestry management plans. For mapping forestry habitat types, objective mapping guides with defined rules are available in Germany (Burkhardt et al., 2004). Within these rules parameters of habitat structures, such as number of forest development phases, number of biotope trees per ha, number of dead wood per ha, or percentage of typical tree types are available for different habitat qualities. For the percentage of typical tree types the habitats were identified for:

- excellent quality (A) ≥ 90 % typical tree types,
- good quality (B) ≥ 80 % typical tree types, and
- medium quality (C) ≥ 80 % typical tree types.

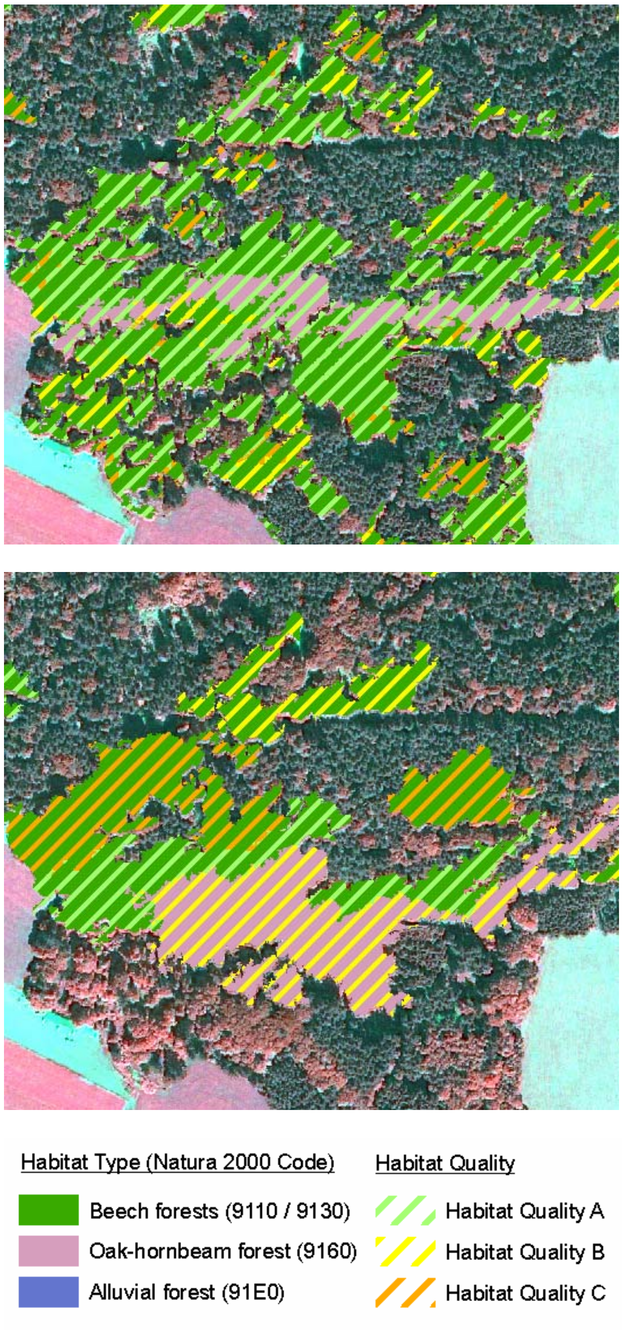

Figure 4. Derivation of different habitat type qualities for level 2 (upper figure) and level 3 (lower figure).

Figure 4 shows the result of this approach. Apart from the fact that it is not possible to derive habitat types automatically, due to parameters which cannot be detected by remote sensing, local knowledge of mapping, and political decisions (Förster et al., 2005a), the so-called orchard problem (Lang and Langanke, 2006) arises. Because of different segmentation scales, altered mixtures of areas and qualities for the habitats are available. Smaller scale parameters tend to define more fragmented areas of a good and excellent quality, while with a scale parameter of 150 larger and coherent areas of good and medium quality will occur.

As can be seen, both results and segmentation levels have certainly advantages and disadvantages, but the approach shows that the mapping guide for German forest lacks the consideration of different landscape levels. An analysis of the correlation between terrestrial mapped habitat types and different landscape levels could be helpful to formulate a more exact definition of habitat type quality.

4. DISCUSSION AND OUTLOOK

The presented investigation shows that for some forest types the classification accuracy can be higher with ancillary information integrated by fuzzy logic. It is indicated that natural site-conditions are more relevant for the classification success. However, further improvements can be made by analysing which kind of ancillary geo-information is most effectual in classification enhancement. Moreover, a comparison of different techniques of integrating geo-data into classifications, such as neural networks or multi-agent modelling, could be useful for a quality assessment of integration techniques. As the presented study illustrates that with a combined very high resolution remote sensing and fuzzy logic approach ancillary data can be successfully included in a multi-scale segmentation process, these classification results have to be extended to woodland types and species of other regions, such as north-east Germany.

One of the most difficult problems is the handling of different landscape scales with VHSR-data. For different reasons, the scale levels are helpful, e. g. for the improvement of the classification quality via sub-level and super level (see 2.4.2). However, the results vary with an altering scale. Not always supplies the finest scale and classification the adequate information. This problem arises especially with the approach of defining NATURA 2000 habitat types and qualities (see 3.2). From a classification based point of view it cannot be said which landscape level (in this example level 2 and level 3) shows a better picture of the quality of biodiversity. It is more a question of defining upper and lower biotope area sizes of habitat types for the authorities of environmental policy. Moreover, this question has to be defined by experts of plant sociology, because for each habitat type it is necessary to define whether coherent large areas have to be covered by a species or small sized habitats of a good quality are required. Another approach could again be found in the utilisation of available GIS-data. The polygons of an existent biotope map are certainly useful for biotope quality assessments or monitoring purposes (Frick et al., 2005).

However, the greatest uncertainty in the classification process is still due to the fact, that the remote sensing based parameters (as texture, spectral value, and object shape) for forest types have very broad ranges of occurrence. These parameters have mixtures, depending on, for instance distances between tree crowns, which have to be investigated more carefully in the future. Hence, including of silvicultural information could be more effective when taken from the stand level (preferred forest type) to behaviour of single tree level, for instance typical shape or texture of developing tree groups. Simulated data with typical tree growth models, such as SILVA (Pretzsch et al., 2002) could be a starting point to understand and integrate these phenomena.

Performing a classification using additional GIS-data provokes the question for consistent availability of these data. Within the forests of Germany a very good data basis already exists, especially with the information from the forestry-site map. However, the coverage and quality of geo-data will rise. Therefore, the development of integrating techniques of these data into classification processes is essential.

References

Benz, U., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, pp. 239-258.

Burkhardt, R., Robisch, F. and Schröder, E., 2004. Umsetzung der FFH-Richtlinie im Wald - Gemeinsame bundesweite Empfehlungen der Länderarbeitsgemeinschaft Naturschutz (LANA) und der Forstchefkonferenz (FCK). *Natur und Landschaft*, 79(7), pp. 316-323.

Burnett, C. and Blaschke, T., 2003. A multi-scale segmentation / object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 168, pp. 233-249.

Craglia, M., Fullerton, K. and Annoni, A., 2005. INSPIRE: an Example of Participative Policy-making in Europe Openness, Participation and Transparency. *Geoinformatics*(9), pp. 43-47.

Diedershagen, O., Koch, B. and Winacker, H., 2004. Automatic Segment and Characterisation of Forest Stand Parameters using Airborne LIDAR data, Multispectral and Fogis Data, Proceedings of the ISPRS working group VIII/2, Freiburg, pp. 208-212.

Förster, M., Kleinschmit, B. and Walentowski, H., 2005a. Comparison of three modelling approaches of potential natural forest habitats in Bavaria, Germany. *Waldökologie Online*(2), pp. 126-135.

Förster, M., Kleinschmit, B. and Walentowski, H., 2005b. Monitoring NATURA 2000 forest habitats in Bavaria by the use of ASTER, SPOT5 and GIS data – an integrated approach. In: H. Olsson (Editor), ForestSat. Swedish National Board of Forestry, Borås, Sweden, pp. 21-25.

Frick, A., Weyer, G., Kenneweg, H. and Kleinschmit, B., 2005. Knowledge-based approach to vegetation Monitoring with Quickbird imagery, ISPRS Workshop 2005 - High- Resolution Earth Imaging for Geospatial Information, Hannover, pp. 1-8.

Hagner, O. and Olofson, K., 2004. A high resolution geometric-optical forest model for development of single tree direction algorithms. In: C. Kleinn, J. Nieschulze and B. Sloboda (Editors), GGRS. Sauerlaenders, Göttingen, pp. 43-52.

Jansen, M., Schulz, R., Konitzer, A. and Sloboda, B., 2002. GIS based investigations of effects of the LÖWE program in the Harz mountains. In: M. Jansen, M. Judas and J. Saborowski (Editors), *Spatial Modeling in Forest Ecology and Management*. Springer, Berlin, pp. 177-193.

Kleinschmit, B. et al., 2006. Erfassung von Wald-Lebensraumtypen in FFH-Gebieten - Fernerkundung am

Taubenberg und im Angelberger Forst. *LWF Wissen*, 51, pp. 1-39.

Lang, S. and Langanke, T., 2006. Object-based mapping and object-relationship modeling for land use classes and habitats. *PFG*(1), pp. 5-18.

Maselli, F., Conese, C., Filippis, T. and Romani, M., 1995. Integration of ancilliary data into a maximum likelihood classifier with nonparametric priors. *Journal of Photogrammetry and Remote Sensing*, 50(2), pp. 2-11.

Pretzsch, H., 2002. *Grundlagen der Waldwachstumsforschung*. Parey, Berlin, 414 pp.

Pretzsch, H., Biber, P. and Dursky, J., 2002. The single tree-based stand simulator SILVA: construction, application and evaluation. *Forest Ecology and Management*, 162, pp. 3-21.

Stolz, R., 1998. Die Verwendung der Fuzzy Logic Theorie zur wissensbasierten Klassifikation von Fernerkundungsdaten. *Münchner Geographische Abhandlungen*, B 26, pp. 177.

Stolz, R. and Mauser, W., 1996. A fuzzy approach for improving landcover classification by integrating remote sensing and GIS data. In: E. Parlow (Editor), *Progress in Environmental Remote Sensing Research and Applications*. Balkema, Rotterdam, pp. 33-41.

Verbeke, L., Van Coillie, F. and De Wulf, R., 2005. A directional variant of the local maximum filter for stand density estimation from IKONOS imagery. In: H. Olsson (Editor), ForestSat. Swedish National Board of Forestry, Boras, Sweden, pp. 49-53.

Walentowski, H., Ewald, J., Fischer, A., Kölling, C. and Türk, W., 2004. *Handbuch der natürlichen Waldgesellschaften Bayerns*. Geobotanica, Freising, 441 pp.

Walentowski, H., Fischer, M. and Seitz, R., 2005. Fir-dominated forests in Bavaria. *waldökologie online*(2), pp. 18-39.

Zadeh, L.A., 1983. The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems. *Fuzzy Sets and Systems*, 11, pp. 199-227.

Zhang, Y., 2002. Problems in the fusion of commercial high-resolution satellite images as well as Landsat 7 images and initial solutions. *Int. Arch. Photogram. Remote Sens.*, 34(4).