# Integration of Geo-Information in Classification Processes of Satellite Imagery for NATURA 2000 Monitoring

vorgelegt von
Diplom-Geoökologe
Michael Förster
aus Potsdam

Von der Fakultät VI – Planen Bauen Umwelt
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
Dr.-Ing.
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Volkmar Hartje
Berichterin: Prof. Dr. Birgit Kleinschmit
Berichter: Prof. Dr. Hartmut Kenneweg

Tag der wissenschaftlichen Aussprache: 20.11.2008

Berlin 2008

D 83

# Summary

The effective and meaningful usage of already available geo-data will be of major importance for satellite-based classification processes. Remote sensing is not longer just data-provider but is one part of a spatial data-basis which is increasingly advancing in terms of homogeneity, actuality, and accessibility. Especially when monitoring processes are forced by policy directives, provided data-sources from other national or European initiatives should be included efficiently. The monitoring of the NATURA 2000 guideline - an EU-wide network of nature protection areas – is one of the first approaches of environmental planning which requires a joint usage of up to date remote sensing data and geo-information.

The objective of this thesis is to develop and evaluate methods of integrating existing geo-data and ancillary information into a knowledge-based classification process of very high resolution satellite imagery in the context of biodiversity monitoring. Moreover, for investigating the success of the integration process it is important to analyse the significance of the geo-data and the knowledge-base for the classification accuracy. For this purpose, different types of integration techniques in combination with satellite data of the sensors QuickBird, Spot5 and Aster were applied to two forested NATURA 2000 areas in Bavaria, Southern Germany.

With applying different integration methods it is shown that a support of monitoring tasks in the context of NATURA 2000 is possible. Therefore, depending on pre-defined scales of the Habitats Directive two techniques were applied in this thesis. A modelling technique and an integrated classification used different types of satellite data and aimed at different biodiversity-indicators. Furthermore, the thesis demonstrates that the classification accuracy of the investigated forest types can be enhanced by including ancillary data. The classification precision is influenced by the type of geo-data and the applied integration technique. Moreover, two methods of evaluating the significance of multiple classification accuracies were presented in this work. Especially the Significance Analysis of Microarrays (SAM) can identify significant influences of different types of additional data. This contributes to a better evaluation of classification results.

The thesis concludes with overall recommendations for a NATURA 2000 classification strategy as well as suggestions for an efficient combination of remote sensing data and geo-data. Future research should focus on transferability of these methods to other biogeographic regions, meeting the challenge of climate change in terms of monitoring particularly vulnerable habitats, and evaluating other techniques of integrating ancillary data.

# Zusammenfassung

Die effektive und sinnvolle Nutzung von bereits erhobenen Geodaten wird in Zukunft eine große Rolle für satellitenbezogene Klassifikationsprozesse spielen. Fernerkundung ist nicht länger nur ein Datenlieferant sondern Teil einer Datenbasis, welche sich bezüglich ihrer Homogenität, Aktualität und Verfügbarkeit zunehmend verbessert. Speziell wenn Monitoringprozesse durch Vorgaben der Politik gefordert werden, sollten Daten welche durch andere nationale oder Europäische Initiativen bereits erhoben wurden effizient einbezogen werden. Das Monitoring der NATURA 2000 Richtlinie – eines EU-weiten Netzwerkes von Schutzgebieten – ist einer der ersten Ansätze im Bereich der Umweltplanung, welcher eine gemeinsame Nutzung von aktuellen Fernerkundungsdaten und Geoinformationen benötigt.

Die Zielsetzung dieser Dissertation ist es daher, Methoden zu entwickeln und zu evaluieren, welche vorhandene Geodaten und Zusatzinformationen in einem wissensbasierten Klassifikationsprozess von räumlich sehr hoch auflösenden Satellitenbildern integrieren. Dies geschieht im Kontext des Biodiversitätsmonitoring. Um den Erfolg des Integrationsprozesses zu ermitteln ist es daher erforderlich, die Signifikanz des Einflusses einzelner Geodaten und der Wissensbasis für die Klassifikationsgenauigkeit zu analysieren. Für diese Aufgabe wurden verschiedene Arten von Integrationstechniken in Kombination mit Satellitendaten der Sensoren QuickBird, Spot5 und Aster auf zwei bewaldete NATURA 2000 Gebiete in Bayern (Süddeutschland) angewandt.

In der Arbeit wird aufgezeigt, dass unterschiedliche Integrationsmethoden die Aufgaben des Monitoring im Kontext von NATURA 2000 unterstützen können. Deshalb werden innerhalb der Dissertation auf Grundlage von vordefinierten Maßstabsebenen der FFH-Richtlinie zwei Techniken angewandt. Die vorgestellte Modellierung und die integrierte Klassifikation verwendeten verschiedene Satellitendaten um unterschiedliche Biodiversitätsindikatoren zu bestimmen. Weiterhin wird nachgewiesen, dass sich die Klassifikationsgenauigkeit der untersuchten Waldtypen durch Verwendung von Zusatzdaten verbessert. Die Exaktheit der Klassifikation ist dabei von der Art der Geodaten und der verwendeten Integrationstechnik abhängig. Diese Ergebnisse wurden mit zwei Methoden, unter Zuhilfenahme verschiedener Variationen von Klassifikationsgenauigkeiten, auf ihre Signifikanz geprüft. Speziell die Signifikanzanalyse von Mikroarrays (SAM) kann Einflüsse verschiedener Arten von Zusatzdaten identifizieren. Die Methode trägt daher zu einer besseren Evaluierung der Klassifikationsergebnisse bei.

Die Dissertation schließt mit Empfehlungen für Klassifikationsstrategien des NATURA 2000 Monitoring und gibt Hinweise für eine effektivere Kombination von Fernerkundungsdaten und Geodaten. Weitere Untersuchungen sollten sich verstärkt auf die Übertragbarkeit der Methoden auf andere biogeographische Regionen konzentrieren. Auch die Herausforderungen des Klimawandels und die Folgen für besonders gefährdete Habitate wie auch die Prüfung und Verwendung weiterer Integrationstechniken von Zusatzdaten sollte Gegenstand zukünftiger Forschung sein.

# Contents

# List of Figures

**CHAPTER IVa**



**CHAPTER IVb**



**CHAPTER V**



**CHAPTER VI**

# List of Tables

# List of Abbreviations

| Abbreviation | Expression | Description[1] |
|---|---|---|
| AF | Afforestation | Class |
| ALL | All geo-data included in a rule-based classification | If ALL is combined with a single type of geo-data or a single class named in this list, then the rules related to this parameter are all included in the classification process (defined rules). |
| AS | Aspect | Type of geo-data |
| ATKIS | Authorative Topographic Cartographic Information System | Basic topographic data-set in Germany |
| BA | Black Alder | Class |
| BE | Beech | Class |
| BERN model | Bioindication for Ecosystem Regeneration towards Natural conditions - model | Fuzzy logic-based model to derive potential natural vegetation |
| BEY | Beech-young | Class |
| CCD | Charge-Coupled Device | |
| CIR | Colour InfraRed | |
| CSM | Conceptual Soil Map | Type of geo-data |
| CUR | Curvature | Type of geo-data |
| $d(i)$ | Relative difference | Ratio of change in data expression to standard deviation in the data – in this work: measure for significance of classification accuracies |
| $d_E(i)$ | Expected relative difference | Average of the performed permutations of $d(i)$ |
| DEM | Digital Elevation Model | Type of geo-data |

[1] Expressions which are known to the scientists specialised in remote sensing and NATURA 2000 are not described in detail here.

| Abbreviation | Expression | Description |
|---|---|---|
| DTM | Digital Terrain Model | Type of geo-data |
| EON 2000 | Earth Observation for NATURA 2000 | EU funded research project (2001 – 2004) |
| FCS | Favourable Conservation Status | The overall objective of the Habitats Directive is to achieve and maintain this quality condition for all habitats and species of community interest. |
| FDR | False Discovery Rate | Percentages of variables which have no significant influence in the classification accuracy |
| FRV | Favourable Reference Values | Measure of the reference value for favourable conservation status, which takes into account trends within the reporting period. FRV include a monitoring at the scale of biogeographic regions. |
| FSM | Forest Site Map | Type of geo-data |
| FSM1 | Forest Site Map – attribute available water | Type of geo-data |
| FSM2 | Forest Site Map – attribute available nutrients | Type of geo-data |
| FSM3 | Forest Site Map – attribute soil substrate | Type of geo-data |
| GCP | Ground-Control Points | |
| GDI-DE | GeoDatenInfrastruktur Deutschland | German initiative for the development of a spatial data infrastructure |
| GSD | Ground Sampling Distance | |
| HSR | High Spatial Resolution | Sensors with a ground sampling distance of 1 – 4 m are defined as HSR (Möller 2002) |
| IHS fusion | Intensity – Hue – Saturation fusion | |
| INSPIRE | Infrastructure for Spatial Information in Europe | European Directive for the homogenisation of spatial data infrastructure |
| IMF | Input Membership Function | Expression used in fuzzy logic method to describe the input to a rule |
| KIA | Kappa Index of Accuracy | |
| LA | Larch | Class |

| Abbreviation | Expression | Description |
|---|---|---|
| NDVI | Normalised Difference Vegetation Index | |
| NO | Single types of geo-data excluded from all classes | Abbreviation used for the rule-based classification (e.g. NO-CUR for no curvature rule included) |
| OMF | Output Membership Function | Expression used in fuzzy logic method to describe the output of a rule |
| ONLY-SB | Classification is based solely in spectral values | Defined rule |
| ONLY-B1-B4 | Classification is based on single spectral bands | B1 = blue; B2 = green; B3 = red; B4 = near infrared (defined rules) |
| $P(x \mid \omega_i)$ | spectral probability of a class $\omega_i$ to be assigned to a classified object $x$ | |
| $\Pi_{sp}(\omega_i \mid x)$ | spectral possibility of a class $\omega_i$ to be assigned to a classified object $x$ | |
| PCA | Principal Components Analysis | |
| pSCI | proposed Sites of Community Importance | List of Sites of Community Importance submitted by the Member States to the EU |
| RGB | Red Green Blue | Colour composite of images |
| SAC | Special Areas of Conservation | Sites which are areas which are especially valuable for and rich in biodiversity according to the specifications of the Habitats Directive |
| SAM | Significance Analysis of Microarrays | Method to measure the significance of geo-data within a multisource classification |
| SCI | Sites of Community Importance | Proposed Sites of Community Interest approved by the European Commission. |
| SLO | Slope | type of geo-data |
| SPIN | Spatial Indicators for Nature Conservation | EU funded research project (2001 – 2004) |
| $\Delta$ | Self-defined threshold of significance | Used in SAM |
| VHSR | Very High Spatial Resolution | Sensors with a ground sampling distance of 0.5 – 1 m are defined as VHSR (Möller 2002) |

# Chapter I:
# Introduction

# 1 Loss of biodiversity – A major threat in a changing world

As rates of habitat and species destruction continue to raise, the need for conserving biodiversity, a term which will be defined and discussed in the following paragraphs, has become increasingly important. The extinction of endangered species diminishes the genetic resources needed for medical advances, secure food supply, and to ensure that the world's ecosystems can provide the necessary functions that are essential for life (UN 2000). Main drivers for the loss of biodiversity are the intensification of agriculture, urbanisation and ongoing infrastructure development, since they further reduce the effective population size of many non-volatile species. Global warming reinforces these threats, causing changes in abiotic habitat conditions such as soil water availability, and leading to shifts in phenology and distribution of plants and animals, changing habitat structures and functions as well as altering species compositions (Neubert et al. 2008a).

The term of biological diversity was introduced by evolutionary biology searching to explain the distribution of species, and why some places are especially rich in flora and fauna. Biodiversity is defined by the variability among living organisms and the ecological complexes of which they are part: this includes diversity within species, between species and of ecosystems (UNEP 1992). Often different biological processes are involved which affect a variety of scales from single species to processes between ecosystems (Hannah et al. 2002). Biodiversity consists of structural, compositional, and functional aspects at the level of genes, species, and biotopes (Zebisch et al. 2004). Approaches to quantify biodiversity during the last decades were made by conservation biology as well as landscape ecology (Lister 1998). Conservation biology has a more structural and functional perspective on biodiversity with a focus on indicators such as species richness or minimum viable populations. However, often the studies of this research field are restricted to certain ecosystems, such as forests (Jansen et al. 2002a) or wetlands (Bobbink et al. 2008), or functional groups and their interactions (Schulze and Mooney 1993). Landscape ecology is driven by a more hierarchical and scale-dependent perspective, which includes research on patch dynamic or network dynamics and fragmentation (Lang and Blaschke 2007). Because of this variety in approaches it is difficult to develop a standardised and simple measurable set of indicators which could be used to identify the decline of biodiversity (Duro et al. 2007). Moreover, contrary to other major ecological challenges as effective usage of water resources or natural hazards (e.g. flooding), there is no immediate danger arising from loss of biodiversity. Therefore, it is very difficult to estimate the quantity of the decline in species and habitats, due to the complex and dynamic nature of ecosystems. Therefore, the

monitoring of biodiversity has long been neglected or narrowed to the protection of single endangered species.

The results of quantitative biodiversity research, although rather based on estimations and modelling than on in-situ measurement, show a massive decline in species, habitats, and ecosystem variability (Dirzo and Raven 2003). Based on this scientific background, political aims for the protection of biodiversity have been defined. The most important standards for biodiversity protection on European level are the Habitats Directive (European Commission 1992) and the Birds Directive (European Commission 1979), which form the legal basis of the NATURA 2000 network. With over 26,000 protected areas covering all the Member States of the EU and a total area of about 850,000 km² divided in 11 biogeographic regions (see Figure 1) it is the largest coherent network of protected areas in the world.

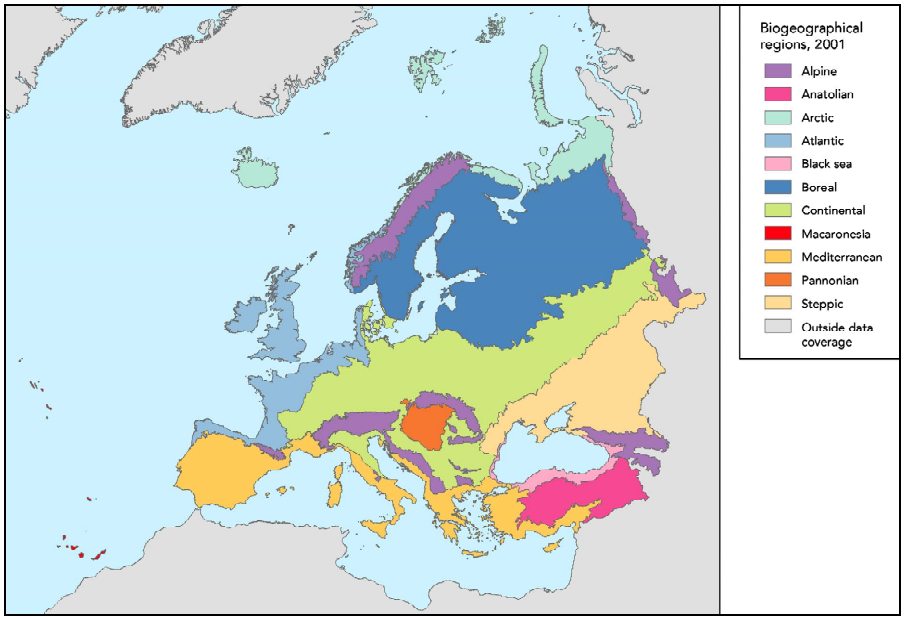



Figure 1: Biogeographic regions defined by the European Commission (source: European Environmental Agency)

The Habitats Directive defines in different annexes the types of habitats (ANNEX I) and species (ANNEX II) whose conservation requires the designation of special areas of conservation or even strict protection (ANNEX IV). An interpretation manual (European Commission 2007b) gives further recommendations how to select and manage the areas which are especially valuable for and rich in biodiversity. These sites are named Special Areas of Conservation (SAC). Each Member State submits a list of proposed Sites of Community Importance (pSCI's) for selecting appropriate areas. This list is evaluated in order to identify finalised lists of SCI's, each of which must in turn be designated as a SAC by the Member State within six years of its inclusion in the Community List. The SAC has to be evaluated by a scheme which includes three quality conditions. The purpose of

management strategies is to achieve a “Favourable Conservation Status” (FCS) instead of an “Unfavourable – Inadequate” or “Unfavourable – Bad” conservation status.

The interpretation manual provides valuable information on the species and habitats important to protect Europe’s biodiversity, but lacks a clear description of methods used for selecting and mapping proposed sites as well as a detailed definition of the FCS. Therefore national environmental agencies, such as the Federal Agency for Nature Conservation in Germany published more detailed information (Petersen et al. 2003). Only recently, with the introduction of a more specific mapping and monitoring guideline as an extension of the Habitats Directive – DocHab 04-03/03 rev.3 (European Commission 2005) the concept of the FCS was more clearly defined by the EU.

In conclusion, the complex issue of biodiversity conservation is represented by a very practical framework of protected areas with a specified variety of species and habitats. Whether the NATURA 2000 framework is an effective tool to stop the loss of biodiversity still remains a matter of discussion (Dimitrakopoulos et al. 2004).

## 2 Remote sensing and the monitoring of NATURA 2000

The legal basis of the NATURA 2000 network requires a standardised monitoring of species and habitat types and a reporting every six years. For this reason operational, economically priced and as far as possible automated techniques are required to complement or substitute cost-intensive terrestrial monitoring. A traditional method is the mapping of samples of the monitoring-site to extrapolate the biogeographical extend of species or habitats. The disadvantage of this technique is the spatial definition of sample-sites in combination with often insufficient knowledge about the extrapolation of the samples, due to inadequate knowledge of distribution pattern of species (Whittaker et al. 2005). Therefore, the combination with a top-down method, such as satellite-based classification, could overcome these difficulties. The potential of remote sensing for detecting and monitoring of NATURA 2000 habitats according to the Habitats Directive has already been proven by the EU projects Spatial Indicators for Nature Conservation – SPIN (Langanke et al. 2004) and Earth Observation for NATURA 2000 – EON 2000 (Sell et al. 2004). The rapidly developing sensor techniques and image processing methods in combination with the use of ancillary geo-data offers new opportunities to apply remote sensing for the NATURA 2000 monitoring. Especially very high spatial resolution satellites (VHSR) such as IKONOS (launched 1999) and QuickBird (launched 2001) are suitable for the monitoring of NATURA 2000 quality parameters (Frick 2007). However, with upcoming sensors of high temporal (RapidEye, launched 2008) and spectral (EnMap, proposed launch 2011) resolution

as well as RADAR satellites (e.g. TerraSAR-X), there will be further advance in the monitoring of biodiversity (Gillespie et al. 2007).

According to the monitoring guideline of the EU the observation of the NATURA 2000 network is not limited to the SAC's, because the overall situation of the biodiversity is required to be assessed and monitored. As an enhancement to the Favourable Conservation Status, the Commission introduced the concept of Favourable Reference Values (FRV), which are more closely defined by a scientific working group (European Commission 2006). In contrast to the FCS within NATURA 2000 sites, the FRV takes into account that a habitat or species could decline in range and quality over a longer time than the six-years monitoring cycle. Hence, a quality assessment seamlessly covering the biogeographic regions is required (Neukirchen 2005) in addition to the monitoring of the actual sites. Therefore, the development of a monitoring concept for quality parameters of the NATURA 2000 sites (SAC) is required as well as a method at the biogeographic level, covering larger areas. On the biogeographic level, quality parameters to be monitored are defined by the European Commission by the terms range (area has to be sufficiently large to allow the long term survival of the habitat), area (region considered minimum necessary to ensure long-term viability of a habitat type), and population (population size large enough to allow the long-term survival of the habitat). For these standardised indicators, a monitoring with remote sensing techniques is suggested (European Commission 2005).

The estimation of the spatial distribution and the type of NATURA 2000 habitats in a biogeographic region can be based on spatial modelling of potential natural vegetation. This can be based on expert knowledge about the requirements of the habitat types in respect to site specific factors, such as soil type or relief. For this task, besides the study of historical sources, a modelling of potential natural vegetation as well as a monitoring with methods of earth observation could be applied.

With monitoring tasks at the SAC-level and at the biogeographic level, the monitoring process of NATURA 2000 turns out to be a two-scale process, which probably requires the implementation of a set of different methods. The classification of highly detailed biodiversity information as well as the broad scale scanning of changes in species and habitats are very challenging tasks. Therefore, different approaches and satellite data have to be tested and evaluated in order to find suitable and effective monitoring methods.

# 3 Effective integration of geo-information into classification processes

An effective combination of remote sensing classifications with other spatial data will increase the potential of monitoring results. In this thesis the term **geo-data** is used for this geographical data-source. Geo-data are defined in this thesis as data derived with secondary data acquisition methods[2]. Geo-data consist of geometric, topologic, thematic, and temporal attributes which form a spatial object as a data-set. **Information** is defined as knowledge about the relevance of the geo-data, which generates a new value (de Lange 2002). In this work, geo-data and additional information is often combined via a rule-base. As an example one can imagine a soil map as geo-data, which is combined with the information that a specific soil type is preferred by a certain forest type as potential natural habitat. This non-spatial information should be effectively integrated into the geo-data to generate an added value, the so-called **geo-information**.

## 3.1 Standardisation of geo-data

A heterogeneous data-basis generally affects the methods used for spatial modelling. Especially in the context of NATURA 2000, many sources of data (e.g. national species inventories, data collected for other directives or site-specific data such as soil maps) could be used for monitoring purposes. In practice, they are often not standardised, only partially available or outdated.

With the development of a standardised and pan-European available geo-data-framework, introduced by the EU Directive on Infrastructure for Spatial Information in Europe – INSPIRE (European Commission 2007a), the availability and utilisation of geo-data will be significantly enhanced in respect to harmonised spatial data specifications, increased availability and searchability of metadata and measures facilitating the sharing and re-use of spatial data and services between public authorities (Cragila et al. 2005). If spatial modelling techniques are used for a standardised monitoring of NATURA 2000 sites, the basis of geo-data (e.g. soil data) has to be Europe-wide harmonised. Therefore, a more transparent approach for different available data and methods is required and could be reinforced by this European Directive (Lux 2007). INSPIRE is additionally supported by national initiatives, such as the "GeoDatenInfrastruktur – Deutschland" (GDI-DE), founded by the federal cadastral offices in Germany.

---

[2] Because of this definition remote sensing data (as primary data acquisition method) is not included in the specific definition of geo-data in this thesis.

Especially with the development of a harmonised European geo-data infrastructure there is a rising awareness of including such additional geo-information in remote sensing classification processes. Although there is a variety of promising approaches of integrating geo-data (Blaschke 2001), remote sensing is often seen as a data and information provider, which develops stand-alone maps from satellite data without realising that there is a broad variety of existing geo-data and information already available. In terms of information on environmental protection this is even more important as there is a variety of geo-data and information sources about an ecosystem, which is closely linked to required classification products. Therefore it is important to investigate the effects of the usage of additional geo-data for the classification of remote sensing data.

## 3.2 Multisource data for improving classification performance

Ancillary geo-data may be combined with remotely sensed data in a multisource classification to improve its performance (Hutchinson 1982). Especially approaches utilising Digital Elevation Models (Maselli et al. 2000), soil data (Baban and Yusof 2001), information about road density or coverage (Zhang et al. 2002), and census data (Mesev 1998) are commonly applied for land-cover or specific urban classifications.

Generally, the success of a multisource classification depends on a variety of criteria (see Figure 2). Driven by the requirements of a certain application there will be a choice of classes valuable to the user. With increasing complexity, the difficulties in differentiation of classes increase (Zubrow 2003). Therefore a more detailed classification key induces often lower classification accuracy.

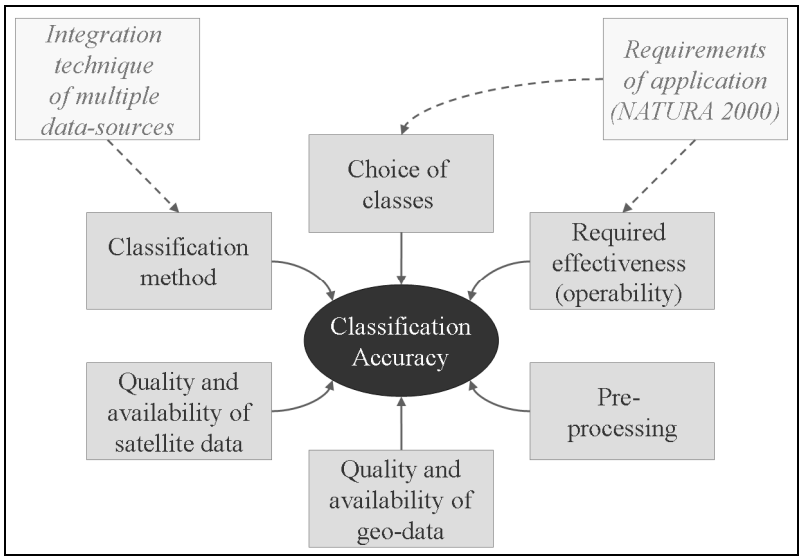



Figure 2: Criteria for classification success of an analysis based on multisource data

A similar challenge is connected with automating the process. With higher operability and less manual refining a high accuracy of the derived product can be challenging (Ehlers 2002). Moreover, the environmental and geometric quality and the availability of satellite data can influence the classification result as well as the effort which is required for the pre-processing of an image (Kleinschmit et al. 2007).

The stated points apply similarly to a classification based solely on satellite imagery. However, for multisource analyses the availability and quality of additional geo-data is crucial. Difficulties still exist in data integration due to the differences in data structures, data types, spatial resolution, and geometric characteristics (Wang and Howarth 1994) as well as temporal resolution. Most of all, the classification success depends on the methodological approach. In addition to a stand-alone analysis, the integration technique of multiple data-sources is of major importance and requires careful consideration.

Major approaches for combining various ancillary data and remote sensing imagery for image classification improvement were summarised and categorised by Lu & Weng (2007) and Hinton (1999). Ancillary data may be used to enhance image classification in three ways, through **pre-classification stratification (1.)**, **post-classification sorting (2.)**, and **internal classification (3.)**.

1. Pre-classification stratification involves the division of a scene into smaller areas based on specific criteria. This enables land-cover types that are spectrally similar to be classified independently (Harris and Ventura 1995). For example, an internal differentiation of an agricultural area into spectrally similar field crops is easier to implement when arable land is masked by means of existing geo-data. Successful approaches utilising this method were made based on topography (Bronge 1999), illumination and ecological zones (Helmer et al. 2000), and shape index of the patches (Narumalani et al. 1998).

2. Post-classification sorting allows individual pixels or objects to be refined based on decision rules derived from the ancillary data (Harris and Ventura 1995). It is easily implemented and efficient, because it can be restricted to “problem” classes. An example would be the usage of topographic information to exclude wrongly classified pixels above a defined altitude (Golden and Lackey 1992). Often this method is applied using derived remote sensing information, such as a filtering-based co-occurrence matrix (Zhang 1999) or a kernel-based reclassification (Barnsley and Barr 1996).

3. The internal processing of ancillary data during the classification process can be further differentiated. The most common groups are **parametric (a)**, **non-parametric (b)**, and **knowledge-based or rule-based (c)** classifier.

a) Parametric classifications use the additional information as a separate channel in the feature space (Spooner 1991) or as a Bayesian *a priori* probability of classes based on their estimated areal composition (Strahler 1980). However, these approaches are likely to violate the assumption of Gaussian distribution used for a conventional statistical classifier, such as Maximum Likelihood and require a careful verifying of these presumptions (Benediktsson et al. 2007).

b) With non-parametric classifiers, the assumption of a normal distribution of the dataset is not required. Therefore, these classifiers are especially suitable for the incorporation of additional geo-data (Lu and Weng 2007). Among the techniques utilised with non-parametric classifiers are Artificial Neural Networks (Kanellopoulos et al. 1997), advanced tree-type classifiers, such as Random Forests (Gislason et al. 2004), and Support Vector Machines (Foody and Mathur 2004). Although non-parametric methods can provide a significant improvement of the classification accuracy, the inclusion of knowledge about classes and their relation to additional data-sources is only considered indirectly by improving the algorithms while including more data.

c) Knowledge-based methods build rules with *a priori* knowledge, to relate the information of the ancillary data to the spatial distribution of certain target classes (Maselli et al. 1995). These methods can be distinguished in (1) knowledge-based modelling of geo-information and remote sensing image analysis, (2) interactive systems, which combine methodological steps of geoinformatics and digital image analysis, and (3) internally integrated classification algorithms. The latter technique often utilises a known non-parametric method extended with a knowledge-base.

Especially for the detection of a specific and complex composition of vegetation patches (as it is the case for the monitoring of biodiversity) knowledge about the relation between abiotic and topographic conditions and the actual biotic representation is crucial for the classification success. Therefore knowledge-based methods were applied in this thesis as modelling approaches (Chapter III) and integrated classification approaches (Chapter IV).

## 3.3 Validation of the classification process

The validation of thematic maps derived by remote sensing image classification analyses is often utilised by accuracy assessments, based on training samples. Amongst others, the Kappa coefficient and the results of an accuracy assessment matrix are the most frequently applied algorithms classification evaluation (Wilkinson 2003). In both cases, information

about the percentage of correctly classified samples for each class and the complete result will be given.

Unfortunately, the influence and significance of different geo-data sources and rule-sets on the classification accuracy is not examined with these methods. When classifying multisource data, knowledge about the effects of including different sources of geo-data would be helpful in terms of

- using this information to omit or weight certain types of geo-data within the classification process, depending on their significance,
- evaluating and enhancing single parameters of rule-base in the case of a knowledge-based method,
- gaining knowledge about the effectiveness of different integration methods of geo-data into an image classification.

Some advances in measuring the significance of difference classification results was made by the usage of pairwise *t* tests, which proved to be a good significance measure between different multisource classification settings (Sader et al. 1995) and by a variable importance estimate of the Random Forest method (Breiman 2001). However, both methods were not used in the context of knowledge-based methods so far. Hence, the evaluation of the significance of single rule-sets associated to a data-source was not included in these analyses. Therefore, different methods of evaluating the significance of the classification results and their included geo-data and rule-base were applied in this thesis (Chapter V)

# 4 Research areas

The investigation focuses on two case study areas, both situated in the pre-alpine region of Bavaria in Southern Germany (see Figure 3). Both areas are forested NATURA 2000 sites with a high percentage of deciduous tree types. The areas were selected for the following reasons:

- Distributed across three biogeographic regions (alpine, atlantic and continental) Germany has so far proposed 4,622 SAC (or pSCI) to the EU. The sites reported represent 9.3 percent of the country's land surface (BfN 2008). Covering most parts of the land's surface, Germany has a high responsibility for habitats in the Continental biogeographic region. Before anthropogenic influence much of the Continental region was dominated by deciduous forest. The climatic conditions and soils are particularly well suited to broadleaved forest, such as beech which is at the centre of its distribution here and oak and hornbeam (Sundseth 2005). Although

most of the natural forest has been replaced by agriculture, the countries which are situated within the Continental region own still the highest percentage of the habitat group "Forests of Temperate Europe" (habitat group: 91xx). Especially the habitat types "*Luzulo-Fagetum* beech forests" (habitat type 9110), "*Asperulo-Fagetum* beech forests" (habitat type 9130), "Oak or oak-hornbeam forests" (habitat types 9160 and 9170), and "Alluvial forests with *Alnus glutinosa* and *Fraxinus excelsior*" (habitat type 91E0) are occurring in all countries of the Continental biogeographic region (European Commission and European Environment Agency 2002).

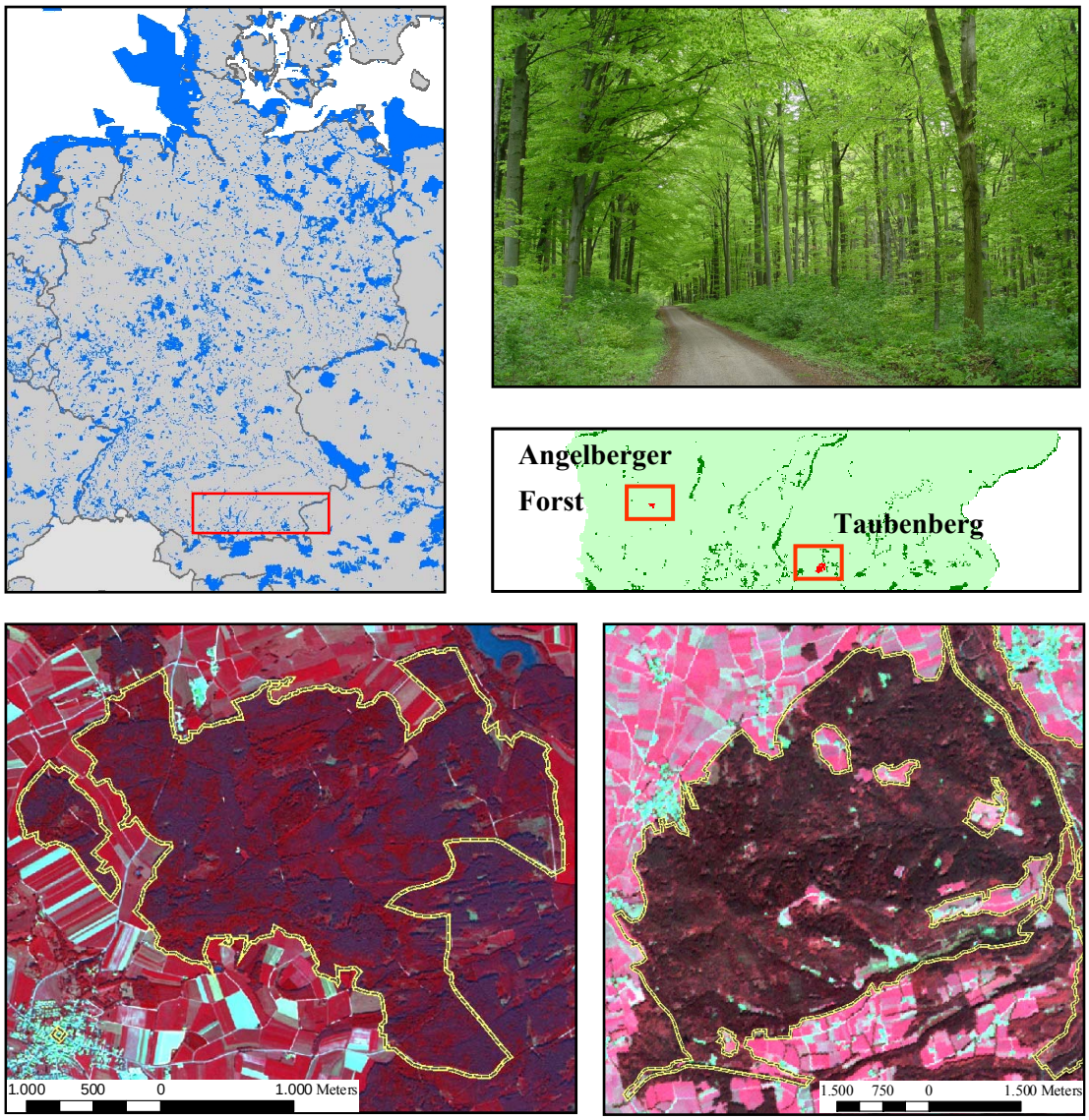



Figure 3: Location and impressions of the research areas: distribution of the SAC`s in Germany (upper left in blue), location of the research areas in Southern Germany (middle right), SPOT5 satellite image of the area Angelberger Forst (lower left), ASTER satellite image of the area Taubenberg (lower right), and photograph of beech forest at the Angelberger Forst (upper right)

- Remote sensing of forests has a long tradition (Hildebrandt 1969). With beginning of the availability of aerial photographs images were used for silvicultural planning and forest inventory, merely by developing manual interpretation keys (AFL 2003; Ahrens et al. 2004). Since the introduction of satellite-based remote sensing, semi-automatic classification techniques have been developed for different purposes, such as woodland monitoring (Hoffmann 2001), mapping of quality parameters like the existence of deadwood or forest structure (Hese 2001; Heurich 2006), or the detection of forest drought damages (Kleinschmit and Coenradie 2005b). With the introduction of VHSR-satellite data more recent approaches in remote sensing such as object-based classification were frequently applied in forestry (Coenradie 2003; de Kok 2001). Moreover, approaches using LIDAR data are shown to be especially useful for forest inventory and monitoring (Wulder et al. 2007).
- Forest inventories are necessary for long term planning and management in woodland areas. Therefore, the spatial data-basis of forests in Germany is traditionally very good. Forest inventories, such as the Federal Forest Inventory (Bundeswaldinventur) are based on systematic sampling. Satellite data can be used for the stratification of forest inventories (Dees et al. 1998). Vice versa, inventories can be utilised to validate remote sensing classification results. Moreover, the Forest Site Map is a valuable information source concerning spatial specifications of soils and hydrology (AK Standortkartierung 1996). Forest inventories are often available for forests in Europe. Because of the differentiation in public and private forests in Germany and different interpretation of mapping manuals by federal authorities, the data is still heterogeneous.

In summary, forested NATURA 2000 areas combine the requirement of a regular monitoring, a long methodological experience in remote sensing classification, and a very detailed as well as heterogeneous spatial data-basis. Hence, these areas are suitable study sites for investigating the integration of geo-data in classification processes.

As test areas two sites („Angelberger Forst“ and „Taubenberg“) in the pre-alpine region of Southern Bavaria were chosen (see Figure 3). Within these NATURA 2000 sites, different semi-natural mixed forest types exist, including especially beech forests (9110, 9130), alluvial forests (91E0), and bog woodland (91D0).

The **Angelberger Forst** (SCI no. DE-7829-301) is situated in the landscape “Donau-Iller-Lech-Platte” (D64) within the Middle Swabian Upland at altitudes ranging from 580 to 650 m. Tree species occurrence in the actual forest vegetation indicates a significant human impact by cultivation of conifers. Today, the Angelberger Forst comprises 24 % broadleaved

forest, against 66 % conifer-broadleaved forest. The broadleaved forest is dominated by *Luzulo-Fagetum* (habitat type 9110: 17.9 %) and *Asperulo-Fagetum* (habitat type 9130: 3.5 %). *Stellario-Carpinetum* (habitat type 9160: 1.3 %) occurs at periodically moist locations. Very moist habitats, mostly along streams, are covered by alluvial forests with *Pruno-Fraxinetum* and *Carici remotae-Fraxinteum* (habitat type *91E0: 2.0 %). The distribution and sequences of the forest habitats are clearly arranged following the gradients of soil moisture and supply of base/acidity. A higher degree of anthropogenic influence can be identified by a higher share of coniferous trees within the mixed forest. The Angelberger Forst is nearly completely governmentally owned, which assures a good availability of digital data. Because of the relatively homogeneous conditions in terms of data homogeneity and vegetation sequences, this area was assumed as a simple classification case.

The **Taubenberg** (SCI no. DE-8136-302) is situated ca. 15 km north of the Alps close to Miesbach/Upper Bavaria and covers an area of ca. 1,800 ha. The altitude ranges from 620 to 896 m. Nearly the whole hill is built from the tertiary material “Obere Süßwassermolasse”. The foot of the hill and the surroundings are covered by old glacial moraines and river terraces of the Riss era. The Taubenberg region is dominated by *Asperulo-Fagetum* (28.9 %), while *Luzulo-Fagetum* (1.9 %) occurs less frequent. It is important to note that *Asperulo-Fagetum* contains a considerable percentage of fir-mixture. Moreover, Alluvial forests (7.9 %) and smaller areas of *Tilio Acerion* (habitat type *9180), and acidophilous *Picea* forests (e.g. *Vaccinio-Piceetea*; habitat type 9410) are present. The area has a very distinct microrelief, because the substratum is differently fragile to alteration. Hence, the study site shows a high small scale variation concerning exposition, slope, micro climate, and supply of nutrients. Therefore, a higher share of coniferous forest can have silvicultural as well as natural reasons.

Taubenberg is an important drinking water reserve of the city of Munich which owns around two thirds of the area. Concerning data availability, different owners lead to a more heterogeneous situation. Not all data exist in digital format. Moreover, the data-sets of the Forest Site Map are mapped at different time steps. Because of the relatively challenging conditions in terms of geo-data and natural resources this area was assumed to be a complex classification case.

# 5 Objective and research questions

The first overall objective of this thesis is to develop and evaluate methods to integrate existing geo-data and ancillary information into a knowledge-based multisource classification process of very high resolution satellite imagery in the context of biodiversity monitoring. This objective is addressed by applying techniques of different complexities in

terms of spatial modelling (explicit use of geo-data to derive potential natural vegetation) and classification approach (internal use of additional information for a integrated classification) in order to evaluate their suitability for effective implementation in mapping and monitoring schemes.

The second overall objective is the analysis of the significance of additional geo-data and knowledge of the classification success. In remote sensing digital image processing the preparation of an accuracy assessment based on validation samples is frequently utilised. Unfortunately, the derived values give no information about the contribution of single parts of the original data-set (e.g. spectral bands). Therefore, this objective is addressed by developing and adapting techniques suitable for the description of classification significance.

Based on the two overall objectives a number of research questions are posed:

- Which type of remote sensing data and which method is especially suitable for the monitoring biodiversity in the framework of NATURA 2000? Which classification technique is best to be utilised for the different monitoring tasks at different scales of the Habitats Directive?
- How can multisource data be included in a classification process to improve the classification results?
- By what technique can the significance of different types of geo-data be quantified? Which of the available geo-data concerning this case-study are significant for the classification success?

## 6 Structure of this thesis

The structure of the thesis is summarised in Figure 4. Chapter I introduces the research subject, the research questions and the case study areas, while Chapter II provides an insight into the pre-processing of VHSR satellite data.

Chapters III to VI relate to the research questions outlined above. Firstly, three different modelling approaches for classifying potential NATURA habitat types were applied and compared for both case study areas (Chapter III). Thereafter, an object-based fuzzy logic classification approach utilising VHSR QuickBird data is presented for the research area Angelberger Forst (Chapter IVa). This approach (utilised for both research areas) was then compared with a per-parcel classification approach classifying heathland habitats in Northern Germany (Chapter IVb). In order to test the significance of different data-sources, the classification method of Chapter IVa was applied for different sets of information and

was subsequently compared (Chapter V). Finally, Chapter VI synthesises the results of the four preceding chapters and provides recommendations for future research.

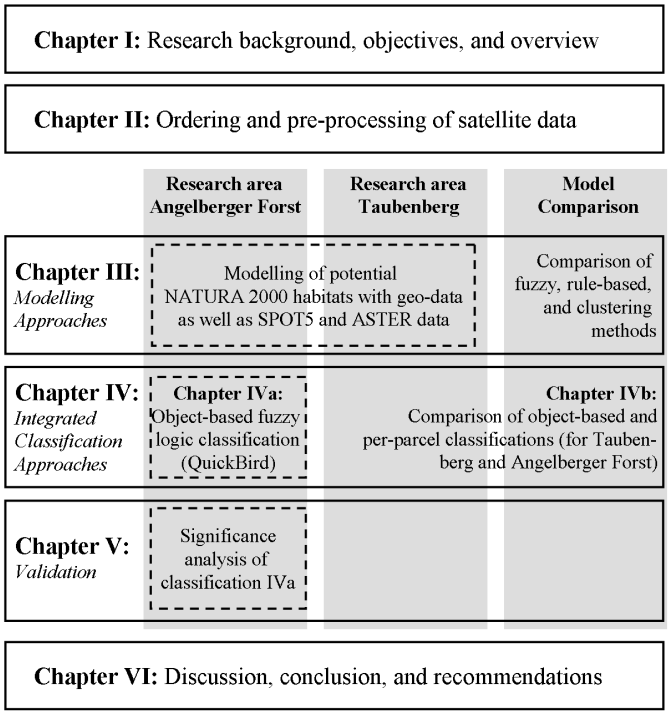



Figure 4: Outline of the thesis

Chapters II to V were written as stand-alone manuscripts to be published or submitted in international peer-reviewed journals or books (book chapter). The four chapters were published or submitted as follows:

Chapter II: Kleinschmit, B., Förster, M., Frick, A., & Oehmichen, K. (2007). QuickBird Data – experiences with ordering, quality and pan sharpening. PFG (1/07), 73-84

Chapter III: Förster, M., Kleinschmit, B., & Walentowski, H. (2005). Comparison of three modelling approaches of potential natural forest habitats in Bavaria, Germany. Waldökologie Online (2), 126-135

Chapter VIa: Förster, M., & Kleinschmit, B. (2008). Object-based classification of QuickBird data using ancillary information for the detection of forest types and NATURA 2000 habitats. In T. Blaschke, S. Lang & G. Hay (Eds.), Object-Based Image Analysis (275-290). Berlin: Springer

Chapter VIb: Förster, M., Frick, A., Walentowski, H., & Kleinschmit, B. (in press). Approaches of utilising Quickbird-Data for the Monitoring of NATURA 2000 habitats. Community Ecology

Chapter V: Förster, M., & Kleinschmit, B. (submitted). Significance Analysis of Different Types of Ancillary Geoinformation into an Object-Based Classification Process. International Journal of Remote Sensing

The stand-alone manuscripts are developed and discussed in teamwork. Therefore, more than a single Author is responsible for the presented research. However, apart from Chapter II the thesis was written by the corresponding author (first author). Therefore, the presented material (of Chapter I, II.4, and III to VI) was written originally by the author and was then subsequently revised by the co-authors. More information about the involved institutions and colleagues is given in the acknowledgement section of this thesis.

The data processing was partly realised by cooperating institutions. Those parts of the work which were not processed by the author are named here:

- In Chapter II, data quality examples are supplied from different research projects (e.g. SARA ’04).
- The multivariate clustering and the fuzzy-logic modelling approaches utilised in Chapter III are processed and supplied from cooperating companies.
- The knowledge-based method for heathland habitats used for comparison of methods in Chapter IVb was developed by the Luftbild Umwelt Planung GmbH.

Each chapter is structured into the subsections background/introduction, study area, methods, results, discussion, and conclusion resulting in a limited amount of recurring material throughout the thesis. Considering the inner structure of the stand-alone manuscripts, these redundancies are retained to allow the reader the undisturbed study of single aspects of this work.

The contents of the pre-published articles have remained unchanged in this thesis. As far as minor details in the articles are outdated, single corrections concerning especially coverage of NATURA 2000 areas and development and launch of new satellite systems are updated and described in footnotes.

# Chapter II:
# Ordering & Pre-Processing

"QuickBird Data – experiences with ordering, quality and pan sharpening"

*PFG – Photogrammetrie – Fernerkundung – Geoinformation 2/2007: 73-83*

Birgit Kleinschmit, Michael Förster, Annett Frick & Katja Oehmichen



*1432-8364/07/2007/0073*
*received July 2006*
*accepted October 2006*

## Abstract

The QuickBird sensor is one of the first commercial satellites that provides a sub-meter resolution. This article presents experiences with the ordering, the quality and pan sharpening of QuickBird data, which were acquired for different purposes in various regions of Germany and Asia.

The ordering process and the characteristics of the four offered products are described. The image characteristics depend mainly on the off-nadir view angle. The influence of slant effects and inclination are shown. Other data quality characteristics of QuickBird images are an induced overcharge in the sensor's charge-coupled devices (CCD) for highly reflective materials like metal or glass and "rainbow" pixels which occur along objects with high contrast. A big advantage of the available 11-bit data range is the possibility to differentiate further details in areas overthrown by shadow. Another challenging effect is the high and artificial texturing of areas with low reflection that should be very homogeneous.

Moreover, the quality of seven different pan-sharpening algorithms of three software products was tested. The study introduces the pan-sharpening accuracy assessment, which considers the spectral reliability of the fused data in comparison to the original image and the desired higher spatial frequency of the merged data. The enhanced I fusion proved to be the most successful in pan sharpening QuickBird images.


## 1 Introduction

The specification "very high spatial resolution" (VHSR) is not well-defined but commonly used for a geometric resolution of multispectral sensors with a ground sampling distance (GSD) of up to 4 m (Ehlers 2002). Examples for panchromatic and multispectral sensors operating as VHSR systems are QuickBird, OrbView 3, Ikonos 2 or Eros A1 (see Table 1). Among these QuickBird, which was launched in October 2001, is one of the first commercial satellites that provides sub-meter resolution imagery. Its panchromatic band collects data with a 61 cm resolution at nadir while the multispectral ground sampling distance is 2.4 m at nadir. The company DigitalGlobe (Longmont, Colorado, US) offers different types of QuickBird high resolution imagery products supporting a wide range of applications such as mapping purposes, monitoring of environmental aspects (floods, earthquakes, oil spills), land management forecasting and fire-risk assessment.

WorldView I, the successor of QuickBird, is scheduled for 2007[3] and will provide a panchromatic resolution of 46 cm at nadir. In 2008 WorldView II is anticipated to launch. It has a multispectral resolution of 1.84 m together with four additional colour bands. (Digital Globe 2006b).

Table 1: Examples for VHSR systems (Jacobsen 2005, modified).

| System | Launch date | GSD [m] pan /MS | Radiometric Resolution | Swath [km] |
|---|---|---|---|---|
| IKONOS 2 – USA | 1999 | 0.82 / 3.24 | 11 | 11 |
| EROS A1 – Israel | 2000 | 1.8 pan | 11 | 12.6 |
| QuickBird-2 – USA | 2002 | 0.61 / 2.44 | 11 | 16.5 |
| OrbView 3 – USA | 2003 | 1 / 4 | 11 | 8 |
| FORMOSAT-2 – Taiwan | 2004 | 2 /8 | 12 | 24 |
| Cartosat 1 – India (stereo) | 2005 | 2.5 pan | 10 | 30 |
| TopSat – UK | 2005 | 2.5 / 5 | n.a.[4] | 15 /10 |
| ALOS – Japan (stereo) | 2006 | 2.5 / 10 | 8 | 35 |
| EROS-B1 – Israel | 2006 | 0.82 pan | 10 | 7 |
| ResourceSat DK-1 – Russia | 2006 | 1 / 3 | n.a. | 28 |
| KOMPSAT-2 – South Korea | 2006 | 1 / 4 | 8 | 15 |
| WorldView I – USA | 2007 | 0.46 pan | 11 | 16 |
| OrbView 5 – USA[5] | 2007 | 0.41 / 1,64 | n.a. | 15 |
| Pleiades – France | 2008 | 0.7 / 2.8 | n.a. | 20 |
| WorldView II – USA | 2008 | 0.46 / 1.84 | 11 | 16 |
| EROS-C - Israel | 2009 | 0.7 / 2.8 | n.a. | 11 |

# 2 Ordering Data

The distribution of QuickBird satellite data is organised by a world wide network of international resellers. The master distributor for Europe and North Africa is Eurimage, headquartered in Rome, Italy, but there are also several local reseller (Eurimage 2006a). The period of time between the initial data order and the actual delivery can vary greatly and depends on several factors. The first step in the ordering process is the decision for one of the available QuickBird products. These products mainly differ in the amount of pre-

[3] WorldView I was launched September 18., 2007 (after publication of this manuscript)
[4] n.a. = not available
[5] OrbView 5 was renamed to GeoEye 1 and launched September 6., 2008 (after publication of this manuscript)

processing that is done prior to delivery. At the moment DigitalGlobe offers the following products (Eurimage 2006b):

- Basic imagery
- Standard and standard ortho-ready imagery
- Ortho-rectified imagery
- DigitalGlobe Digital Ortho Quarter Quad (DG DOQQ: only available for the United States).

Basic imagery is the least processed of the QuickBird Imagery Products. It is radiometrically and sensor corrected, but only geometrically corrected by inner orientation and not mapped to a cartographic projection and ellipsoid. This "quasi" raw data is delivered together with image support data files that provide information about attitude, ephemeris, geometric calibration, camera model, rational polynomial coefficients etc. allowing the customer to perform sophisticated photogrammetric processing such as ortho-rectification and three-dimensional feature extraction.

Standard imagery is delivered with radiometric and sensor corrections. Additionally, it is mapped to a cartographic projection using a coarse digital elevation model (DEM). According to the European data distributor Eurimage it is not suited for producing ortho-images, since the distortion introduced by the coarse DEM cannot be removed later on (Volpe 2003). However, some advances were made in accurate ortho-image generation from QuickBird data (Eisenbeiss et al. 2004). For customers intending to produce ortho-images, a Standard Ortho-ready product can be ordered, which does not use the coarse DEM for geometric correction. In this case further processing with rational polynomial coefficients (RPC) and detailed elevation information is possible in order to achieve good accuracies comparable to those obtained from Basic imagery.

Ortho-rectified imagery is equivalent to the standard imagery, but uses a DEM and ground-control points (GCP) provided by the customer for geometric correction. Therefore the accuracy depends on the number and quality of the provided ancillary data (DEM and GCPs). According to the project objectives (e.g. ortho-images, stereo analysis, classifications), the available ancillary data and the intended data processing steps, the best suited product level and its related options should be chosen. If there is no additional data, the standard product delivering a positional accuracy of 23 m (CE 90 %[6], RMSE[7] 14 m,

[6] Circular error with 90 % confidence (not in the manuscript)
[7] RMSE = Root Mean Square Error (not in the manuscript)

excluding terrain distortions) is recommended (Digital Globe 2006a). Otherwise the amount and quality of the costumer delivered data defines the achievable accuracy level.

QuickBird data can be ordered either out of the comprehensive DigitalGlobe archive or by submitting a new collection request. When ordering out of archive, there is a rush option available. Otherwise there are three different tasking options, namely standard, priority and rush, which differ in multiple acquisition opportunities (including minimal/maximal order sizes), customer defined tasking parameters and prices.

Data turnaround times depend particularly on the chosen tasking option and product level, e.g. ortho-rectified imagery will need more time than the basic product. The delivery of the data can potentially be delayed for weeks or even months (Digital Globe 2006c). For certain applications there may be additional constraints, such as data acquisition during the vegetation period for forestry mapping and agricultural purposes. Furthermore the acceptable cloudiness, off-nadir angle and the size of area and other restrictions will influence the time until delivery as well. So the QuickBird revisit time depends on the latitude of the area of interest and the selected maximum off-nadir angle (see Table 2). Orders specifying large areas with a small off-nadir angle range will require multiple passes and several revisits.

Table 2: QuickBird revisit time in days as function of geographic latitude and nadir angle (Digital Globe 2005).

| **Nadir angle per Latitude** | **0° to 15°** | **0° to 25°** | **0° to 45°** |
|---|---|---|---|
| 0 | 11 | 6 | 3 |
| 10 | 11 | 6 | 3 |
| 20 | 9 | 5 | 3 |
| 30 | 9 | 5 | 2 |
| 40 | 8 | 5 | 2 |
| 50 | 7 | 4 | 2 |
| 60 | 7 | 4 | 1 |
| 70 | 5 | 3 | 1 |
| 80 | 3 | 2 | 1 |

Since there are quite a few applications for QuickBird satellite imagery, unexpected events such as natural disasters or military interests may result in a sudden increasing demand for up-to-date QuickBird data that can further prolong the delivery time. In such a situation where the demand exceeds the acquisition capacity, it seems that smaller orders tend to get less priority than bigger ones. To circumvent the delay for research projects intensive communication is necessary, which may be easier when working with a local reseller.

# 3 Image Quality

The image quality depends mainly on the off-nadir view angle. Larger nadir angles are increasing the pixel size on ground and a longer path through the atmosphere. Scenes captured close to the nadir have a better quality. As soon as the view angle exceeds 15° slant effects occur, which can also affect the classification or interpretation processes. In Figure 1 the differences in two subsets from scenes with 5.6° and 20.5° off-nadir are shown, the subsets have the same scale. In urban areas with very high buildings the inclination is another negative effect (see Figure 4 – left).

Figure 1: Subset of a QuickBird-scene with 5.6° off-nadir view angle on the left (Potsdam) and 20.5° on the right (Lieberose), pan-sharpened image (RGB: 4,3,2).

Because of the sensors' very high radiometric resolution there is no over-saturation of large areas. Nevertheless, highly reflective materials like metal or glass can induce an over-charge in the sensor's CCDs resulting in white cones (see Figure 2).

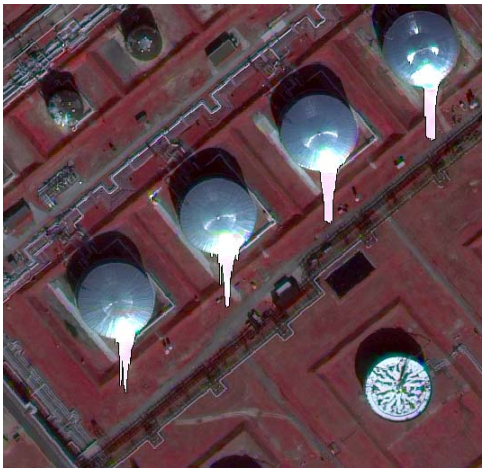


Figure 2: Over-charged areas, pan-sharpened image (Schwedt, RGB: 4,3,2).

A big advantage of the 11-bit data-range is the possibility to differentiate further details in areas overthrown by shadow. In Figure 3 trees in a house-shadow can be interpreted after a

histogram stretch. It depends of course on the kind of urban structure, in areas with very dense and high buildings no scatter light falls into the shadowed areas and no further information can be extracted (see Figure 4 – left).

Figure 3: Shadow-area after a histogram stretch, pan-sharpened image (Potsdam, RGB: 4,3,2).

A negative characteristic of QuickBird imagery are 'rainbow' pixels that occur along objects with high contrast (see Figure 4 – Middle). This effect is due to the separate processing of the single multispectral bands, slight shifts among the bands lead to the assignment of wrong neighbours during the resampling process[8]. DigitalGlobe suggests the use of other convolution kernels for resampling, but with cubic convolution, for instance, almost every image error can be smoothed over. So this is no solution for imagery that is to be used in digital classification.

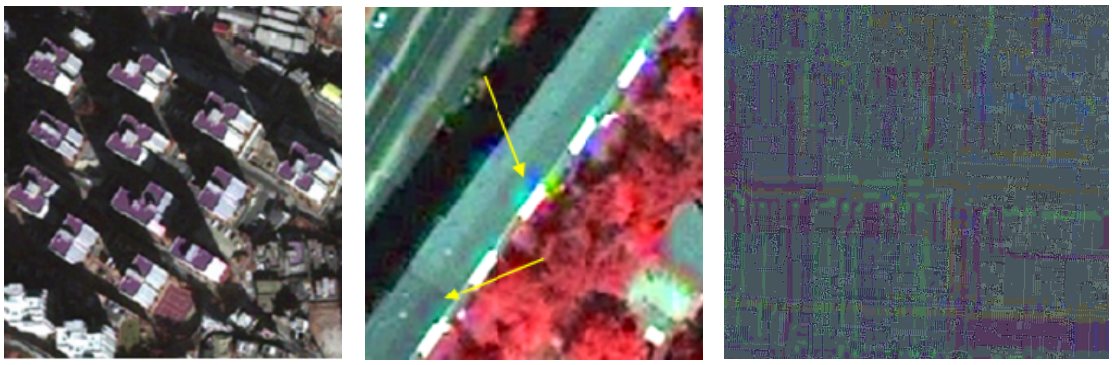


Figure 4: Difficulties with QuickBird data: Left: Problem of shadow and inclination of buildings in very high and dense areas (Seoul, RGB: 3,2,1); Middle: ‚Rainbow' pixels along edges with high contrast, pan-sharpened image (Potsdam, RGB: 4,3,2); Right: Artificial texturing in a water body, pan-sharpened image (Falkensee, RGB: 4,3,2).

[8] internal technical Memo, Eurimage

Another negative effect is the high and artificial texturing of areas with low reflection that should be very homogeneous. In Figure 4 (right) a strangely textured water body is shown. DigitalGlobe finds the source of this error in the downlink process from sensor to earth[1].

Since the resampling process and the applied kernels are object of constant research and change within DigitalGlobe it can result in different standard imagery, though captured on the same day, but processed at a later time. This can lead to serious consequences when additional data is ordered. In Figure 5 a subset is shown where the source of both sides is the same scene (captured in September 2004), the left part was ordered in 2004 whereas the right part was ordered in 2005. Both subsets were processed as standard imagery with nearest neighbour resampling. Both shadow area and tree area are not only shifted but also differently sized. This can affect the extraction of quantitative parameters.

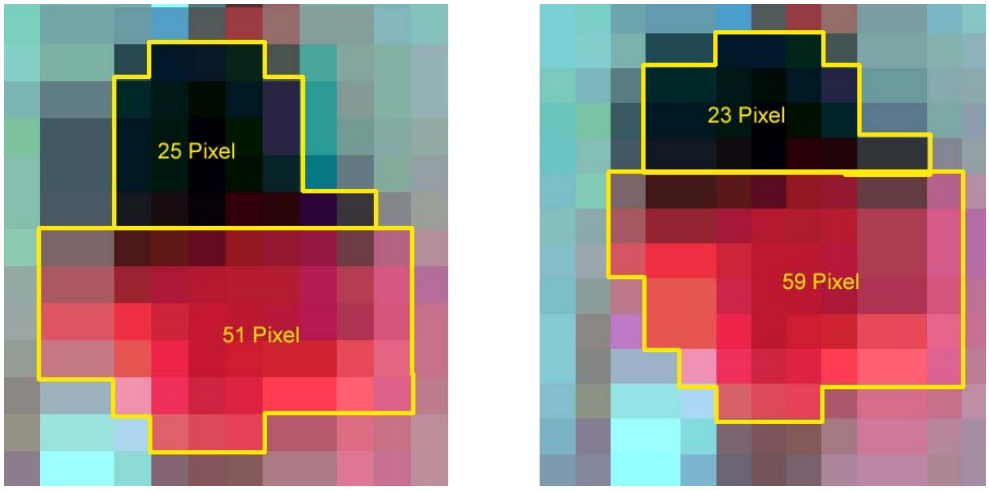



Figure 5: Differences in the resampling process, left: resampled 2004, right: resampled 2005, multispectral image (Lieberose, RGB: 4,3,2).

One of the big disadvantages is the fact, that up to 20 % of cloud coverage have to be accepted. Under certain circumstances this can render an image useless if the most interesting part of the ordered area is covered by clouds or their shadow.

# 4 Pan-sharpening algorithms for data fusion

Commercial image-analysis software packages provide standardised algorithms to fuse panchromatic images of high spatial resolution with multispectral images of lower resolution. In some cases, these algorithms are adapted to certain sensor types, such as QuickBird. There are different quality parameters, depending on the purpose of the image analysis. In this study, the aim of the merging tools is defined to preserve the spectral information, while enhancing the spatial variability. Therefore, additive pan-sharpening

algorithms, such as the Brovey transform (Vrabel 1996) were not considered here. The pan-sharpening tests were examined on a QuickBird image of a pre-alpine area in Bavaria.

To test the quality of the information fusion the image was separately pan sharpened with seven different merging algorithms of three software packages:

ERDAS IMAGINE 8.7 (Service Pack 2):
- Principal component resolution merge (PCA)
- Wavelet PCA resolution merge
- Modified I resolution merge (Siddiqui 2003)

PCI 9:
- Enhanced I fusion (Zhang 2002; Zhang and Hong 2005)

ENVI 4.2
- Gram-Schmidt spectral sharpening (Laben 2000)
- Principal component spectral sharpening (ENVI PC)
- Colour normalised spectral sharpening (ENVI CN)

To examine the dependency of the pan sharpening algorithms on different spectral and textural materials, the analysis was carried out for subsets of three different land-use types (agricultural, forest, urban). In a first step, the statistical features (average, median, minimum, maximum) of these algorithms were compared to the original QuickBird image. For two merging tools, PCA and ENVI CN, the average grey values for the subset differ significantly from the multispectral values of the original image, especially in Band 4 (see Table 3). If these standard statistical parameters are not adapted to the spectral behaviour of the original scene, a later interpretation is likely to produce misclassifications. Therefore, these algorithms were not used for further investigations.

Table 3: Exemplary analysis of average grey values of the test area with predominantly agricultural usage. Similar results were found for other land-use types.

| | **Band 1** | **Band 2** | **Band 3** | **Band 4** |
|---|---|---|---|---|
| PCA | 44.77 | 40.60 | 31.41 | 95.57 |
| Wavelet PCA | 44.03 | 39.90 | 29.71 | 101.02 |
| Modified HIS | 44.46 | 40.35 | 30.08 | 101.64 |
| Enhanced I fusion | 44.56 | 40.43 | 30.26 | 101.33 |
| Gram-Schmidt | 44.57 | 40.43 | 30.26 | 101.29 |
| ENVI PC | 44.57 | 40.44 | 30.26 | 101.36 |
| ENVI CN | 56.27 | 51.04 | 37.38 | 132.80 |
| Original | 44.57 | 40.44 | 30.26 | 101.28 |

The statistical parameters of the ENVI PC and the Wavelet PCA resolution merges achieved statistical results, which were close to the original multispectral image. However, these

algorithms had visible colour distortions, which are described especially for wavelet transforms (Zhang 2002). A possible reason for this effect is a poor co-registration of the pan and the multispectral bands, which have a slightly different view angle and recording time when receiving the data (Terhalle 2005). These colour distortions could lead to misclassification in further analyses of the data (see Figure 6). Consequently, these pan-sharpening tools were not further examined.

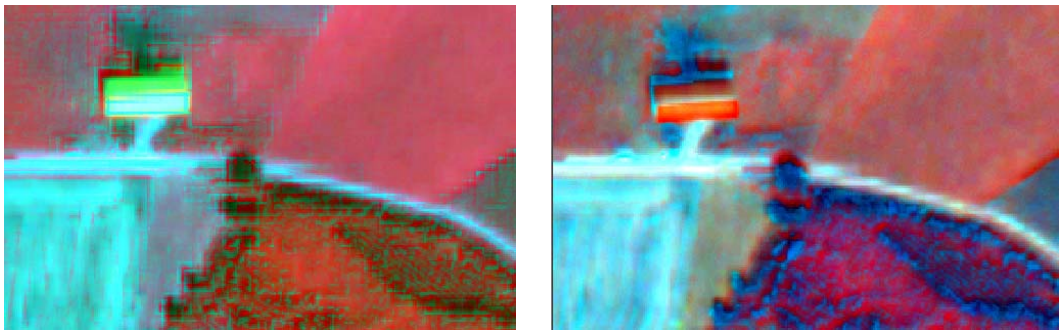


Figure 6: Pan-sharpening results (Angelberger Forst, Bavaria, RGB 4,2,1) with colour distortions of the Wavelet PCA merge (left) and the ENVI PC merge (right).

The remaining three pan-sharpening tools showed visually and statistically reasonable results (see Figure 7). Since an objective visual comparison is only possible to a limited degree, an assessment of the pan-sharpening quality had to be found.

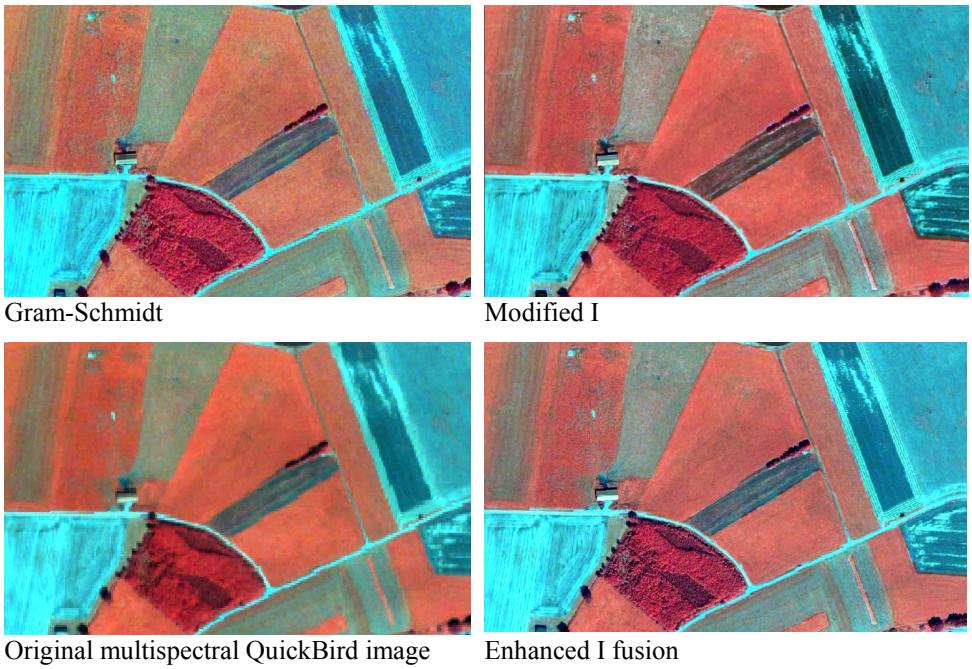



Figure 7: Pan-sharpening results of the enhanced I fusion, the modified I and the Gram-Schmidt method for an agricultural subset of the image (pre-Alpine agricultural area, Bavaria, RGB 4,2,1).

Therefore this study introduces the **pan-sharpening accuracy assessment**, which considers the spectral reliability of the fused data in comparison to the original image and the desired higher spatial frequency of the merged data.

Firstly, the pan-sharpened image will be subtracted from the original multispectral image. If a fused pixel has the same value as the original value, the result is zero. Averaged over a scene, a low value shows high spectral reliability.

In a second step, the higher spatial frequency is taken into account. The result of the subtraction is additionally processed with a focal minimum filter (Kernel 5 x 5). This process is necessary because – although the spectral behaviour of the scene should be constant – a spatial variability of grey values is necessary for a higher resolution image. Therefore, in a surrounding of 5 by 5 pixels the minimum difference value of the pan sharpened and the original image was calculated. With these two easily processed steps, the pan-sharpening accuracy assessment supplies valuable information on the fusion quality. Additionally, areas of spectral deviation can be visualised. In Figure 8, two results of the pan-sharpening accuracy assessment are shown. The areas of the image with high differences to the original multispectral image have higher values (shown in brighter tones), and indicate a lower pan-sharpening accuracy. The Gram-Schmidt algorithm shows differences in areas with very high reflectance values as can be seen with sealed surfaces in Figure 8.

Figure 8: Pan-sharpening accuracy assessment of band 4 of the Gram-Schmidt algorithm (left) and the Enhanced I fusion (middle) of an urban area (original image of Weyarn, Bavaria, right – RGB 3,2,1). The Enhanced I fusion shows significantly lower values, which indicates a better fusion result.

The pan-sharpening accuracy was statistically analysed for average values of different land uses (see Figure 9). Of the three chosen land covers, forested areas are best pan sharpened with all three algorithms, while urban areas are the most difficult sites to process (see Figure 9). Nevertheless, for all subsets the enhanced I fusion proved to be the most successful in pan sharpening QuickBird images. Especially in agricultural and urban areas, the average values of the spatial accuracy for all spectral bands had smaller differences compared to the original multispectral image. In spatial terms, high reflectance areas, such as sealed surfaces

or fully vegetated areas seem to be constantly overestimated by the Gram-Schmidt and the Modified I algorithm.

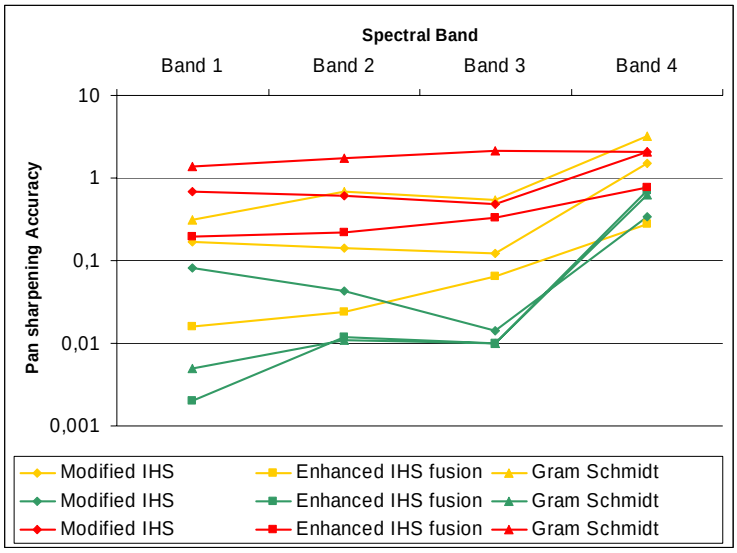



Figure 9: Average values of the pan-sharpening accuracy assessment of three subsets of different land-use (orange = agriculture, red = urban, green = forest) for the three pan sharpening algorithms under investigation. Lower values indicate a better fusion result.

The analysis of pan-sharpening algorithms can only be an intermediate result. New algorithms are already announced[9] (Ehlers and Klonus 2004; Terhalle 2005) or in a scientific development phase (Su et al. 2004; Tu et al. 2005).

[9] The announced algorithms were included into the product ERDAS Imagine 9.1. The names are "High Pass Filter Resolution Merge" and "Ehlers Fusion" (not in the manuscript).

# Chapter III :
# Modelling Potential NATURA 2000 Habitats

"Comparison of three modelling approaches of potential natural forest habitats in Bavaria, Germany"

*Waldökologie online 2/2005: 126-135*

Michael Förster, Birgit Kleinschmit & Helge Walentowski



*ISSN: 1614-7103*

*received October 2005*

*accepted October 2005*

## Abstract

In the context of the EU Habitats Directive, which contains the obligation of environmental monitoring, nature conservation authorities face a growing demand for effective and competitive methods to survey protected habitats. Therefore the presented research study compared three modelling approaches (rule-based method with applied Bavarian woodland types, multivariate technique of cluster analysis, and a fuzzy logic approach) for the purpose of detecting potential habitat types. The results can be combined with earth observation data of different geometric resolution (ASTER, SPOT5, aerial photographs or very high resolution satellite data) in order to determine actual forest habitat types. This was carried out at two test sites, situated in the pre-alpine area in Bavaria (Southern Germany). The results were subsequently compared to the terrestrial mapped habitat areas of the NATURA 2000 management plans. First results show that these techniques are a valuable support in mapping and monitoring NATURA 2000 forest habitats.


## 1 Introduction

NATURA 2000 sites cover approximately seven per cent of the territory of Germany[10]. The EU Habitats Directive – council directive 92/43/ECC – (European Commission 2003a) requires a standardised monitoring of the habitat types and a reporting every six years. For this reason, an operational, objective, economically priced and as far as possible automated application is required. The rapidly developing remote sensing sensor technique and also new image processing methods offer new possibilities to apply remote sensing data for NATURA 2000 monitoring. The possibilities of remote sensing techniques for detecting and monitoring of biodiversity within the scope of the Habitats Directive has already been proven by EU projects SPIN (Langanke et al. 2004) and EON 2000+ (Sell et al. 2004). Consequently this study is a contribution to an effective implementation of these methods for operational use at the regional level (of German federal states). Therefore these methods have to be cost-effective and highly standardised to be a support in terrestrial mapping processes.

[10] The value was stated in the manuscript according to the year 2005. The recent percentage is stated at Chapter I - p. 10 (not in the manuscript).

With the introduction of a more detailed monitoring guideline as an extension of the EU Habitats directive – DocHab 04-03/03 rev.3 – (European Commission 2005)) the EU states that assessing and evaluating the conservation status of habitats and species within the NATURA 2000 network is not sufficient. Therefore an aggregation of the evaluated monitoring sites to the biogeographical level is necessary (Neukirchen 2005). The estimation of the spatial distribution and the kind of NATURA 2000 habitat types in a biogeographical region can be based on the modelling of potential natural vegetation associations. This can be done by the use of expert knowledge about the requirements of the habitat types in respect to site specific factors, such as soil type or relief. For this task, besides the study of historical sources, a modelling of potential natural vegetation as well as a monitoring with remote sensing techniques is required. Furthermore the modelled results of potential natural vegetation can give a support for mapping aerial photographs or classification of satellite images.

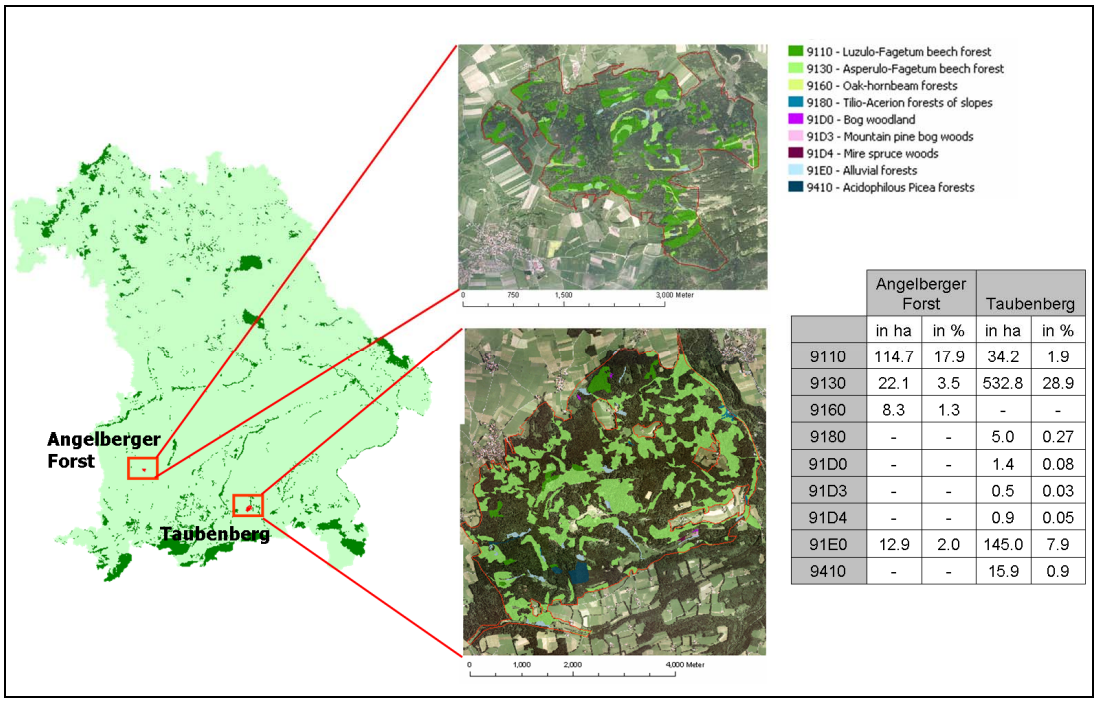
Up to now, area-wide information about the potential natural vegetation is available only at a small scale (Seibert 1968) or in form of statistical data (Walentowski 2001)[11]. In order to derive spatially more detailed facts, a rule-based, a multivariate, and a fuzzy logic approach were applied.

## 2 Test areas

As test areas two sites („Angelberger Forst“ and „Taubenberg“) in the pre-alpine region of Southern Bavaria were chosen (see Figure 1).

These test sites sized 640 and 1,850 hectares respectively, are nearly completely wooded. Within these NATURA 2000 sites, different semi-natural mixed forest types exist, including Beech forests (9110, 9130), Alluvial forests with *Alnus* and *Fraxinus* (91E0), and bog woodland (91D0). To evaluate the results of this study forest management plans (Seitz and Kessler 2004) or publications including mapped woodland habitats (Walentowski et al. 2005) were available for the two selected areas . Figure 1 shows the locations of the test sites and the existing forest habitat types.

[11] This statement referred to the situation in Bavaria. However, there are more studies and information about modelling potential natural vegetation (e.g. Kenneweg, H., Schardt, M., & Sagischewski, H. (1996). *Beobachtung von Waldschäden im Gesamtharz mit Methoden der Fernerkundung*. (pp. 229). Berlin: TU Berlin).




| | Angelberger Forst | | Taubenberg | |
|---|---|---|---|---|
| | in ha | in % | in ha | in % |
| 9110 | 114.7 | 17.9 | 34.2 | 1.9 |
| 9130 | 22.1 | 3.5 | 532.8 | 28.9 |
| 9160 | 8.3 | 1.3 | - | - |
| 9180 | - | - | 5.0 | 0.27 |
| 91D0 | - | - | 1.4 | 0.08 |
| 91D3 | - | - | 0.5 | 0.03 |
| 91D4 | - | - | 0.9 | 0.05 |
| 91E0 | 12.9 | 2.0 | 145.0 | 7.9 |
| 9410 | - | - | 15.9 | 0.9 |

Figure 1: Test sites in the pre-alpine area of the federal state Bavaria. The "Angelberger Forst" area is dominated by habitat type 9110 (*Luzulo-Fagetum*) while the "Taubenberg" area is dominated by habitat type 9130 (*Asperulo-Fagetum*). For the area of the "Taubenberg" it is important to notice that habitat type 9130 consists of a considerable percentage of fir-mixture. Within European NATURA 2000 areas *Abies* forests should be recorded carefully as special habitats. Because of their transitional character between temperate beech forests (habitat type 9130) and boreal spruce forests (habitat type 9410) the forest type is assigned to habitat type 9130 (Walentowski et al. 2005).

## 3 Spatial modelling of potential natural forest associations

Although the forest composition in Central Europe is heavily influenced by human woodland practices, the existence of certain forest types depends on the natural conditions. Those factors, such as soil type, height and steepness of an area, climatic conditions, or availability of water are even included in planting decisions with the use of Forest Site Maps. Consequently, spatial modelling of potential natural forest types from natural location-factors and knowledge about the growing conditions of different tree types can help to determine potential natural forest (Döring and Jansen 2005; Jansen et al. 2002b). Moreover, at a biogeographic scale the approach can supply suitable NATURA 2000 areas. In combination with remote sensing data a spatial modelling is even capable of monitoring different habitat types (Förster et al. 2005a; Frick et al. 2005).

For the modelling of the potential natural forest associations a Digital Terrain Model (DTM 5 and DTM 25), a Conceptual Soil Map (1 : 25.000) as well as a Forest Site Map were used (see Table 1). To exclude regional knowledge and receive results transferable to other NATURA 2000 sites, the data was processed without a field survey.

Table 1: List of existing spatial data. Because of the higher percentage of private forest in the „Taubenberg“ area, the availability of information is more difficult.

| Data Type | Angelberger Forst | Taubenberg |
|---|---|---|
| Map of habitat types | available | available |
| Digital Terrain Model | available (DTM 25) | available (DTM 5 and DTM 25) |
| Climate data | Climate atlas Bavaria | Climate atlas Bavaria |
| Conceptual Soil Map | available | available |
| Forest Site Map | available | not available |
| Forest Organisation Map | available | not available |
| Management Plan | available | not available |

## 3.1 Rule-based method

The purpose of this easily comprehensible and rule-based method was to model potential natural forest associations in areas with identical natural woodland composition (Walentowski et al. 2004). The test sites are situated in the region of mountainous mixed forest (Taubenberg) and beech forest (Angelberger Forst). For habitat types, which could exist in this natural woodland composition, a register of location-factors was developed, including soil type, relief type, water balance, and site- related additional attributes, such as the location of very dry areas. Furthermore sites with a high (H) suitability and sites where the existence of the habitat type is generally possible (P) or excludable (E) were distinguished. Based on the existing suitability the geo-data were combined to a set of rules (see Figure 2):

1. Modelling of locations with high (H) and excludable (E) suitability for a habitat type. As an example, the possibility of the existence of bog woodland (91D0) on brown soils will be excluded, while the existence on peaty soils is highly possible.

2. Modelling of the possible occurrences (P) of a habitat type. A case in point would be the suitability of *Luzulo-Fagetum* (9110), which has a wide range of possibilities to occur. This habitat type can grow on different relief types (southern exposition, steep slopes, hilltops) as well as on different soil types (sand, gravel). These geo-factors are not spatially exclusive. Therefore a site can be chosen as a possible habitat type due to one or several parameters. In this case, the number of possible occurrences was summed up.

3. The calculated site qualities for the habitat types were combined. Firstly a potential forest association is chosen for areas with high suitability (H). At sites without H the habitat type with the largest number of possible occurrences was selected.

4. At each grid cell the dominating habitat type will be chosen as the potential natural forest association. The result is a complete spatial database of the potential natural vegetation.

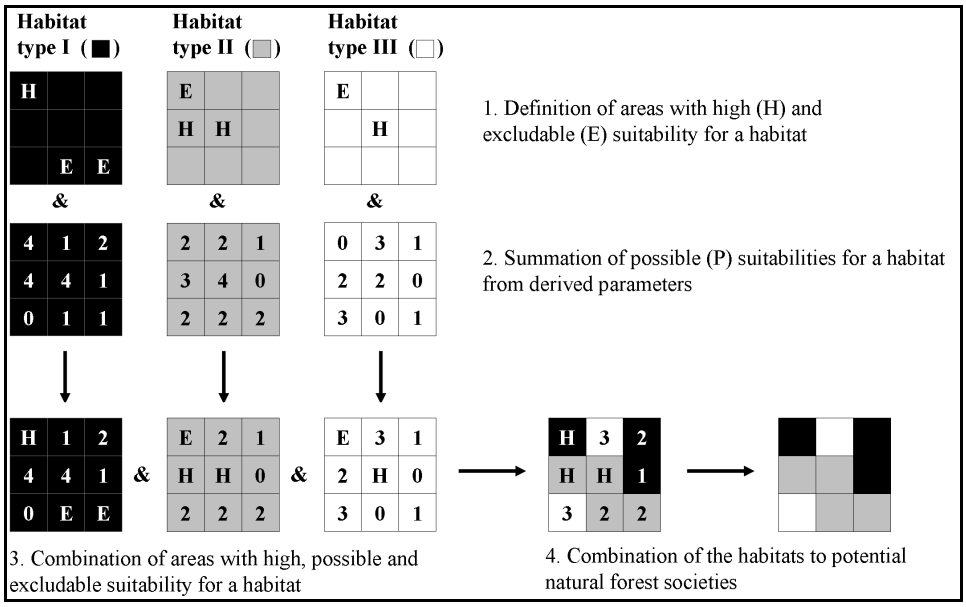



Figure 2: Scheme of the rule-based model.

## 3.2 Multivariate clustering of relief types

The multivariate clustering method of relief types (developed from SciLands inc.) assumes that distribution patterns of various site specific factors are significantly affected by changes of the terrain. Classifying the terrain into landscape ecologic relevant morphographic units can give important evidences concerning the distribution of woodland associations. Therefore three autonomous relief type categories of the terrain were calculated:

- Subdivision of the relief in summit areas, bottom areas, slopes, and closed depressions.
- Differentiation of areas with convergence of discharge (e. g. vales), areas with divergence of discharge, and intermediate areas, which mediate between convergence and divergence.
- The terrain is organised into areas with a slope perpendicular to contour lines and into areas which are steepened or flattened relative to the extremes.

The calculated subdivisions were summarised to clusters. The applied approach is a combination of the iterative minimum distance method (Forgy 1965) and of the hill climbing routine (Rubin 1967). The differentiated clusters were assigned to the possible habitat type in the pre-alpine area. The method relies exclusively on the digital terrain model (DTM). Hence

only habitat types which strongly depend on relief parameters, such as Tilio-Acerion forests of slopes, screes and ravines (9180), can be processed. Because of the higher intensity of the relief only the NATURA 2000 area „Taubenberg“ was used for the multivariate clustering of single habitat types, such as 9180 or 9150.

## 3.3 Fuzzy logic method

The relation between a suitable site and a vegetation association is often described as a “blurred relation” (Glavac 1996). The mathematical branch of fuzzy logic (Zadeh 1978) offers an instrument for handling blurred and only qualitatively described information, so-called fuzzy relations. Due to the fact that there is a great variety of knowledge about the relation between site conditions and plant associations (Ellenberg 1996), the model BERN (**B**ioindication for **E**cosystem **R**egeneration towards **N**atural conditions) was developed to integrate these facts.

In order to better integrate ecosystematic connections, the BERN model was developed on the basis of empirical compilations, performed within a well-monitored region of Germany (Schlutow and Hübener 2004). The BERN model database includes in the first stage only the fundamental niches of the plant species with their blurred thresholds of the suitable site parameters (base saturation, C/N-ratio, soil moisture, length of vegetation period and continentality index (Schlutow and Hübener 2005). The combination of these site-factors which influence the vegetation vitality results in a possibility for plant existence. In the second stage the niche of the whole plant community had been modelled by combining the fundamental niches of the constant plant species with a fuzzy “AND” operator.

The model consists of the following stages:

1. Investigation of the primary natural characteristic of the site conditions (soil moisture, climate, relief type, exposition, soil type).
2. The BERN database contains fuzzy constraints for >700 plant species according to the parameters mentioned above, taken from expert literature of the region (Oberdorfer 1992; Walentowski et al. 2004). The fuzzy constraints are combined by the minimum operator, because the ecological limiting factor (e.g. poor soils) represses the plant growth, even if all other site-factors are in favour for the development of certain vegetation. This leads to a 5 dimensional possibility distribution function of the species for the site conditions.
3. The possibility measures were combined with a fuzzy gamma operator regarding the species of a certain vegetation type.

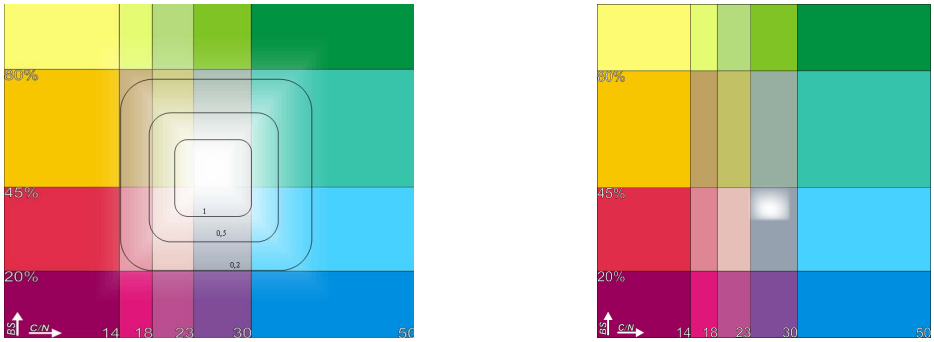



Figure 3: Distribution function of the possibility of occurrence of the tree-type red beech (*Fagus sylvatica*) and the habitat type *Luzulo-Fagetum* in relation to the site-factors base saturation and C/N-Relation (ground soil). The white colour shows the distribution of the possibility of occurrence for red beech from 0 to 1.

As Figure 3 shows for the example of the red beech (left), different distribution functions of the possibility of occurrence of tree types in relation to the site-factors were developed and used competitively. In a second step the tree types were aggregated to habitat types (here *Luzulo-Fagetum* – 9110). The habitat type with the highest possibility of occurrence was assigned to each raster cell.

# 4 Classification of forest types and combination with the modelling results

The implementation of remote sensing in detection and monitoring of NATURA 2000 habitats and site quality key parameters is stated by various authors (Granke et al. 2004; Lang and Langanke 2005). In order to identify real forest habitat types, the modelled potential natural forest associations had to be combined with a classification of the in situ vegetation. To guarantee a possible implementation into the workflow of the local authorities, the acquired data had to be cost-effective and processable using standard methods. The satellites SPOT5 and ASTER were considered suitable for the differentiation of coniferous, deciduous, and mixed forest as they offer a spatial resolution of 5 m to 15 m and spectral bands in the infrared and near infrared region. The satellite scenes of SPOT5 (acquisition date: 07.09.04) and ASTER (acquisition date: 14.09.04) were georeferenced using the DTM. Furthermore a Minnaert correction was carried out with the objective of topographic normalisation (Kleinschmit and Coenradie 2005a). On the basis of scanned and georeferenced true colour air photographs and Forest Organisation Maps[12], training areas for all existing principal tree species and types of mixture were defined. The result was

[12] The term “silvicultural map” which was used in the original manuscripts is subsequently replaced by the term “Forest Organisation Map” as an expression for “forstliche Betriebskarte”. This map presents the results of the forest inventory.

summarised to higher-ranking classes (see Figure 4). On the basis of these combined training areas a Maximum Likelihood classification was carried out. The result was validated by the use of forest management maps and local knowledge of forest officials.

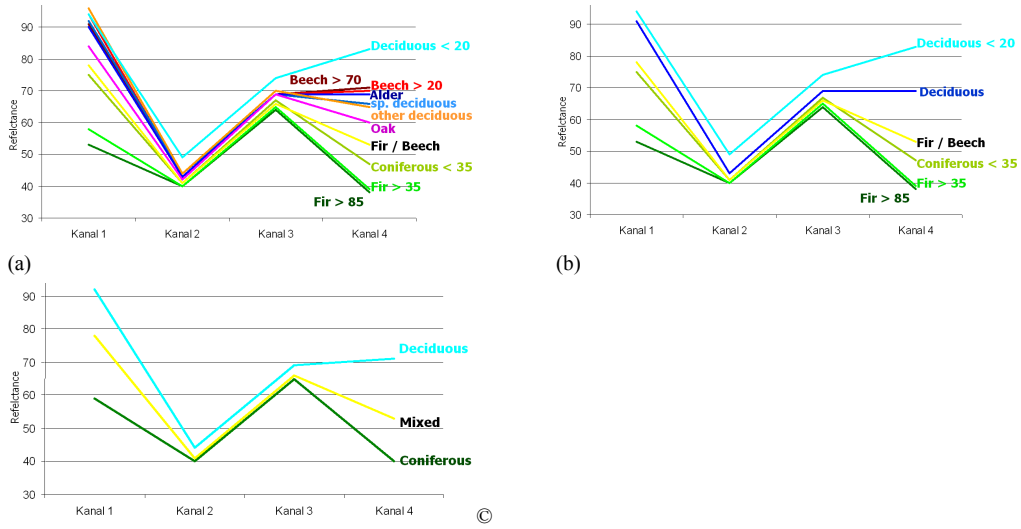



Figure 4: Example of the summarising of tree classes for the test area "Angelberger Forst". The reflectance values were taken from SPOT5 data and assigned to classes with the help of air photographs and Forest Organisation Maps. In the first step (a) different classes of age, tree type and mixes were taken, but reliable information was not possible due to similar spectral behaviour and individual variance of sample. Therefore deciduous tree types were summarised to the class deciduous (b). Because the age information was not necessary for this application, the classes were combined to Deciduous, Mixed and Coniferous (c).

For the test site „Angelberger Forst" the classes deciduous, coniferous, and mixed forest were detected, while for the test site „Taubenberg" the classes deciduous, coniferous – non fir, coniferous – fir and mixed forest were recognised. The results correspond to the experiences of (Blaschke and Felbermeier 2003), who detected similar classes with ASTER and SPOT4. Finally the scenes were segmented in two levels of detail. The majority class in one object was assigned as attribute to each polygon in order to receive more homogeneous classes.

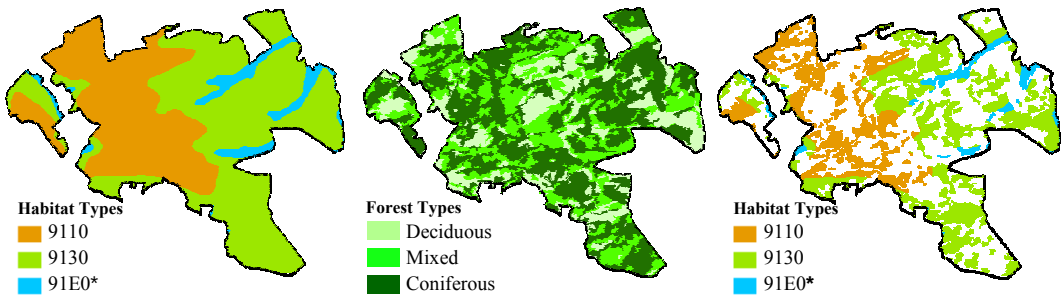



Figure 5: Potential natural forest (left), classification of SPOT 5 data (middle) and combination to potential habitat types (right) for the „Angelberger Forst" (example from fuzzy logic method)

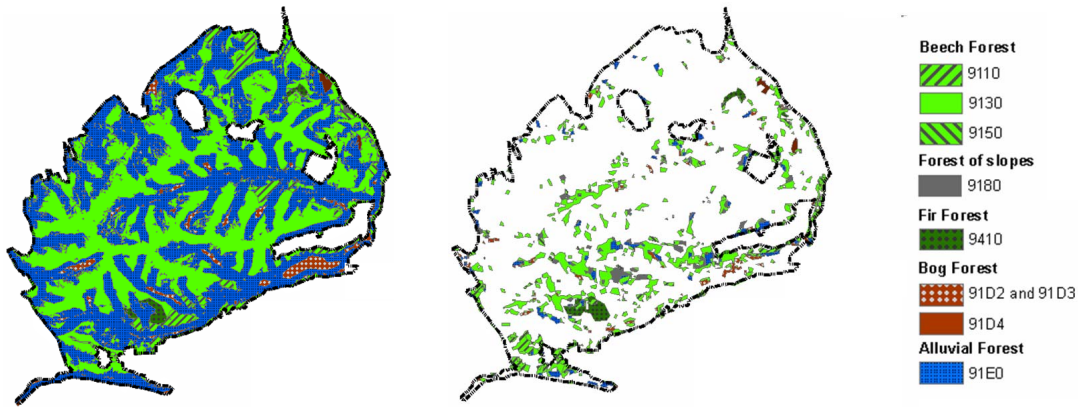



Figure 6: Potential natural forest associations (left) and derived potential habitat types (right) for the test site „Taubenberg“ (example from rule-based method)

The results of the satellite classification were combined with the methods described in section 3 in order to calculate the potential natural forest associations. The potential forest habitat type is only selected if the classified real vegetation corresponds to the topmost forest association (see Figure 5 and Figure 6). An example would be the detection of a potential beech forest association, which is not considered a potential habitat type if the satellite classification result is coniferous forest.

## 5 Comparison of modelled and terrestrial mapped habitat types

The results were compared to the terrestrial mapped habitat types from the forest management plans of the test sites (see Table 2).

A great part of the considered habitat types could be detected by the use of the model approaches. The best results were achieved by using the satellite data with higher spatial resolution (SPOT5) on the test site “Angelberger Forst”. The rule-based method and the fuzzy-logic approach obtained similar results in comparison to the management plan. Generally, apart from *Luzulo-Fagetum* (9110) the fuzzy logic approach achieved slightly higher results. An additional factor for a good correlation is the utilisation of a Forest Site Map (FSM) within the model. It was evident that the models tend to underestimate habitat types with distinct site specific growing conditions, such as Alluvial Forest (91E0) and *Tilio Acerion* Forest of slopes (9180).

In the case of habitat type 9180 the clustering technique of relief types yields better results. Thus it was possible to detect 80 % of the areas with *Tilio-Acerion* forests of slopes, screes and ravines (9180), while the rule-based method achieved only 40 % of this habitat type. Within this research study, the clustering of relief types was only used for this habitat type.

However it can be suggested, that this modelling approach is especially successful in surface-dominated areas such as alpine forests.

Table 2: Exemplary comparison of the results of the three modelling approaches with the forest management plan. Additionally the test site „Angelberger Forst“ was processed with and without a Forest Site Map (FSM). The term “n.a.” stands for “not available” because the habitat type is inexistent in this area or was not processed due to modelling restrictions (green marked areas with accuracies over 80 per cent, orange marked areas with accuracies below 60 per cent).

| **Habitat types** | **Rule-based Method** | | | **Fuzzy Logic method** | | | **Clustering of relief types** |
|---|---|---|---|---|---|---|---|
| | Angelberger Forst with FSM (in %) | Angelberger Forst without FSM (in %) | Taubenberg without FSM (in %) | Angelberger Forst with FSM (in %) | Angelberger Forst without FSM (in %) | Taubenberg without FSM (in %) | Taubenberg without FSM (in %) |
| 9110 | 95.4 | 92.3 | 62.5 | 68.0 | 60.3 | 72.4 | n.a. |
| 9130 | 70.6 | 66.0 | 76.9 | 74.8 | 72.0 | 92.3 | n.a. |
| 9180 | n.a. | n.a. | 40.0 | n.a. | n.a. | 40.0 | 80.0 |
| 9410 | n.a. | n.a. | 66.7 | n.a. | n.a. | 66.7 | n.a. |
| 91D2 – 91D4 | n.a. | n.a. | 80.0 | n.a. | n.a. | 85.7 | n.a. |
| 91E0 | 91.7 | 58.8 | 46.2 | 75.0 | 53.3 | 73.1 | n.a. |
| 9160 | 88.2 | 73.9 | n.a. | 86.7 | n.a. | n.a. | n.a. |

A more detailed comparison, which involved the detected area size of model results and forest management plans, resulted in increased differences. As an example, the areas of Alluvial Forest of the data sets intersect only to 33.2 %. This is due to the model assumption that a habitat type necessarily has the potential natural vegetation. However, a modelled and detected deciduous beech forest can be an artificial planted forest of neophytes. Therefore, it would be impossible to detect these kinds of errors by the use of the methods described above.

The results could be improved if a buffer area surrounding the modelled habitat type would be applied as “suspected potential habitat type”. For the rule-based and the fuzzy-logic method the accuracy could significantly be increased by using the second dominant modelled habitat type (see Figure 7). When combining potential habitat type and potential natural forest association the subordinated forest association could be regarded as possible result.

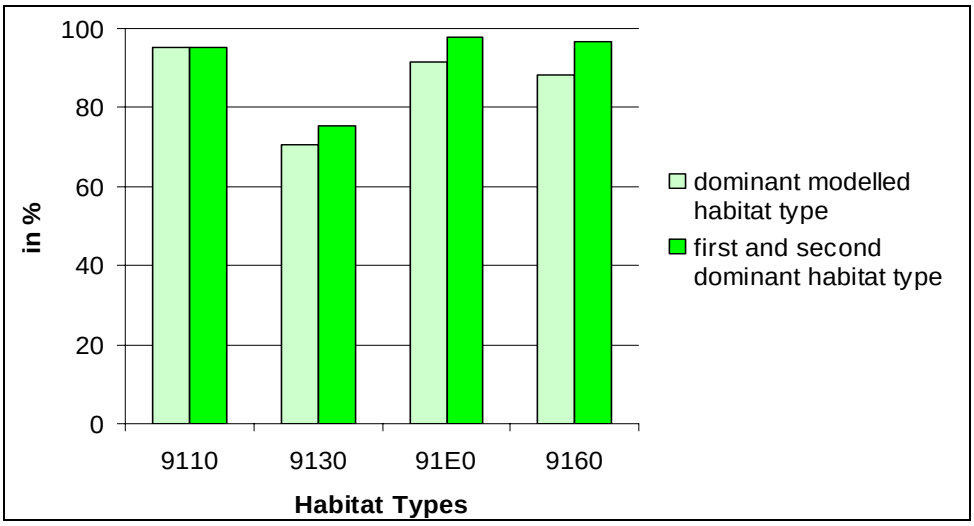



Figure 7: Increased classification accuracy in modelling the second dominant modelled habitat type (example from rule-based method). The results raise up to 7.6 per cent.

To test the efficiency of the modelled approaches, the time expenses for the realisation were recorded. Because of the restricted model boundaries of the clustering of relief types, this was only possible for the rule-based and the fuzzy-logic methods (see Table 3). The tasks were divided into non-recurring steps (e.g. development of a mathematical and botanical basis, data-base design) and recurring steps (e.g. calibration according to special data availability, exceptional natural conditions).

Table 3: Time expenses for the rule-based and the fuzzy logic modelling approach.

| | **Required Time (in hours)** | |
|---|---|---|
| **Non-recurring tasks** | **Rule-based** | **Fuzzy Logic** |
| Development of mathematical and botanical basis<br>Technical realisation of the model | 35 | 30 |
| Preparation of the vegetation data basis | 130 | 160 |
| **Recurring Tasks (for each area)** | | |
| Validation and calibration for test sites | 30-40 | 8-24 |
| Presentation of results (cartography) | 16 | 16 |
| **Sum** | **211 – 221** | **214 – 230** |

Although from an economic point of view this comparison is only valid to a limited degree, a few tendencies do arise. The rule-based approach is more efficient in the development of non-recurring tasks (165 h) while the fuzzy-logic method can manage the recurring tasks more efficiently (24 to 40 h). Therefore with more modelled NATURA 2000 areas the fuzzy-logic approach is more effective. On the other hand, the rule-based method is implemented

in standard GIS software (ArcGIS Model Builder) while the BERN model is a separate software package. This makes the rule-based approach more easily applicable.

# 6 Analysis of the support the models can give to aerial photograph mapping

Besides the modelling of potential natural forest vegetation for the information on the biogeographic level of NATURA 2000, the results of the modelling could be used to support aerial photograph mapping. The mapping of NATURA 2000 areas with the help of very high resolution aerial or satellite images will become of increasing importance. This is due to the monitoring duties of each EU member state obliged to observe the quality of the NATURA 2000 areas in a six-years cycle (European Commission 2003b).

The modelling results were used to evaluate the usage to support aerial photograph mapping. Aerial photographs of the test sites were acquired, orthorectified and classified by the use of the software ArcGIS Stereo Analyst (Seitz and Fischer 2005). The imagery was analysed

- without any additional help,
- with the use of the Forest Site Map,
- with the rule-based model and
- with both ancillary information.

As was to be expected, the mapping without any additional information yielded the poorest results. However, all other results were not convincing. Even the best approach with only the forest site map tends to heavily underestimate oak-hornbeam forests, while not reaching sufficient results for all other habitat type classes (see Figure 8).

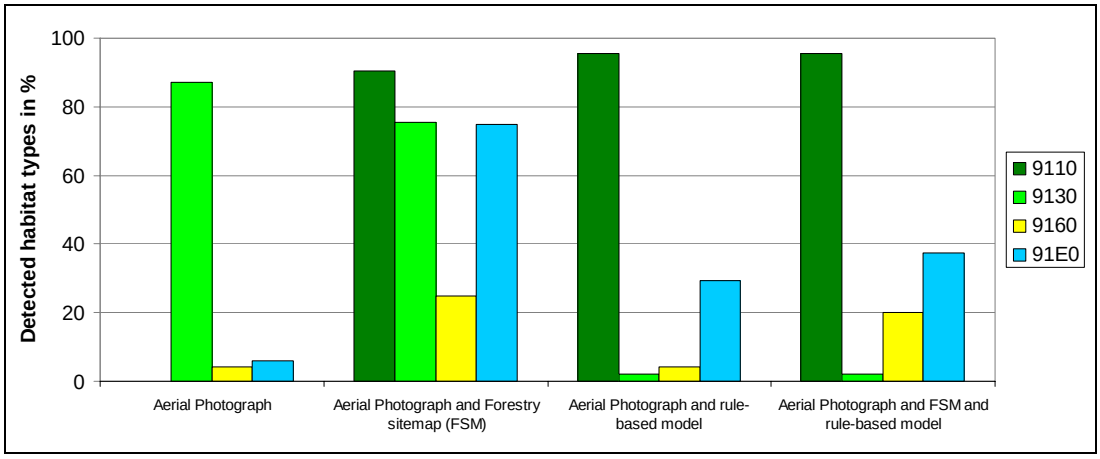



Figure 8: Mapping of habitat types with different additional information (Angelberger Forst).

The inadequate results depended on the quality of the aerial photographs. The imagery was supplied without the near infrared information. Additionally, the scan quality was not of a very high standard. It was recognisable that additional information improved the mapping results. However, a definite conclusion as to what kind of information is useful for the mapping process could not be drawn. It can only be stated that mapping NATURA 2000 forest habitat types can only be successful with colour infrared imagery (aerial photographs or satellite based)[13].

## 7 Discussion and outlook

The results of this study have shown great potentials of the modelling of forest vegetation. They are especially valuable when combined with remote sensing data of high geometric resolution. These results can support mapping as well as monitoring of biodiversity within and outside of the declared boundaries of NATURA 2000 areas. However, there are some points which should be discussed in the context of NATURA 2000. Only a small percentage of the European forests is still in the condition of natural circumstances (UNEP World Conservation Monitoring Centre 2000). The forest has been changed for economic, touristic or hunting reasons, to name only some. However, the assumption of the model is that the vegetation is still natural. Even the areas within NATURA 2000 boundaries are not without human influences. This supposition can be partly clarified by using remote sensing classifications as worked out in this article. However, only a very detailed classification with very high resolution data and near infrared information is able to reduce these errors due to model restrictions. Another possibility would be to include information from forest inventories (Long et al. 2004) into the modelling process to achieve a more realistic result.

Another point of discussion is the method of terrestrial mapping. Often the transitions in between habitat types or between a habitat type and other forest are not clearly to detect. In addition habitat types will be mapped even if the main species covers only a small proportion of an area. The intention in mapping these poorly developed habitat types is to ascribe a higher importance to them, because there is a potential development of the forest association. As a consequence terrestrial mapped habitat types (especially in cases of low percentages of deciduous forest) cannot be detected with remote sensing techniques.

The difficulties explained above lead to the realisation that the dealing with a probability to a certain habitat type would be more reasonable than searching for defined boundaries of

[13] The effect of stereo aerial photographs in comparison to imagery without this advantage was not analysed in this study. However, it can be assumed that there is a increase in classification accuracy when using stereo images (not in the manuscript).

habitat types (see Figure 9). This approach would be more in harmony with the theory of ecotones (Hill et al. 2005). Moreover local knowledge could be included more directly. The rule-based model as well as the fuzzy-logic model could implement such probabilities of occurrence in their computation.

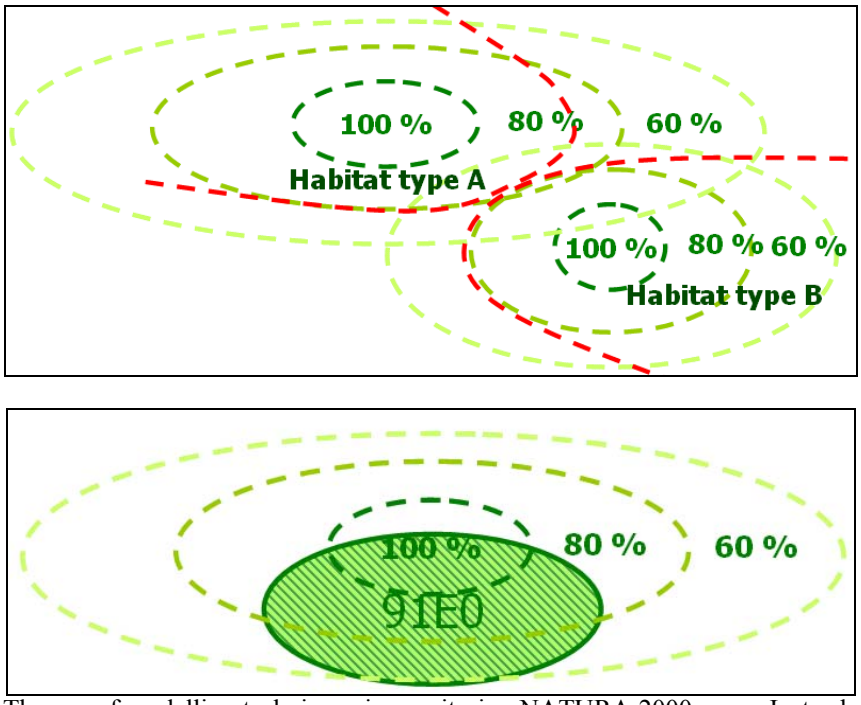



Figure 9: The use of modelling techniques in monitoring NATURA 2000 areas. Instead of sharp borders of assignment to one habitat type, the modelling approaches could calculate regions of probability (upper figure – green dashed lines). Within these regions it is possible for regional authorities with local knowledge to define habitat-type borders (upper figure – red dashed lines). Changes of the habitat type could be monitored by percentages of membership to a defined habitat type (lower figure). When the percentage of membership has changed, the habitat type changed its condition.

The most promising way of improving the results is the use of very high resolution remote sensing data. Approaches with the QuickBird satellite data (Frick et al. 2005) and the HRSC airborne camera (Gähler et al. 2004) could achieve very detailed classification outcomes. Moreover, the direct inclusion of additional geo-data in the remote sensing classification could be applied. A combined application of modelled location-factors and precise classification results is certainly a solution for the great demand of monitoring techniques for NATURA 2000 areas.

**Acknowledgements:** The Authors would like to thank Mr. Seitz and Dr. Fischer from the Bavarian Forest Management Authority who critically evaluated our results. Additionally we would like to thank SciLands inc. (Mr. Bock) for the processing of the cluster analysis and Ökodata GmbH (Mrs. Schlutow and Mr. Kraft) for the processing of the fuzzy logic analysis with the BERN model.

# Chapter IVa:
# Classification of NATURA 2000 Habitats with Additional Geo-Information

"Object-based classification of QuickBird data using ancillary information for the detection of forest types and NATURA 2000 habitats"

*In: Blaschke, Th., Lang, S. & Hay, G.J. (eds.) Object-Based Image Analysis. Springer – Lecture Notes in Geoinformation and Cartography. 2008: 275-290*

Michael Förster & Birgit Kleinschmit



*ISBN: 978-3-540-77057-2*

*received April 2007; revised July 2007*

*accepted October 2007*

## Abstract

The detection of forest types and habitats is of major importance for economic forest management as well as for the monitoring of biodiversity in the context of NATURA 2000. For these purposes, the presented study applies an object-based classification method using VHSR QuickBird data at a test site in the pre-alpine area of Bavaria (Southern Germany). Additional geo-data and derived parameters, such as altitude, aspect, slope, or soil type, are combined with information about forest development and integrated into the classification using a fuzzy knowledge-base. Natural site conditions and silvicultural site conditions are considered in this rule-base.

The results of the presented approach show higher classification accuracy for the classification of forest types using ancillary information than can be reached without additional data. Moreover, for forest types with very distinctly defined ecological niches (e. g. alluvial types of forest), a better characterisation and integration of rules is possible than for habitats with very wide ecological niches. Hence, classification accuracies are significantly higher when these rules are applied. In a second step NATURA 2000 habitat types and selected habitat qualities are derived from the classified forest types. However, the share of habitat qualities varies with an altering scale. This difficulty should be addressed in further research of NATURA 2000 monitoring.


## 1 Introduction

With the development of a standardised and pan-European geo-data-infrastructure (Cragila et al. 2005) remote-sensing applications which integrate GIS information will become increasingly important. Therefore, various studies of combining additional data and knowledge into classification processes (Maselli et al. 1995; Stolz 1998) were undertaken. However, the integration of additional geo-data into very high spatial resolution (VHSR) imagery remains a challenging task.

Fig. 1 shows an exemplary overview of multi-scale segmentation for different forest scales with the corresponding levels of ancillary data and knowledge. Additional information can be differentiated into two categories (see Figure 1). Firstly, spatially explicit knowledge is available. For forestry applications a broad range of this kind of data sources can be used, namely the simulation of geo-data (Hagner and Olofson 2004; Verbeke et al. 2005), the usage of altitude information, especially with LIDAR techniques (Diedershagen et al. 2004), and the integration of Forest Organisation Maps (Förster et al. 2005b) as well as soil and hydrology maps into classification procedures.

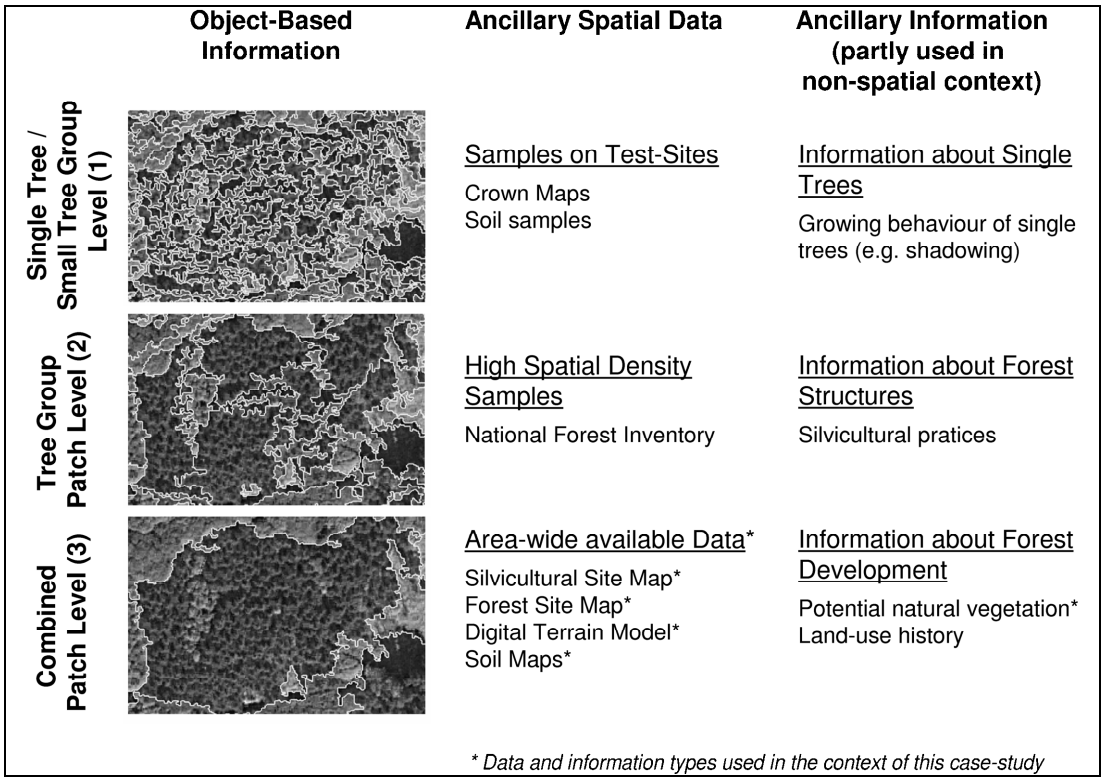



Figure 1: Exemplary overview of multi-scale dependence of object-based information and ancillary GIS-Data and knowledge for a forestry application

Complementary knowledge about processes of the forested landscapes is abundantly available and recorded, i.e. individual tree growth and altitude by measurement of crown diameter. Information about the land-use history, forestry practices (Pretzsch 2002), and the potential natural vegetation (Walentowski et al. 2004) is of high relevance.

Moreover, similar to delineated objects of satellite information (Burnett and Blaschke 2003), ancillary information can be grouped in different scales, which depend on the landscape level the data was assessed or observed. While some information is available on a site-specific scale[14] other knowledge exists only as interpolated data for larger landscape patches.

The objective of this study is to develop a method of integrating different types of geo-data into an object-based classification process. The presented approach combines spectral and textural information of a QuickBird scene with ancillary data for the identification of forest structures and habitats. Because the knowledge of woodland development is often on purpose expressed ambiguously, a fuzzy-logic-based rule set is applied. This example is especially suitable to show chances and challenges of data-integration techniques, because long-term information about forestry practices and ecological woodland development in the study area (Bavaria, Germany) are available.

[14] The information can be likewise available as sample plots (not in the manuscript)

# 2 Data and methods

For the presented study the satellite data are processed with a multi-scale segmentation method (Burnett and Blaschke 2003) by using an object-oriented approach with the software eCognition[15] (Benz et al. 2004). The segmentation levels of different resolution are delineated and assigned to hierarchically organised groups of objects, such as forest habitats, crown combinations and crown types of single-tree species.

The segments are classified as different forest types 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 of the test area is partly 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).

In order to show the practical applicability of the classification, the results are employed to derive NATURA 2000 forest habitat types and qualities. NATURA 2000 is a European directive designed to protect the most seriously threatened habitats and species. The NATURA 2000 network consists of more than 20,000 sites and covers almost a fifth of the EU territory[16]. The directive requires a standardised monitoring of the habitat types and a reporting every six years. From the German point of view, standards for mapping NATURA 2000 forest sites have been developed (Burkhardt et al. 2004).

## 2.1 Study site

As test area the forested NATURA 2000 site “Angelberger Forst” in the pre-alpine area of Bavaria, Germany, was chosen. It covers approximately 650 ha. The test site was selected because a terrestrial mapping of NATURA 2000 sites had been completed. Moreover, a broad variety of different semi-natural mixed forest types exists.

Relatively large areas of this submontane forest are vegetated with Beech, mainly with two special habitats, named *Luzulo-Fagetum* (17.9 %) and *Asperulo-Fagetum* (3.5 %). Periodically moist locations can be covered by *Stellario-Carpinetum* (1.3 %) but with less frequent occurrence. Very moist habitats, mostly along streams, are also vegetated with Alluvial forests with *Alnus* and *Fraxinus* (Alder and Ash; 2.0 %). Additionally, small sites

---

[15] The software eCognition is now renamed to Definiens Professional. Nevertheless, the presented method is transferrable to this newly introduced product (not in the manuscript).

[16] The most recent information (June 2008) is included in Chapter IVb. According to these information the network covers over 26,000 protected areas with a total area of about 850,000 km² (not in the manuscript).

are covered with Larch and Sycamore. However, due to forest management practices a high proportion of coniferous forest (mainly Spruce) exists in this area. The natural allocation mostly depends on soil moisture and acidity, but can equally rely on the relief or anthropogenic influences.

## 2.2 Data

For the presented investigation a QuickBird scene was acquired at 11th August of 2005. The sensors panchromatic band collects data with a 61 cm resolution at nadir while the multispectral (visible and near infrared) ground sampling distance is 2.44 m at nadir (Digital Globe 2006a). A selection of geo-data was supplied by the Bavarian State Institute of Forestry (see Table 1). A Digital Terrain Model (DTM 5 and DTM 25) is used for relief information. The parameters slope, aspect, curvature, and altitude are derived from this data source. Furthermore, a Conceptual Soil Map is used. This map is generated by the Bavarian State Institute of Geology from available data sources (e.g. geological map) and mapped soil samples. The Conceptual Soil Map is available for large parts of Bavaria. In spite the detailed spatial specification (scale of 1 : 25,000), the attributes for the soil types are often summarised and imprecise (an example of a class is: "soil complex of gleys and valley sediments"). As a further important data source a Forest Site Map is utilised. This map contains very detailed information about forest soils and is often used for economic forestry purposes.

Table 1: List of existing spatial data and derived parameters of the study site

| **Type of geo-data** | **Relevant attributes and derived parameter** | **Scale / Availability** |
|---|---|---|
| Digital Terrain Model | Slope<br>Aspect<br>Curvature<br>Altitude | DTM 5 and DTM 25 / parts of Germany |
| Conceptual Soil Map | Soil type | 1 : 25,000 / parts of the federal state Bavaria |
| Forest Site Map | Availability of nutrients<br>Availability of water<br>Soil substrate | 1 : 5,000 / digital for state forests |
| Forest Organisation Map (for training and validation purposes) | Main tree type per stand<br>Tree mixture per stand<br>Age<br>Usage | 1 : 5,000 / digital for state forests |
| Terrestrial Map of Habitat Types (for validation purposes) | NATURA 2000 type<br>Conservation status | 1 : 1,000 to 1 : 5,000 / for large parts of NATURA 2000 areas |
| True colour aerial photographs (for training) | | 1 : 12,400 |

For training of the classes and validation of the results a Forest Organisation Map is used. This map presents the results of the forest inventory. It consists of information about main tree types and tree-type mixtures per forest stand. As training sites a set of 527 sample polygons is selected in non-mixed forest stands with a stratified order for occurring tree types. The sites are allocated randomly within the areas of specific forest stands. The chosen samples are confirmed using current (2003) true colour aerial photographs. In a few indistinct cases, samples are terrestrially mapped together with local forest experts. The knowledge base to build up rule sets for potential forest types was available from a previous project in cooperation with the Bavarian State Institute of Forestry (Kleinschmit et al. 2006). These rules are 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 resolution of 0.61 m (Zhang 2002), all spectral bands of the QuickBird scene are segmented at three landscape scales. These levels are named Single Tree / Small Tree Group Level or level 1 (Scale Parameter[17] (SP) 15, shape factor 0.1, compactness 0.5), Tree Group Patch Level or level 2 (SP 40, shape factor 0.1, compactness 0.5), and Combined Patch Level Structure or level 3 (SP 150, shape factor 0.1, compactness 0.5). The segmentation of all levels is mainly (to 90 %) based on the spectral values of the satellite scene because in heterogeneous structured mixed forests the shape of objects can be obscured by neighbouring trees. The scale parameter of level 1 is chosen to define objects of the size of a tree crown diameter (mean size of objects: 10.5 m²), while level 2 is selected to characterise patches of tree groups (mean size of objects: 27.7 m²). The scale of level 3 is generated to delineate forest habitats similar in size to terrestrially mapped NATURA 2000 areas (mean size of objects: 103.7 m²).

Previous to the forest classification non-forest land uses, such as agriculture or urban area are masked, based on thresholds for shape, texture and spectral mean value of these classes. The classification of the forest types Beech (two age classes), Black Alder, Larch, Spruce (two age classes), Sycamore, and the forest-related objects Clearances and Afforestation is performed on single tree / small tree-group level (1) as nearest neighbour classification of the mean spectral values of the segments. A 30 per cent share of different tree types is defined as

[17] The named specifications are parameters of the multiresolution segmentation method of eCognition. The Scale Parameter determines the maximum allowed heterogeneity within an object. The shape factor defines the textural homogeneity of the objects. Compactness of objects is one of the parameters (the other is smoothness) for the shape factor.

mixed forest within an object. Therefore, the results of level 1 were aggregated to the tree-group patch level (2), where an object is assigned to a single species class if 70 per cent of the sub-objects are classified by one species. Mixed stands are assigned to a newly introduced group "Mixed deciduous" and "Mixed". The third level (Combined Patch Level) is used to improve the classifications of the sub-levels and to derive potential NATURA 2000 habitat types, such as Beech habitats, Alluvial forest habitats, or *Stellario-Carpinetum* habitats.

Shadowed areas are separately classified using a NDVI-threshold. To reduce the share of segments classified as shadow the class is differentiated. A second NDVI threshold is introduced, indicating shadowed vegetation. Then the shadowed segment is assigned to the class of the neighbouring forest object with the longest shared border.

## 2.4 Integration of class rules via fuzzy logic

The occurrence of different forest habitats depends on specific ecological and anthropogenic influences. These conditions allow or prevent species and habitats to exist. In the following these factors are referred to as silvicultural site conditions and natural site conditions. They can be related to geo-factors, which describe and attribute the ecological quality of a specific location (see Table 1). The probability of assignment for each object (values from 0 to 1) classified with the nearest neighbour (see 2.3) is combined with a fuzzy knowledge-base, which consists of silvicultural and natural site conditions.

According to the fuzzy logic theory, a fuzzy set for each class concerning each geo-factor is defined, containing input and output membership function. 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 formalised by fuzzy input membership functions. Subsequently, linguistic variables for all possible forest type classes are defined via membership output functions (e.g. "possible occurrence" or "limited occurrence" for the class "Sycamore"). Furthermore, fuzzy rules have been developed describing the relationship between each linguistic term of each variable and the degree of possibility of the class. As a result of this process, defuzzicated membership functions are derived for the geo-factors[18].

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

[18] An example of the fuzzy rules and a list of all possible rules for the utilised geo-data and forest classes are shown in Chapter V (not in the published manuscript).

be compared. The class with the highest membership will be assigned to the object (maximum – OR – rule, see Figure 2). Consequently, the lowest membership and the class assignment can be defined solely by a geo-factor.

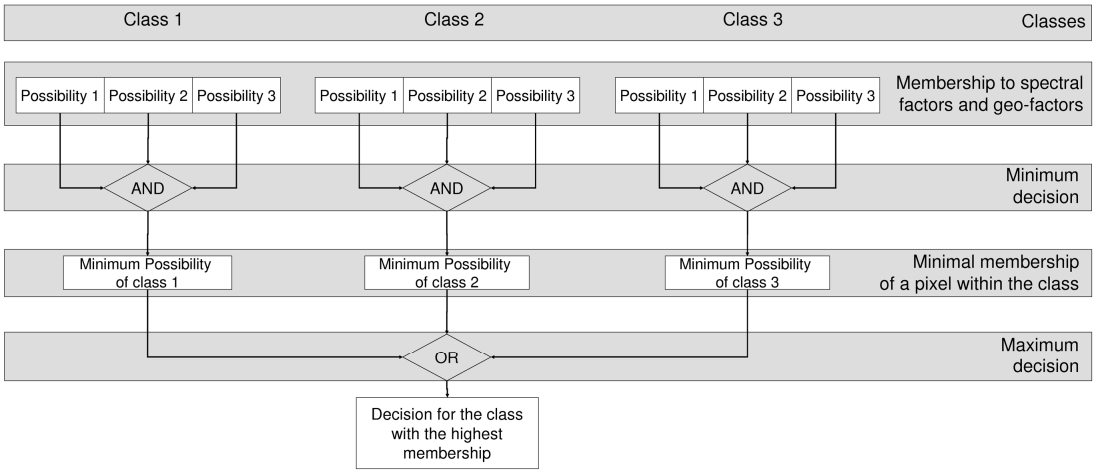



Figure 2: Schematic application of a fuzzy class decision with spectral classification and geo-factor possibilities

### 2.4.1 Natural site conditions

For habitat types which can possibly exist in the study area the list of geo-factors is used, consisting of slope, aspect, curvature, and altitude, soil type from a Conceptual Soil Map, and available water, soil substrate, and availability of nutrients from a Forest Site Map (see Table 1). For all main tree types the natural site conditions are developed based on knowledge of local experts and literature about natural forest associations in Bavaria (Walentowski et al. 2004; Walentowski et al. 2005). Especially for this kind of information the integration via fuzzy logic is useful, because there are often no sharp thresholds. The statements of local experts and literature sources are made in linguistic terms, such as: "*Sycamore can be found in higher pre-alpine regions, especially at very steep slopes*".

As schematically shown in Figure 2, each classified object consists of at least three different classification outputs. Normally, the class with the highest probability will be assigned to the object. The extended fuzzy-classification includes the possibility of occurrence of forest types within certain natural conditions into this classification process. An example for this classification procedure is an object which is classified as Beech "with a high probability". However, there is still a smaller probability to be assigned to Alder or Sycamore. These probabilities are combined with the natural site conditions. If the slope of the object is very steep, it is very unlikely for Beech and Alder to exist with these natural site conditions. In contrast, Sycamore is adapted to these factors and has a high possibility to exist. Of all compared classification probabilities and possibilities of natural site conditions, the lowest

value (on a scale from 0 to 1) is assigned to the possibility of existence for the class, because the most unacceptable factor is the critical value for the forest type to occur. In the case of a very steep object it is likely that the possibility of existence for Beech forest is lower after including ancillary data than the spectral probability of Sycamore and therefore the latter forest type is assigned to the object. Because the natural site conditions give very explicit information about species with narrow ecological niches, which cannot be distinguished by spectral values, these forest types are better recognisable with ancillary data.

### 2.4.2 Silvicultural site conditions

To include forestry practices, two approaches are used. Firstly, silviculturally preferred mixture types of forest stands in Germany were taken from literature (Jansen et al. 2002c) and extracted from Forest Organisation Maps (e.g. 65 % Spruce, 25 % Beech, 5 % Sycamore, 5 % Birch). Statistics of classified tree-type compositions are taken at segmentation level 3 (e.g. 60 % Spruce; 30 % Beech; 10 % Beech – young). If the dominant species and the second dominant species of such a classified mixture type do not differ more than 5 % to a silvicultural preferred mixture type, the sub-level (level 2) included these possibilities for the tree species of the mixture as one ancillary layer (e.g. possibility Spruce = 0.6, possibility Beech = 0.3) and included it in the minimum (AND) decision.

Another approach was undertaken to improve the classification accuracy of elder Spruce (more than 120 years) stands. In classification level 1 small clearances are classified with the nearest neighbour approach. The existence of clearances is used in level 2. If an amount of more than 10 % of the classes 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

For validation 121 objects of level 1, which cover 1.5 ha (0.23 %) of the study site, are chosen in a random-stratified order of occurring classes. These segments are compared to the Forest Organisation Map. This is done on pixel basis (0.6 m resolution), because the classified object can possibly intersect with validation polygons. The Forest Organisation Map is additionally compared to recent (2003) true colour aerial photographs, in case of occurring errors in this map. Level 2 includes two additional summarised classes (“Mixed” and “Deciduous Mixed”), which cover 12 % of the area. Because the process of taking samples is carried out in level 1, these classes are assessed in a separate evaluation realised by visually interpreting aerial photographs.

Firstly, in this chapter the results of the forest type classification of level 1 and level 2 are presented (3.1). In a second step it is shown that NATURA 2000 habitat types and the habitat qualities can be derived from the classified forest types (level 3).

## 3.1 Classification of forest types

The results of the accuracy assessment are shown in Table 2. Significantly higher classification accuracies can be reached with instead of without additional data. Especially the detection of species with a small ecological niche is improved. With a multispectral-based classification, a forest type such as Black Alder is not distinguishable from other deciduous forest while showing the highest classification accuracy with ancillary data. This is due to the influence of the natural site conditions, especially the geo-data and rules for the available water from the Forest Site Map and the curvature derived by the DTM. Other decisive factors can be the substrate (Larch) and the slope (Sycamore). Between level 1 and level 2 there is no advance in classification accuracy due to the introduction of mixture classes. As shown in the separate evaluation of these classes in Table 2, a better overall classification result could be expected, if level 2 were assessed collectively for all classes, especially for mixed forest.

Table 2: Accuracy assessment for forest types (level 1 and level 2)

| **Forest Type** | **Multispectral Classifcation with Ancillary Data** | | **Purely Multispectral-Based 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 |
| Mixed Forest | - | 0.90 | - | 0.85 |
| Deciduous Mixed F. | - | 0.55 | - | 0.45 |

Between the multispectral classification and the classification with ancillary information 18.6 % of the objects changed the class assignment in level 1. The decisive possibility of class assignment changed for 36.4 % of the objects when additional data are applied.

Therefore, it can be assumed that for the classification with ancillary data an amount between 18.6 % and 36.4 % of the objects are based on the factors of the natural site conditions. However, the interaction of classification results and the rule-base are complex as first tests of evaluating the influence of single natural site conditions have shown.

A further improvement could be made without differentiation of age levels (e.g. Beech – young). Natural site conditions are not suited for separation of the same species. Therefore, rule sets for silvicultural site conditions and age classes, similar to the usage of clearances, could be used.

## 3.2 Derivation of NATURA 2000 habitat types

For the classification levels 2 and 3 the attempt to obtain and assess NATURA 2000 habitat types and their qualities was undertaken. At the moment, this information is terrestrial mapped by field survey and combined to forestry management plans. For mapping forest 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 natural forest types are available for different habitat qualities. For the percentage of natural forest types the habitats are identified for:

- excellent quality (A) ≥ 90 % natural forest types,
- good quality (B) ≥ 80 % natural forest types, and
- medium quality I ≥ 70 % natural forest types.

Figure 3 shows the result of this approach exemplary for one type of Beech forest. Unfortunately it is only possible to a limited degree to derive habitat types automatically, due to parameters which cannot be detected by remote sensing (e.g. understorey vegetation), local knowledge of mapping, and political decisions (Förster et al. 2005a).

Moreover, the so-called orchard problem (Lang and Langanke 2006) arises[19]. 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.

[19] The described "orchard problem" is described before as "modifiable areal unit problem" in different context than nature conservation. Therefore, in Chapter VI the original description and reference is used (not in the manuscript).

Both the results and the segmentation levels shown in Figure 3 have certain advantages and disadvantages, but the approach shows that the mapping guide for German forest disregards the need for a spatial reference unit such as landscape levels or a minimum mapping unit. 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.

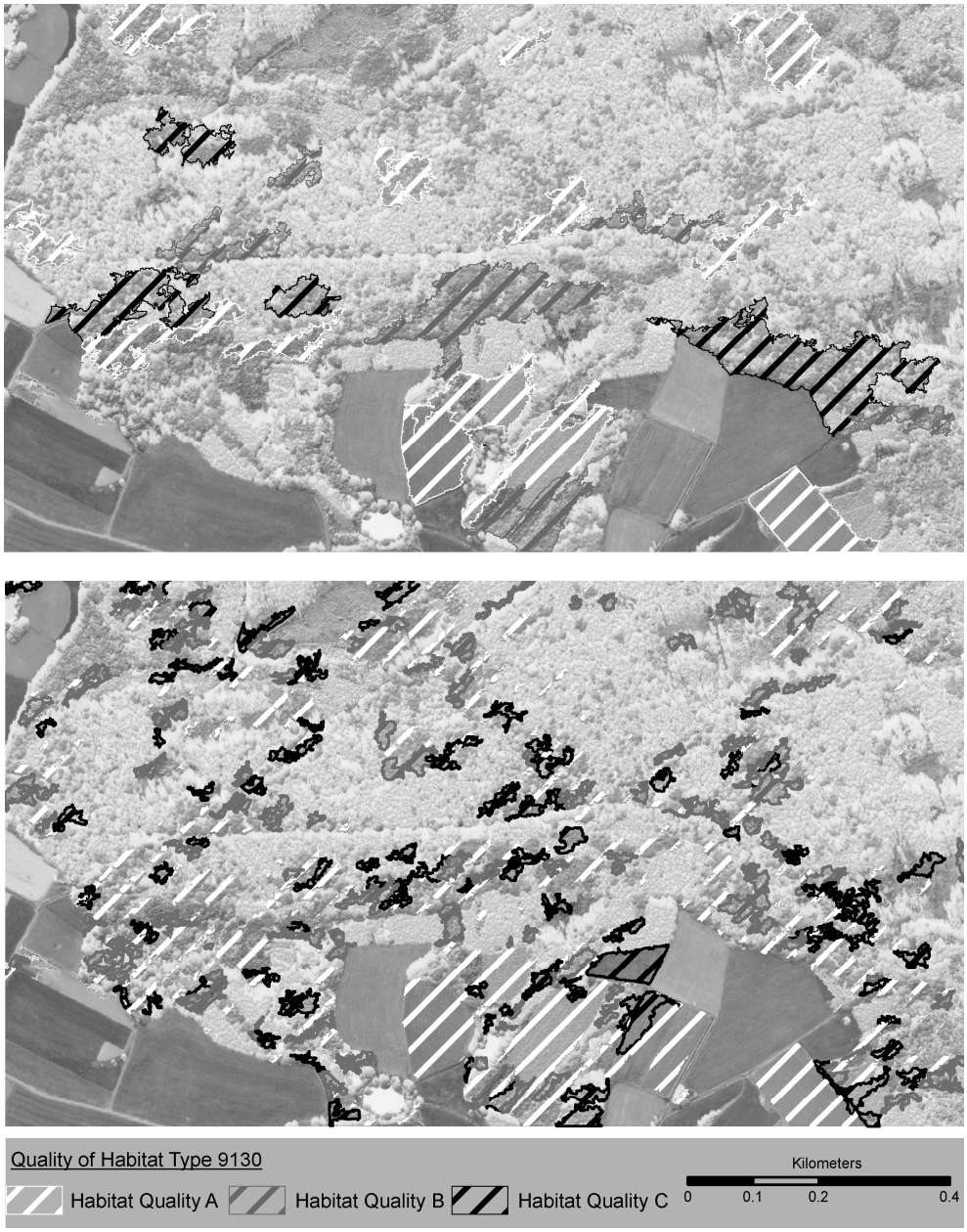



Figure 3: Derivation of different habitat type qualities for level 3 (upper figure) and level 2 (lower figure) for the special Beech habitat type Asperulo-Fagetum (NATURA 2000 nomenclature: 9130)

# 4 Discussion and outlook

The presented study shows that the classification accuracy of the investigated forest types is higher with ancillary information integrated by fuzzy logic than with a pure multispectral classification. In comparing the results with and without additional data an average increase of 13 % of the overall accuracy is detected. Especially the identification of forest types with narrow ecological niches, such as Black Alder, is significantly increased. However, forest types with wide ecological niches, such as Spruce or Beech, are classified with similar result (difference of not more than 10 %), when comparing the methods. Moreover, silvicultural site conditions are integrated into the classification process. However, two tested approaches have only a limited influence on the classification success (less than 1 % of the objects are classified differently). It can therefore be stated that natural site conditions are more relevant for the classification success than silvicultural site conditions.

In a second step, NATURA 2000 habitat types and the habitat quality "share of natural forest types" are derived from the forest-type classification. It is shown that the share of habitat qualities varies with an altering scale. For the two segmentation scales used in this study (level 2 and level 3) the fine scale defines smaller objects of a higher habitat quality, while the coarse scale defines larger areas of a medium habitat quality. At present there is no standard defining a spatial reference size (e.g. minimum mapping units) for the quality of biodiversity. This question should be addressed to ecologists and included in mapping guidelines. If a certain habitat requires a coherent large area a larger segmentation scale should be applied, while small-sized habitats should be classified with a finer object size.

Performing a classification using additional GIS-data provokes the question for consistent availability of these data. Within the state owned forests of Germany a good data basis already exists, especially with the information from the Forest Site Map. However, in private forests in Germany, geo-data is often not even available as analogous map.

A further improvement of this study can be obtained by a careful analysis of the dependence of accuracy on natural site conditions, integration of other additional data (such as LIDAR data), and a more efficient usage of silvicultural site conditions, e.g. by incorporating stand-density data. The presented method has to be transferred to forest types and habitat types of other areas, such as north-east Germany, to validate the reliability of the technique and become more generally applicable. Moreover, a comparison of different techniques of integrating geo-data into classifications, such as neural networks, could be useful for a quality assessment of integration techniques (Baltsavias 2004).

# Chapter IVb:
# Comparison of Classification Methods of NATURA 2000 Habitats with Additional Geo-Information

"Approaches to utilising QuickBird Data for the Monitoring of NATURA 2000 habitats"

*Community Ecology (in press)*

Michael Förster, Annett Frick, Helge Walentwoski & Birgit Kleinschmit



*ISSN: 1588-2756*

*received January 2008; revised July 2008*

*accepted August 2008*

## Abstract

The implementation of standardised methods for the monitoring of NATURA 2000 sites in Europe is still a key topic in environmental research. Effective, economically priced and, as far as possible, automated applications are required. Rapidly developing sensor technology together with advanced image processing methods offer new possibilities for application of remote sensing data to NATURA 2000 monitoring.

The studies presented here combine commonly available GIS data, such as Biotope Type Maps or Forest Site Maps with remote sensing classifications of the very high spatial resolution (VHSR) QuickBird sensor. Two knowledge-based approaches under inclusion of *a priori* object-based information are utilised to detect the extent of habitats as well as their quality according to the German NATURA 2000 mapping guidelines. While one method used a segmentation of forested sites in Bavaria (Southern Germany), the second technique applied available objects to classify heathland habitats in the Brandenburg Region (northern Germany). The results were subsequently compared, in close cooperation with local environmental authorities, with habitats mapped terrestrially for NATURA 2000 management plans.

These results indicate that different remote sensing methods can be a valuable support for terrestrial mapping. Woodland habitats can be detected and specific NATURA 2000 quality parameters (e.g. percentage of natural forest types) are recognisable. In the case of heath-dominated sites, terrestrial mapping can even be replaced by remote sensing of certain habitat types for which it is also possible to obtain adequate measures of quality. Having evaluated the quality of forest and heathland NATURA 2000 habitats, two general challenges when implementing the guideline regionally could be indicated. Firstly, the very general scope of the Habitats Directive contradicts to specific local protection purposes. Secondly, the protection aims given by NATURA 2000 areas are very static. The Directive could be improved by adapting existing management and conservation strategies to pro-actively respond on likely anthropogenic influences.


## 1 Introduction

The EU is committed to the protection of biodiversity; not least by a political commitment to halt biodiversity loss within the EU by 2010 (“2010 Biodiversity Target”), first adopted by EU Heads of State at the EU Summit in Gothenburg in June 2001. As part of the attempt to fulfil this objective, over the last 25 years the EU has built up a vast network of over 26,000 protected areas including all the member states and a total area of around 850,000 km², 

representing more than 20 % of total EU territory. This vast array of sites, known as the NATURA 2000 network, is the largest coherent network of protected areas in the world. At the 9th Conference of the Parties to the UN Convention on Biological Diversity (CBD) held from 19.-30 May 2008 in Bonn (Germany), worldwide attention was recently paid to the European NATURA 2000 network. The legal basis for the NATURA 2000 network derives from the Birds Directive (Council Directive 79/409/EEC on the conservation of wild birds), which dates back to 1979, and the Habitats Directive of 1992 (Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora). Together, these Directives constitute the backbone of the EU's internal policy on biodiversity protection.

As the selection of sites for the NATURA 2000 Network nears completion, attention is increasingly focused on the issue of management in accordance with the provisions of Article 17 of the Habitats Directive (92/43/EWG). As stated above, almost a fifth of the EU territory has to be supervised. The Directive requires that standardised monitoring of the habitat types be undertaken and a report on this submitted every six years. For this reason, an easily operated, economically priced and as far as possible automated application is required which can support cost-intensive terrestrial monitoring.

A traditionally used method is the terrestrial mapping of stratified samples from the monitoring-site. The full geographical extent of the species or habitat is extrapolated from these small scale measurements using modelling techniques. Therefore, the mapping is very detailed and often includes different search-spaces to facilitate up-scaling of the results (Stohlgren et al. 1997). However, terrestrial mapping is very costly, reducing the number of stratified samples obtained. Moreover, there is not always sufficient knowledge about appropriate extrapolation of the results, because of inadequate research on distribution patterns of species (Whittaker et al. 2005). A combination of terrestrial mapping with a top-down method, such as remote sensing classification, could prevent these difficulties arising.

Advances in spatial and spectral resolution of available sensors are now abundantly available. These sensors make it possible to directly measure certain aspects of biodiversity by remote sensing (Turner et al. 2003). While the detection of species still remains a difficult task, often solved with indirect approaches (Jobin et al. 2008; Nagendra 2001), promising possibilities for the classification of habitats with VHSR sensors seem to exist. The potential of remote sensing and Geographic Information System (GIS) techniques for identifying and monitoring of NATURA 2000 habitats according to the Habitats Directive are evaluated on a general basis by the EU projects SPIN (Langanke et al. 2004) and EON 2000+ (Sell et al. 2004). However, results for different habitats show considerable variation in accuracy of detection and monitoring. Especially Mediterranean habitats (Boteva et al. 2004), wetlands,

dry grasslands (Bock et al. 2005) and mires (Küchler et al. 2004; Langanke et al. 2007) have yielded good classification results.

Since the introduction of the Habitats Directive in 1992, the legislative as well as technical specifics have changed. Firstly, more precisely defined monitoring guidelines are now available at national and European levels (Burkhardt et al. 2004; European Commission 2006). In Germany a set of monitoring indicators was developed for different main types of habitats (e.g. forest, wetland, heathland). Taking forest as an example, this assessment matrix includes amongst other parameters the existence of different forest development stages, the quantity of biotope trees, deadwood, percentage of natural forest types, quality of understory herb layer, faunal quality, disturbances of soil, hydrology and forest structure, as well as disturbances caused by forest fragmentation. According to the guidelines of the EU Commission, the monitoring matrix distinguishes between a favourable, unfavourable-inadequate, and unfavourable-bad conservation status. For the example of deadwood, the German monitoring guidelines require the presence of more than three trees per hectare for favourable conservation status, while more than one deadwood object indicates unfavourable-inadequate status and less than one an unfavourable-bad conservation status.

Clearly, not all of the indicators of the assessment matrix are detectable by remote sensing methods. However, rapidly developing sensor technology and image processing methods offer new possibilities for use of remote sensing in combination with GIS for these types of NATURA 2000 monitoring. With the launch of VHSR satellite systems, such as IKONOS or QuickBird, the interpretation of remote sensing images has developed from a pure pixel-based semi-automatic classification, depending on the spectral information, into a more object-based classification taking into account vector or segment information (Blaschke and Strobl 2001). In Figure 1 the workflows of three different classification strategies are schematically shown for the example of single deciduous tree detection. A pixel-based classification (cf. Figure 1a) is based on the spectral values. Because the shape of the observed object is not taken into account, a so-called “salt and pepper” effect occurs, which can be reduced with filter techniques (Meinel et al. 2001). The results of pure pixel-based approaches are difficult to interpret for VHSR data, because the shape of objects is often fragmented and proximity relations (e.g. tree / shadow) can be lost (Burnett and Blaschke 2003).

The methods explained in this section are illustrated with two examples of possibilities for detecting habitat objects by resolving the complex data basis. An object-based classification (cf. Figure 1b) starts with the creation of vectors. After this segmentation, the objects are classified. Proximity and different sizes of object can be included in the classification process. A knowledge-based classification (cf. Figure 1c) carries out a pixel-based

classification, but resolves the results in an existing vector data-set, which can be edited to reveal significant changes in geometry.

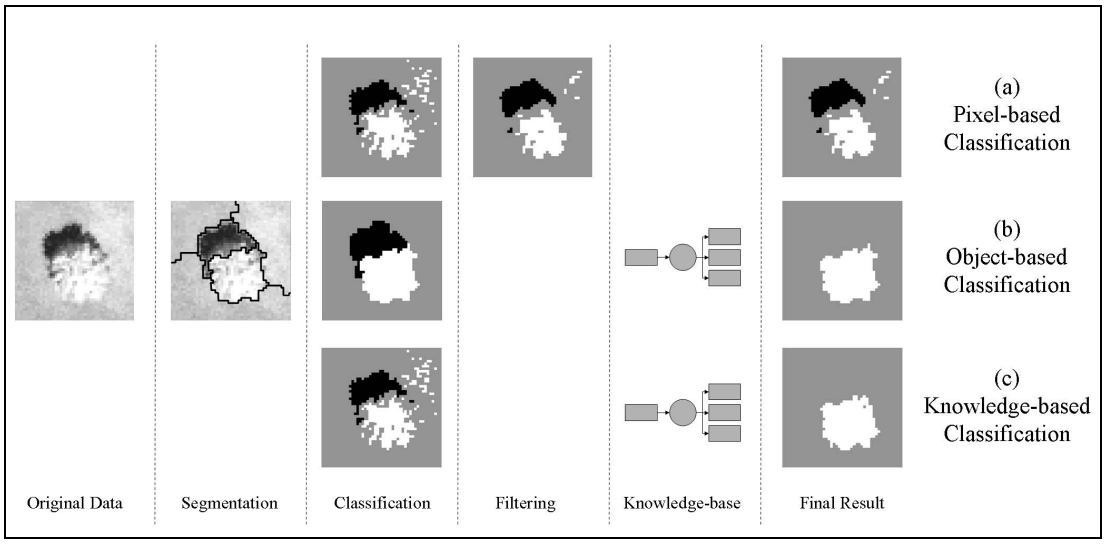



Figure 1: Schematic comparison of different classification strategies for the example of a single tree (*Fagus sylvatica* f. *Purpurea*; data source: QuickBird Image; near infrared band; 48 x 48 pixel)

The merging of classification results as objects (as in Figure 1b and Figure 1c) is a highly important step in the process of habitat classification, because the minimum classification unit of VHSR data is not identical with the classification mapping unit usually applied in, for example, NATURA 2000 mapping keys. The average size of a mapped habitat is much larger than the pixel size, therefore a combination of pixels to objects better reflects the target size of the habitats. To give an example, the average size of terrestrial mapped habitats in the study site "Angelberger Forst" is 7,350 m², while the QuickBird pixel size is approximately 0.37 m². Obviously, if a single vegetation type has to be detected (e.g. beech as a proxy for the habitat type *Luzulo Fagetum*) the spectral variance within the single pixels of the crown is less important than the shape of the tree crown (see Figure 1). Apart from this aggregation of pixels to form an object, or a habitat defined by a mapping key, the habitat can also consist of more than one detected object. A heathland habitat may be composed of a pre-defined percentage of vegetation classes like dry heath, grassland, scrub and open sand within a certain area (cf. Figure 6 and Figure 7). These examples of aggregation to an object (a) and subdivision of an existing vector (b) illustrate that although VHSR data have detection limitations (e.g. understory vegetation), the high spatial resolution necessitates simplification, as demonstrated below.

The objective of the studies presented here is to illustrate and compare the strengths and weaknesses of different methods for classifying VHSR data obtained with the sensor QuickBird. The approaches presented here combine spectral and textural information of

QuickBird images with ancillary data, to identify forest and heathland habitats. Firstly, a per-parcel-based hierarchical classification approach is used for the detection of heath. Secondly, an object-based approach is applied for the detection of forest habitat types.

## 2 Ecological background of the study sites

Four NATURA 2000 areas in Germany were selected as test sites (see Figure 2 for geographical location and Table 1 for area specification). The sites selected in the pre-Alpine area of Bavaria are mainly forested, while those in the north-eastern lowlands of Brandenburg are predominantly heathland. Apart from the interest of the regional environmental agencies in these sites, they were selected because they represent a wide range of NATURA 2000 habitats. Moreover, data suitable for validation are available for these areas in the form of terrestrial mapping results and existing management plans.

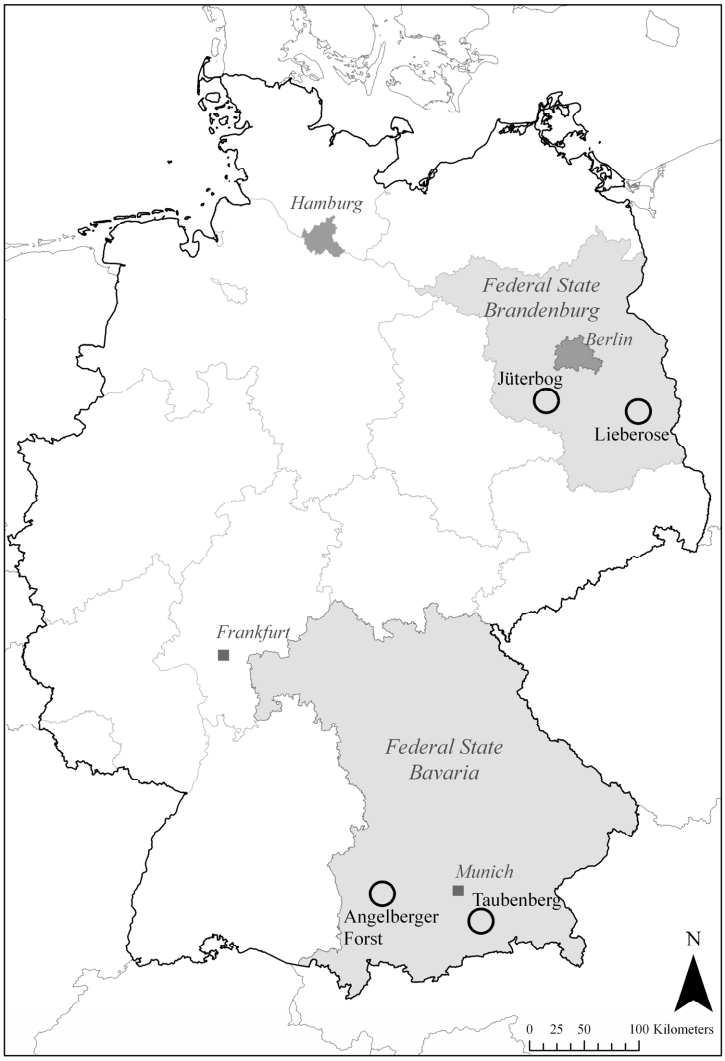



Figure 2: NATURA 2000 test sites in Germany

Table 1: Main characteristics of the test sites and the utilised QuickBird data. Environmental Quality is assessed on the portion of imagery not affected by clouds, but influenced by particles such as haze, fog, or smoke. The distributor of the QuickBird product divides the image quality in excellent (0 – 10 % imagery affected), good (11 – 25 %), and fair (26 – 50 %).

| Study Area | Area in km² | Acquisition date | Cloud cover | Environmental quality | Off Nadir angle | Land use characteristics | Existing NATURA 2000 habitat types | Predominantly existing species or communities |
|---|---|---|---|---|---|---|---|---|
| Jüterbog | 82 | 04 August 2003 | 0% | excellent | 14.6° | dry, open, forest | Dry sand heaths (2310)<br>Inland dunes (2330)<br>Oligotrophic to mesotrophic base poor standing waters (3130)<br>Natural eutrophic lakes (3150)<br>European dry heaths (4030)<br>Xeric sand calcareous grasslands (6120*) | *Caluna* and *Genista*<br>*Corynephorus* and *Carex*<br>*Littorelletea uniflorae and/or Isoeto-Nanojuncetea*<br>*Maganopotamions and/or Hydrocharitions* |
| Lieberose | 163 | 06 September 2004 | 0% | fair | 20.5° | dry, open, forest | | |
| Angelberger Forst | 7 | 11 August 2005 | 1% | fair | 11.3° | forest | *Luzulo-Fagetum* beech forests (9110)<br>*Asperulo-Fagetum* beech forests (9130)<br>Oak-hornbeam forest (9160)<br>Forest of slopes, screes and ravines (9180)<br>Alluvial forests (91E0)<br>Acidophilous Picea forests of montane levels (9410) | *Luzulo-Fagetum*<br>*Asperulo-Fagetum*<br>*Stellario-Carpinetum*<br>*Tilio-Acerion*<br>*Alnus glutinosa*<br>*Fraxinus excelsior*<br>*Vaccinio-Piceetea* |
| Taubenberg | 18 | 05 July 2005 | 0% | excellent | 7.0° | forest | | |

## 2.1 Bavarian sites

Two forested Sites of Community Interest (SCI) were chosen as test areas: Angelberger Forst and Taubenberg (Kleinschmit et al. 2006). The sites cover respectively approximately 650 ha and 1,847 ha. In both areas a wide variety of semi-natural, mixed forest types exists. The natural distribution of forest types primarily depends on soil moisture and acidity, but can be determined by relief or anthropogenic influences, too.

The **Angelberger Forst** (SCI no. DE-7829-301) is situated in the landscape "Donau-Iller-Lech-Platte" (D64) within the Middle Swabian Upland at altitudes from 580 to 650 m ASL. The natural vegetation is submontane deciduous forest. Tree species occurrence in the actual forest vegetation indicates a significant, human impact by cultivation of conifers. Today, the Angelberger Forst comprises 24 % broadleaved forest, against 66 % conifer-broadleaved forest (non-habitat quality). The actual broadleaved forest area is dominated principally by *Luzulo-Fagetum* (habitat type 9110: 17.9 %) and *Asperulo-Fagetum* (habitat type 9130: 3.5 %). The requirements of *Stellario-Carpinetum* (habitat type 9160: 1.3 %) are satisfied by periodically moist locations, but occur less frequently. Very moist habitats, mostly along streams, are covered by alluvial forests with *Pruno-Fraxinetum* and *Carici remotae-Fraxinteum* (habitat type *91E0: 2.0 %). Additionally, some small sites carry *Larix decidua* and *Acer pseudoplatanus*.

The **Taubenberg** is situated about 15 km north of the Alps close to Miesbach, Upper Bavaria, and covers an area of approx. 1,600 ha. The altitude ranges from 620 to 896 m ASL. The hill is formed from an alluvial fan of the "Obere Süßwassermolasse" (Upper Tertiary epoch). The basic material was deposited 10 – 15 million years ago, when the rivers transported debris from the uplifting Alps to their foothills and heaped up massive deltas. The unconsolidated fluvial sediments were transformed to conglomerates by diagenesis. The distinctive meso-relief is caused by soft clay marls lying between the hard conglomerate banks, visible in spring horizons and landslips. Where the impervious clay prevents a rapid infiltration of rainfall, the large surface flow has caused gully erosion and created ravines. The foot of the hill and its surroundings are covered by glacial moraines and river terraces of the Riss Era. Mean annual precipitation is around 1,500 mm (800 mm from May to September), mean annual temperature ranges from 6.2 to 7.6°C. The area has recently been designated as a protection area according to the Habitats and the Birds Directives of the European Union (SCI no. DE-8136-302). The Taubenberg region is an important drinking water reserve for the city of Munich. In accordance with the water protection function, forest management attempts to emulate natural processes. Clear-cutting is avoided. Large areas are in transition from pure spruce stands to mixed mountain forests of spruce, fir and beech.

*Abies alba* participates substantially in stand regeneration of many of these transitional stands. With regard to beta diversity of fir forests, the Taubenberg is of outstanding importance, not only for the "Alpine foothills" (nature unit D 66), but for the whole biogeographic region. A nutrient gradient from very poor to rich soils causes a remarkable species turnover within *Abies alba* forests at the local scale. Four different Silver fir associations were recorded (Walentowski et al. 2005). Of special relevance is the most extensive occurrence of *Myrtillus* type spruce-fir forests within the Bavarian Alpine foothills. They correspond to the spruce-fir forests of the Swiss Midland. Assigned to Annex-I-Habitat types the Taubenberg region is dominated by *Asperulo-Fagetum* (28.9 %), while *Luzulo-Fagetum* (1.9 %) occurs less frequent. It is important to note that *Asperulo-Fagetum* contains a considerable percentage of fir-mixture. The proportion of alluvial forests (7.9 %) is even higher than in the Angelberger Forst. Smaller areas (1 to 15 ha) are covered by *Tilio Acerion* (habitat type *9180), and acidophilous *Picea* forests (e.g. *Vaccinio-Piceetea*; habitat type 9410).

## 2.2 Brandenburg sites

**Jüterbog** – Forst Zinna/Keilberg (SCI no. DE-3944-301) and **Lieberose** – Endmoräne /Staakower Läuche (SCI no. DE-4051-301) are former military training areas with widespread open and dry habitats as well as large forest areas. The sites cover approximately 6,750 ha and 11,300 ha. Natural conditions on both sites were shaped during the glacial period. One of the last actively drifting inland sand dunes in northern Germany can be found in these areas.

In both areas, decades of intensive military exercising have caused mainly nutrient-poor biotope types to develop. The area is situated in a landscape subunit strongly influenced by glacial events and with very sandy soils. Military activities drove the deforestation of the area. Particularly, the repeated destruction of soil and vegetation cover by tanks and explosives helped to keep large areas open, allowing sandy dunes, heaths and dry grasslands with *Vaccinio-Genistetalia* and *Corynephoretalia canescentis* to evolve, which would not occur if the area had remained undisturbed. Biodiversity in those areas is very high and a large number of endangered species live there. Particularly species adapted to nutrient poor sites were able to colonise the degraded land. At landscape level, lowland heath occurs as a patchwork of fragments of varying sizes embedded in a matrix of other biotopes and managed land. Like other semi-natural biotopes, the heathland habitat is not stable, because of successional changes/dynamics and interactions with adjacent patches of different habitats. A continuous state of flux exists. The remarkable ß-diversity is caused by a mosaic of spatial niches and successional phases which provide distinct microhabitat conditions for

very specialised species. The main matrix of the habitat complex is made up of inland dunes with open *Corynephorus* grasslands (habitat type: 2330) and dry sand heaths as well as European dry heath with *Calluna* and *Genista* (habitat types: 2310 and 4030).

Since the sites were abandoned by the Russian army in 1994, natural succession threatens the non-forested habitats, because regular destruction of larger vegetation by military activities no longer takes place. The best way of preserving the existing species richness would be the regular removal of larger vegetation, as has been intensively discussed within the nature protection agencies of the federal state Brandenburg. However, there are few means of impeding the process because of the pervasive contamination with unexploded ammunition. Nevertheless, large areas still exist which up to now were not affected by succession. Hence, a regular monitoring of these NATURA 2000 sites is very important, because changes in the plant communities occur quite rapidly and unpredictably. Since the danger caused by explosive objects restricts a terrestrial monitoring over large areas, a remote sensing method is an especially valuable alternative in such cases.

## 3 Satellite data and classification methods

For the presented investigation four QuickBird data sets were acquired between 2003 and 2005 (see Table 1). The sensor's panchromatic band collects data with a 61 cm resolution at nadir with a multispectral (visible and near infrared) ground sampling distance of 2.44 m.

Different types of additional geo-data are used for specific tasks in the classification process (see Table 2). The topographic data of the German ATKIS (Authorative Topographic Cartographic Information System), biotope type and land use maps developed from interpretation of colour infrared (CIR) aerial photographs and habitat type maps from terrestrial survey are used for the knowledge-based approach. The object-based method used additionally a Digital Terrain Model (DTM 5 and DTM 25), soil information from a Conceptual Soil Map, forest specific land-use and soil information derived from a Forest Site Map as well as topographic maps and information from terrestrial survey. The geometrical quality of the data sources varies up to approximately one meter. Therefore, the geometrical accuracy was assumed to be sufficient for detection of natural vegetation. Both methods were evaluated with the help of CIR or True Colour aerial photographs and terrestrial survey.

Table 2: Additional data sources for the knowledge-based and the object-based classification methods. Further descriptions of the data can be found in (Frick 2007) for the knowledge-based method and (Förster and Kleinschmit 2008) for the object-based method.

| Data source | Knowledge-based | | Object-based | |
|---|---|---|---|---|
| | Date | Usage | Date | Usage |
| topographic data (ATKIS) and topographic maps | 2004 | image pre-processing | 2005 | image pre-processing |
| Digital Terrain Model (DTM) | | | 2004 | image pre-processing<br>fuzzy logic-based integration |
| biotope types and land use data | 1992-1993 | knowledge-base | 1996 | fuzzy logic-based integration |
| soil data (Forest Site Map – only Angelberger Forst Region), Conceptual Soil Map | | | 2004 | fuzzy logic-based integration |
| habitat type maps from terrestrial survey | 2003-2004 | evaluation of classification / knowledge-base | 2002-2004 | Evaluation of classification |
| stereoscopic CIR aerial-photographs | 1998 | evaluation of classification | | |
| True color aerial photographs | | | 2003 | evaluation of classification |

## 3.1 Knowledge-based classification

The complex nature of imagery with a very high geometric resolution requires advanced methods for analysis. Promising results have been achieved with knowledge-based approaches (Friedl and Brodley 1997; Peddle 1995). The method can be adapted to such different domains and objectives as, for example, the classification of agricultural fields (Cohen and Shoshany 2002), the detection of urban objects (Gerke 2002) or the delineation of areas of peat extraction (Pakzad 2001). Knowledge can thus be gathered from existing data. The inclusion of *a priori* information in the form of geo-data has been applied very successfully to solving various problems and improving classification algorithms (Eiumnoh and Shresta 2000; Maselli et al. 1995; McIver and Friedl 2002). It seems important to note that several studies show that the use of only one data source often leads to an unsatisfactory classification accuracy (Hahn and Baltsavias 1998).

The aim of this part of the study was to develop a method for the analysis of heathland habitats with Leica Geosystems ERDAS Imagine. The structure of the knowledge-based method is shown in Figure 3. This classification procedure integrates *a priori* information (e.g. biotope type and land use maps) in a rule-base. Firstly, the images are pre-processed, including a conversion to spectral radiance, georeferencing and pan-sharpening (Zhang

2002). As a second step a pre-structuring of the image was performed, which is based on ratios (e.g. NDVI), principal components, and textural information. This kind of information is derived from the original images and is subsequently included in the rule-base together with information from ancillary data (see Figure 4). Therefore, three different types of rules were taken into account:

- formalised experience of the human interpreter (e.g. trees have shadows)
- spectral characteristics from a large set of samples and the pre-stucturing of the image
- additional land-use data (e.g. biotope type and land use maps)

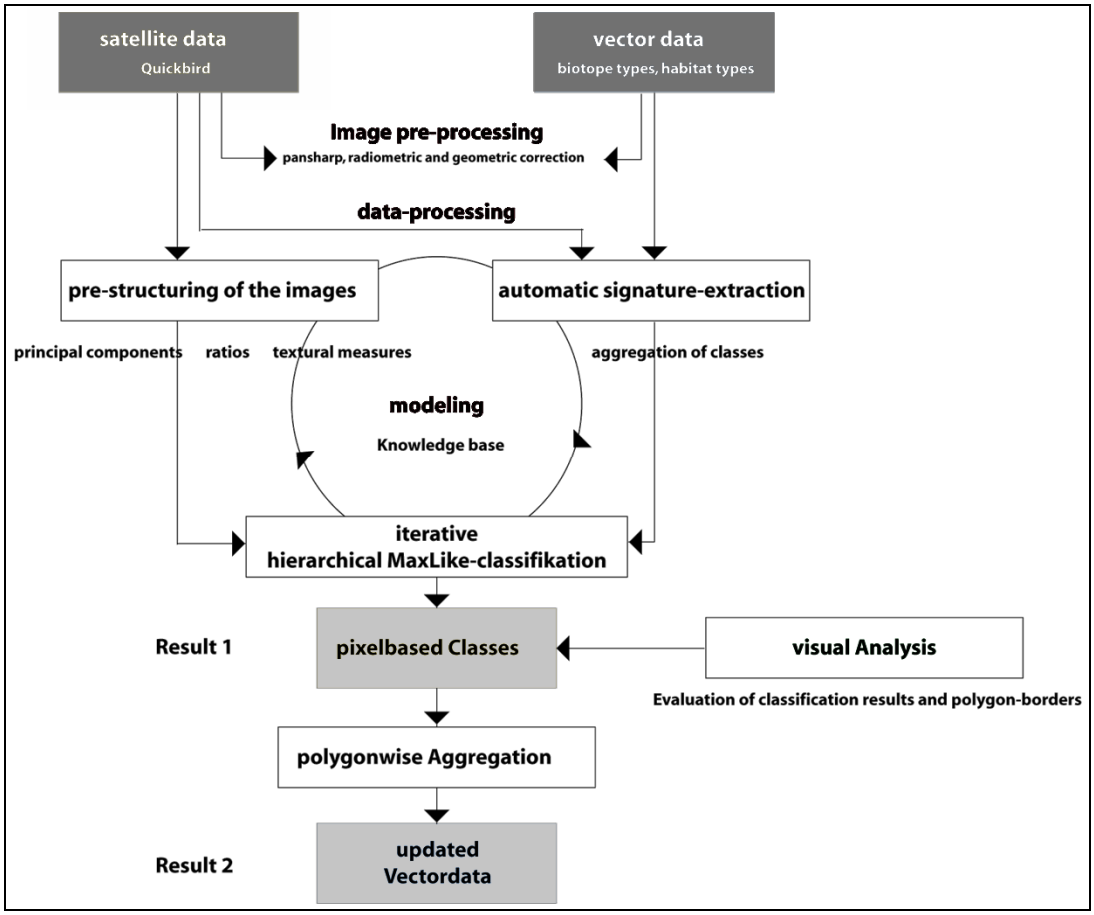


Figure 3: Process scheme of the knowledge-based classification (Frick 2005)

The knowledge-base explained above is not used for classifying the whole image, but to select signatures for training areas (automatic signature extraction in Figure 3). With these extracted signatures, the image and derived information are classified either with a Maximum Likelihood or an ISODATA algorithm (iterative hierarchical MaxLike classification in Figure 3). Pixel-based classes of distinct vegetation units are obtained as a first result. Habitat borders are then derived through visual interpretation of the satellite images or taken from the first terrestrial inventory (see Figure 6). This vector-data was then

integrated in the pixel-based classification results. This second result can be used for evaluation of the conservation status of a chosen area according to the German NATURA 2000 mapping guidelines.

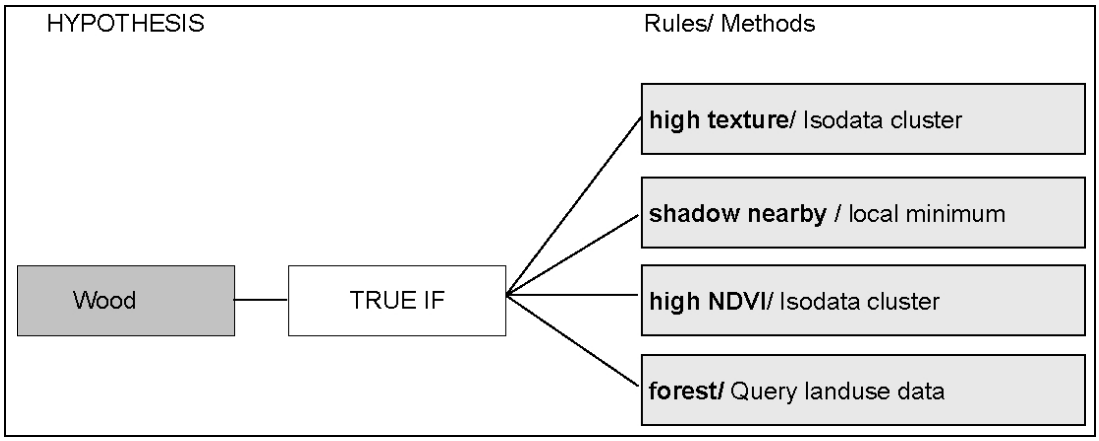



Figure 4: Example for a rule-base to detect training areas for signature extraction

A limitation of the automated signature extraction is that certain sub-classes not present in *a priori* data and/or in the rules can not be detected. For instance, if no water habitat was included in the *a priori* data, no signatures would be extracted for the class water. These limitations can be reduced by the integration of visual image analysis (see Figure 3).

## 3.2 Object-based classification

An object-based approach is applied to the detection of forest habitat types. Object-based classification is a relatively new method based on the segmentation of the image into polygons that are homogeneous with regard to spectral or spatial characteristics (Jensen 2005). An advantage of object-based classification is that the objects can be detected by their texture and pixel spatial continuity, e.g. with reference to those of neighbouring objects (Burnett and Blaschke 2003). Furthermore, when they include segmentations at different levels of detail, object-based methods can represent classes at different landscape scales and utilise these levels to increase classification precision. For the present study the satellite data was processed with a multi-scale segmentation method (Benz et al. 2004) in an object-oriented approach, using the software Definiens Professional (Baatz and Schäpe 2000). This method produces good visual results and low over-segmentation compared to other segmentation approaches (Neubert et al. 2008b).

The classification process is shown in Figure 5. In a first step, a geometric correction and the pan-sharpening of the original data to a resolution of 0.61 m is undertaken (Zhang 2002). The object-based classification performs a hierarchical segmentation of the QuickBird images, following a bottom-up approach that defines objects sequentially (Figure 5).

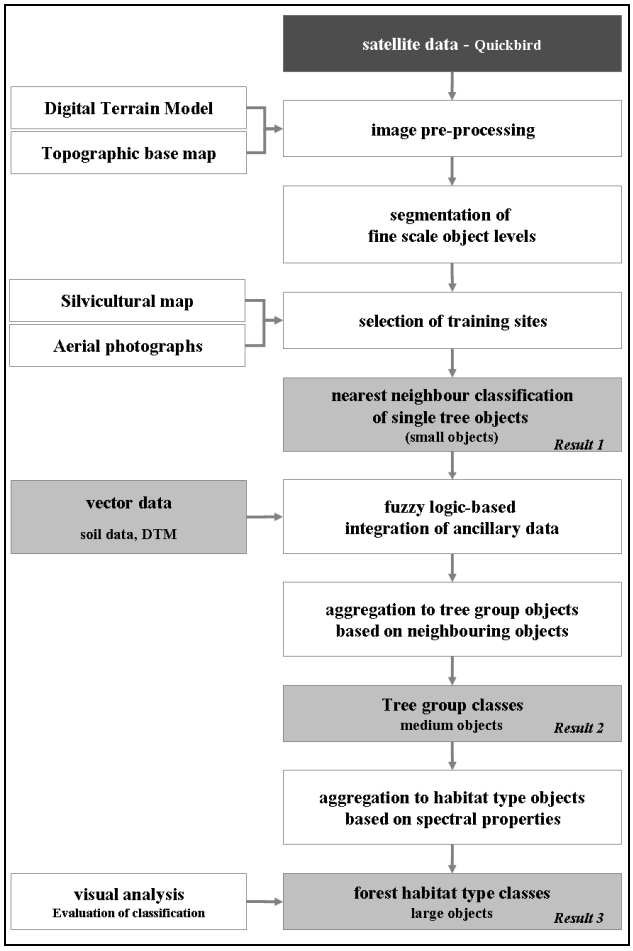



Figure 5: Process scheme of the object-based classification

The process commences with segmentation at the finest scale (Single Tree / Small Tree Group Level – see Table 3) followed by a classification procedure before the next level of segmentation, which defines larger objects (e.g. tree groups at medium scale).

Table 3: Segmentation Parameters developed to identify forest habitats[20]

| Segmentation level | Scale | Color | Shape | Compactness / Smoothness | Medium object size (in m²) |
|---|---|---|---|---|---|
| 1: small objects (single tree / small tree group) | 15 | 0.9 | 0.1 | 0.5/0.5 | 10.5 |
| 2: Medium objects (tree group) | 40 | 0.9 | 0.1 | 0.5/0.5 | 27.7 |
| 3: Large objects (habitat size) | 150 | 0.9 | 0.1 | 0.5/0.5 | 103.7 |

[20] The Definiens Professional segmentation parameter "Scale" is a dimensionless quantity. According to the guidelines of Definiens Professional are the not further defined parameters "Color" and "Shape" percentages complementing to 1.0. "Compactness" and "Smoothness" are sub-parameters of "Shape" which complement to 1.0 (not in the manuscript).

To enhance the accuracy of classification, additional vector data are included (see Table 2) by means of fuzzy-rules, which are combined with the nearest neighbour classification, based on the selection of training sites, using a fuzzy knowledge-base (Stolz and Mauser 1996). The occurrence of different forest habitats depends on specific ecological and anthropogenic influences. These conditions allow or prevent existence of species and habitats. They can be related to geo-data, which describe and quantify the ecological quality of a specific location. The probability of assignment for each object (values from 0 to 1) classified with the nearest neighbour is combined with a fuzzy knowledge-base. In combining the fuzzy sets and the hierarchical nearest neighbour classification results of the segmentation level 1, the approach uses the minimum (AND-) rule, which specifies that the most unacceptable factor is the critical value for the occurrence of the forest type. In a next step the minimum possibility of each possible class is compared. The class with the highest membership will be assigned to the object (maximum – OR – rule). The extended fuzzy-classification includes in this process the possibility of occurrence of forest types within certain natural conditions.

A 30 per cent share of different tree types within an object is defined as mixed forest. Therefore, the results of level 1 were aggregated to the tree-group patch (level 2), where an object is assigned to a single species class if 70 per cent of the sub-objects are classified as one species. Mixed stands are assigned to a newly introduced group “Mixed deciduous” and “Mixed”. The third level (Combined Patch Level) is used to improve the classifications of the sub-levels and to derive potential NATURA 2000 habitat type classes, such as Beech habitats, alluvial forest habitats, or *Stellario-Carpinetum* habitats.

# 4 Results

## 4.1 Knowledge-based classification

The knowledge-based classification approach described here is applicable for the evaluation of heath habitats. The derived habitat classes can be grouped satisfactorily to form indicators, in accordance with the prescribed national quality standard. Similarly to the example for forest given in the introduction, the indicator set for European dry heath consists of area size, wood cover, percentage of open sandy spots, percentage of moss covered area, percentage of grass covered area and percentage of heath covered area. Figure 6 shows the classification results for a reference habitat on the study area Jüterbog. According to the German assessment matrix for the category „completeness of typical structures” the conservation status of the habitat used in this example is in favourable condition, because areas of tree

cover and grass are quite small, but conversely the conservation status is inadequate because heath forms a compact layer and only 4.1 % of the area is covered by the class „open sand”.

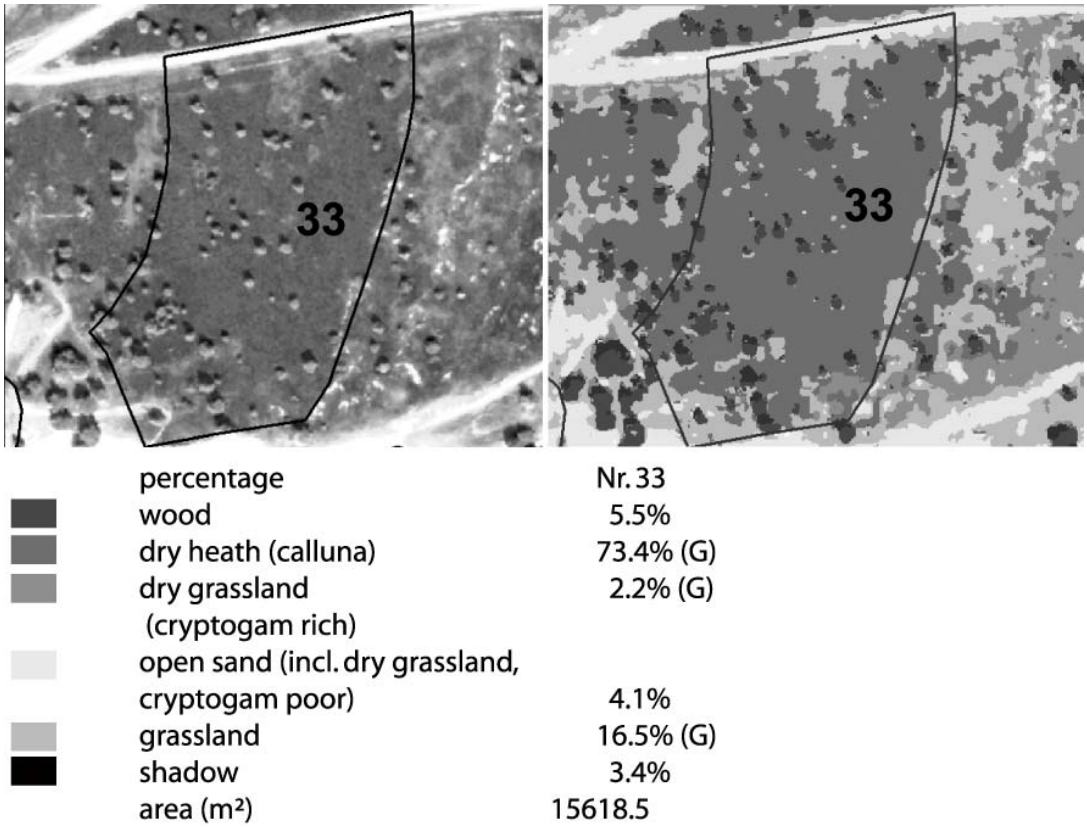



Figure 6: Classification results for the habitat type European dry heath (4030) in Jüterbog. The area borders are taken from the first terrestrial inventory (left: QuickBird subset, pansharpened; right: classification result overlaid with habitat borders).

The main problem inherent to all remote sensing applications still remains, that classes overlaid by others cannot be classified and quantified. Since the terrestrial NATURA 2000 inventory also considers overlapping layers, e.g. if open sand or moss is hidden beneath the dwarf shrubs, the total sum can add up to far more than 100 % cover. The terrestrial inventory of certain indicators therefore can not be compared directly to a satellite or aerial photograph-based inventory. To overcome this limitation at least roughly, the following assumption is made: the classes present within the habitats occur with the same distribution in wooded areas, hence the percentage of non-forest classes is interpolated (marked in Figure 6 with G).

The transferability of the classification procedure and the knowledge-base was tested for the same habitat types on the study area Lieberose. The imagery has quite different characteristics than for Jüterbog; the viewing angle is larger and the resulting image quality was not as good as for the Jüterbog-area. The acquisition was made in September whereas

Jüterbog was captured in August. The classification results show the main problem in Lieberose: the heath is giving way to spreading, grass-covered areas or is being overgrown by woodland.

The accuracy assessment is realised with 525 points for Lieberose and 358 points for Jüterbog (see Table 4). Most of the classes were evaluated by visual examination of CIR aerial photographs. The points used for assessing the classes were selected with a stratified random sample, ensuring that all classes are assessed, even if they are only classified as a small percentage of the overall area, while the classes occurring more often are assessed with a greater number of points, proportional to their abundance. Classes that can only be assessed by terrestrial investigation (e.g. *Corynephorus* grassland) were evaluated with recent maps used for terrestrial survey in NATURA 2000 monitoring.

Table 4: Accuracy assessment for the habitat type European dry heath (only a subset of vegetation classes is shown, e.g. water is omitted since it does not occur at this site)

| **class** | **Jüterbog** | | | **Lieberose** | | |
|---|---|---|---|---|---|---|
| | **Prod. Acc.** | **Users Acc.** | **No. of Points** | **Prod. Acc.** | **Users Acc.** | **No. of Points** |
| open sand | 1.00 | 0.88 | 16 | 1.00 | 0.95 | 21 |
| nutrient-rich grassland | 0.93 | 0.83 | 30 | 1.00 | 0.77 | 44 |
| dry grassland , cryptogam-poor | 0.85 | 0.90 | 19 | 1.00 | 1.00 | 21 |
| dry grassland , cryptogam-rich 1 | 0.89 | 0.89 | 27 | 0.83 | 0.95 | 21 |
| dry grassland , cryptogam-rich 2 | 0.88 | 0.83 | 18 | 0.80 | 0.80 | 20 |
| dry sandy heath (calluna) | 0.95 | 0.95 | 21 | 1.00 | 0.92 | 26 |
| wood | 0.92 | 1.00 | 23 | 1.00 | 0.91 | 82 |
| sealed | 0.78 | 0.82 | 22 | 0.95 | 0.95 | 20 |
| **Overall Accuracy** | **0.91** | | | **0.88** | | |
| **Kappa** | **0.90** | | | **0.87** | | |
| **P** | **< 0.00001** | | | **< 0.00001** | | |

With an overall accuracy of approximately 90 % (Jüterbog = 91 %, Lieberose = 88 %) the method shows results which recommend its further use for NATURA 2000 monitoring. Especially woody structures and open sand are detectable at above average accuracy rates. As can be expected, the differentiation of the two types of grassland (e.g. cryptogam-rich and cryptogam-poor) is difficult, because the spectral and textural features of these classes are very similar. The slightly lower accuracy rates of the Lieberose area could be related to the poorer quality of the image. However, the method is transferrable to this region. The classification accuracy obtained permits the evaluation of the conservation status for most of the heathland indicators defined by the German NATURA 2000 mapping guidelines at an adequate level of quality (as explained for the example of Figure 6 above).

## 4.2 Object-based classification

As a first step the different tree type classes within the forested NATURA 2000 areas were evaluated. For validation purposes 121 level 1 objects in Angelberger Forst and 82 level 1 objects in Taubenberg were chosen within a random-stratified scheme of occurring classes. The segments were compared with the Forest Organisation Map. Additionally, to check for errors in this map it was compared to recent (2003) true colour aerial photographs. The results of the accuracy assessment are shown in Table 5.

Table 5: Accuracy assessment for tree types classified with the object-based classification for the NATURA 2000 areas Taubenberg and Angelberger Forst. The No. of objects is quoted, but the accuracy assessment was calculated on a grid-cell basis.

| **Class** | **Angelberger Forst** | | | **Taubenberg** | | |
|---|---|---|---|---|---|---|
| | **Prod. Acc.** | **Users Acc.** | **No. of objects** | **Prod. Acc.** | **Users Acc.** | **No. of objects** |
| Beech | 0.81 | 0.78 | 37 | 0.77 | 0.70 | 19 |
| Beech – young | 0.32 | 0.31 | 19 | 0.29 | 0.26 | 11 |
| Spruce | 0.75 | 0.74 | 22 | 0.78 | 0.84 | 15 |
| Spruce old | 0.33 | 0.32 | 9 | 0.28 | 0.27 | 6 |
| Black Alder | 0.96 | 1.00 | 8 | 0.86 | 0.92 | 5 |
| Afforestation | 0.95 | 1.00 | 11 | 0.88 | 0.93 | 8 |
| Larch | 0.93 | 1.00 | 9 | - | - | - |
| Sycamore | 0.85 | 0.88 | 6 | 0.73 | 0.79 | 4 |
| Fir | - | - | - | 0.63 | 0.77 | 9 |
| Picea | - | - | - | 0.80 | 0.77 | 5 |
| **Overall Accuracy** | **0.77** | | **121** | **0.71** | | **82** |
| **Kappa** | **0.71** | | | **0.68** | | |

With an overall accuracy between 70 and 80 % (Angelberger Forst = 77 %, Taubenberg = 71 %), the method shows results which are slightly less satisfactory for NATURA 2000 monitoring than the classification of heathland classes. Especially classes which are adapted to very small ecological niches and are not abundant in this region (e.g. Black Alder) are detected at above the average accuracy rates. These encouraging results are mainly due to the incorporation of additional soil and terrain information. Therefore, the slightly lower classification accuracies of the Taubenberg area can be explained by the absence of one of the soil maps (Forest Site Map). Nevertheless, the method proved to be transferable. As for the heathland habitats, the classification accuracy obtained makes it possible to evaluate the conservation status for NATURA 2000 indicators of forest habitats.

For the segmentation and classification level 3 (see Table 3), NATURA 2000 habitat types and their qualities were assessed. This is shown with an example for a *Luzulo-Fagetum* habitat (9110, see Figure 7). According to the German evaluation scheme for this category, parameters such as number of forest development phases, number of biotope trees per ha, amount of dead wood per ha, or percentage of natural forest types have to be evaluated. To give an example, for the parameter "percentage of natural forest types" the habitats are identified as:

- favourable conservation status (A) ≥ 90 % natural forest types,
- unfavourable-inadequate conservation status (B) ≥ 80 % natural forest types, and
- unfavourable-bad conservation status I ≥ 70 % natural forest types.

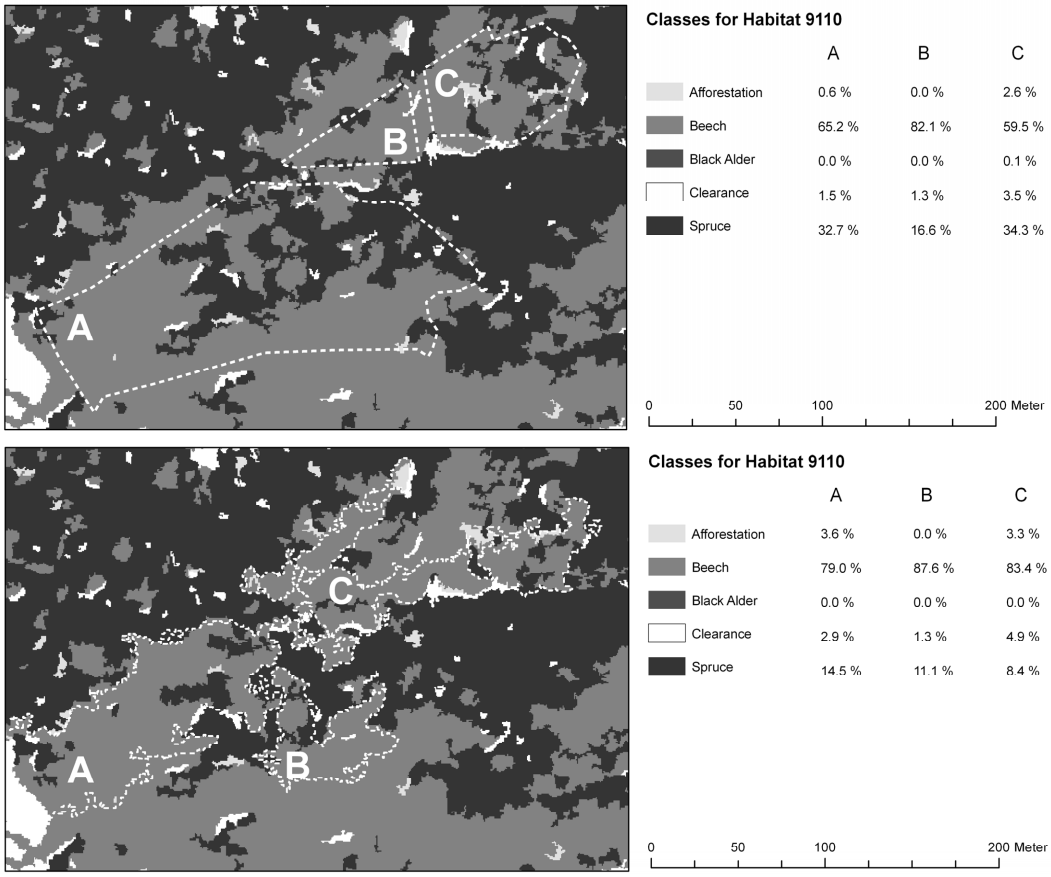



Figure 7: *Luzulo-Fagetum* habitat – exemplary subset for three reference habitats on the study area Angelberger Forst. The percentage of classified tree types is shown for terrestrially mapped habitats (upper image) and for segmented habitats (lower image) 70 % of typical tree types are required for defining a habitat according to this parameter.

Similarly to the heathland habitats, it is only possible to a limited degree to derive habitat types automatically, because of existence of parameters which cannot be detected by remote sensing (e.g. understory vegetation). Moreover, it is less likely that privately owned forests will be mapped as NATURA 2000 habitats than those that are state owned, even if the former are equally suitable as nature protection areas. These merely administrative decisions may play a role when defining habitat types, but obviously can not be detected with remote sensing (Förster et al. 2005a).

If the shape of the polygon of the detected habitats in level 3 is compared to the borders mapped terrestrially, the polygons are not located at the same place (see example in Figure 7). This is obviously due to the different mapping methods (remote sensing and terrestrial). The areas delineated with the object-based method are more detailed, differently structured, and the detected areas are smaller. The finer delineation of the object-based process results in a higher share of the main tree type in the polygon. Hence, even the assessed quality of the classified habitat objects can differ from terrestrially mapped ones (e.g. polygon A and C). There are different reasons for this observation. Terrestrial mapping (especially in forested areas) might face difficulties in estimating the actual extent of an area, especially when the criteria are related to the size of a habitat (e.g. percentage of natural forest type). Moreover, transitions between habitat types are rarely clearly detectable as a habitat border. Furthermore, terrestrial mapping tends to include large areas, even if a target species covers only a small percentage of an area. This is due to the potential development of habitats, which can be seen in the field (e.g. amount of shoots) but not detected by remote sensing.

# 5 Discussion and outlook

Two methods based on remote sensing to evaluate NATURA 2000 habitats for the purpose of an objective monitoring are presented for forest and heath areas in Germany. Both methods achieved good results in classifying habitats and land uses. It is also possible to use these methods to provide information about the quality of habitats and their alteration through time. These results indicate that different remote sensing methods can be a valuable support for terrestrial mapping. In the case of heath-dominated sites even a substitution of terrestrial mapping by remote sensing for certain habitat types and their qualitative attributes is possible. Both methods are therefore capable of improving existing monitoring approaches, because they can be applied objectively and in a standardised manner.

## 5.1 Comparison of the methods

Because the target habitats and defined classes of the methods are very different, it is not possible to compare these directly. However, Table 6 gives a comparative overview of the two techniques.

Table 6: Comparison of the usability of the knowledge-based method and the object-based method for the monitoring of NATURA 2000 areas.

| **Criteria** | **Knowledge-based method applied in heathland habitats** | **Object-based method applied in forest habitat** |
|---|---|---|
| Classification Accuracy | 88 % to 91 % | 71 % to 77 % |
| Amount of indicators of the monitoring guideline which are fully detectable with remote sensing | 8 indicators for habitat 2310<br>8 indicators for habitat 2330<br>8 indicators for habitat 4030 | 3 indicators for habitats 9110 and 9130<br>4 indicators for habitats 91E0, 9410, 9160, 9180 |
| Transferability | possible for all lowland dry heathland habitats | Possible for submontane forested areas |
| Utilisation of additional data | required | Especially soil data required – increasing classification accuracy up to 13 % (see Förster and Kleinschmit 2008) |
| Basic geometry of the habitat type | Usage of existing defined habitat borders / geometry or visual interpretation | Definition of new borders or extend of a habitat type with delineated segments possible |
| Operational usability in the regional environmental agency | Fully operationally used in the federal state of Brandenburg | Applied as a case study in test areas in the federal state of Bavaria |

Although the classification accuracies are shown, it is very difficult to draw conclusions from these, because it is a challenging task to detect different forest types, especially in a mixed deciduous area. Nevertheless, the accuracy of the knowledge-based method makes it very suitable for efficient monitoring and it is a technique which also permits the detection of a high number of indicators. One difficulty for the object-based method is that the indicators for forest habitats are defined less quantitatively than qualitatively, making it more complicated to derive these by remote sensing. Additionally, in forest habitats some indicators are related to the understory vegetation (e.g. quality of the forest ground vegetation), a problem which exists somewhat less in heathland habitats. Both methods are transferable without problems within the range of the classified habitats, as shown by successful application in two different areas. It is important to stress that for both techniques additional geo-data and information are crucial for classification success. Depending on the

habitats, different types of geo-data might be important. While soil data is very important for the classification of pre-alpine forest habitats (Förster and Kleinschmit 2008) existing biotope maps are required for heathland habitats. For monitoring purposes the existing and terrestrially defined habitat borders are normally utilised. Both methods work quite well with these spatial specifications (see Figure 6 and Figure 7). However, the object-based method can define new extents of a habitat type by means of its delineated segments. The derived segments have a spatial extent similar to the terrestrial ones, but a more detailed granularity, which gives a better expression of the habitat's borders. Therefore, the object-based method could provide useful support in identification and monitoring of a habitat's size. This would greatly support the terrestrial delineation of habitat areas and could increase efficiency of monitoring.

A further improvement of the methods used in this study can be obtained by the integration of other remote sensing data, such as LIDAR or hyperspectral information. The methods which are described should be applied to other habitat types in other areas, especially in other biogeographical regions, to validate the reliability of the technique and improve its general applicability. Moreover, a comparison of other techniques of integrating geo-data into classifications, such as neural networks, could be useful for a quality assessment (Baltsavias 2004)

## 5.2 Ecological implications of the results for heathland and forest

The knowledge-based procedure has been successfully validated for the habitat types **dry heaths** and inland dunes, using results from many areas, and is now used operationally in the German federal state of Brandenburg. The application of VHSR satellite data and the classification procedure developed here can substantially support the monitoring of heath habitats. It is possible to automatically derive indicators for assessing their structure and existing impairments. The habitat borders have to be verified visually. The automated classification has the advantage that large areas can be assessed in a short time and the accuracy and stability of the visual interpretation of habitat borders can be increased. Percentages of important indicators can be calculated at a high level of detail. This represents a basis not only for the monitoring of the effects of conservation measures, but also for their detailed planning.

The classification of the study sites indicated a distinct decline in the quality of the habitats inland dunes (2330) and European dry heaths (2310 and 4030). For European dry heath, the conservation status is inadequate in most of the defined habitat areas with respect to the indicators "share of open soil" and *cryptogam*-poor grassland, present on less than 5 % of the

area. Moreover, the cover of woody structures and trees is increasing (often above 25 %), which also indicates inadequate conservation status. Under such conditions the still abundant species *Calluna* and *Genista* are threatened. For inland dunes the situation is slightly less difficult. The percentage of open sandy areas is still much more than 10 % of the total habitat area (a criteria for a good conservation status). But here too, the cover of trees and bushes is increasing above 10 % in some areas, an indicator of unfavourable conservation status.

For **forest**ed areas NATURA 2000 habitat types are derived from a tree-type classification. It is shown that the share of habitat qualities can differ significantly from the results of terrestrial mapping. In comparison to terrestrial mapping, the object-based approach delineates the areas in much greater detail. At present there is no standard which defines a spatial reference size (e.g. minimum mapping units) for the quality of biodiversity. This question should be addressed by ecologists and included in mapping guidelines. If a certain habitat requires a coherent large area, a larger segmentation scale should be applied, while small-sized habitats should be classified with a finer object size.

The NATURA 2000 forest habitat types, especially the indicators “amount of woodland development types” and “percentage of natural forest types” are well suited to detection with remote sensing methods. For the two study areas the classification results show that for habitat types with more distinctly defined ecological niches, such as *Pruno-Fraxinetum* habitats, the percentage of natural forest types is above 90 %, indicating a favourable conservation status. Habitat types covering large areas, such as *Asperulo-Fagetum* beech forests, are more often in an unfavourable condition for conservation purposes, containing less than 80 % of beech. However, for these large areas, more than three different woodland development types exist, which conversely indicates a favourable conservation status. An important consideration in forested European NATURA 2000 areas, is that *Abies alba* forests, which are particularly well represented in the Taubenberg region, should be recorded carefully as special habitats. Because of their partly site-determined, ecologically transitional character between temperate Beech forests (habitat type 9130) and boreal Spruce forests (habitat type 9410), this forest type is at risk of being neglected in the European network NATURA 2000: The Habitats Directive does not list *Abies alba* forests as a distinct habitat type according to Annex I. How should they then be dealt with in protected areas? It is certainly not acceptable to classify them as Non-EU-habitat and thus equate them to coniferous plantations. Alternatively, they could be assigned to the annex I-habitat types 9110, 9130 and 9410, which appear to be closest in synsystematical terms (Walentowski et al. 2005).

Having evaluated the quality of some forest and heathland NATURA 2000 habitats, two general problems with the implementation of the NATURA 2000 guidelines became obvious.

Firstly, the very general scope of the Habitats Directive contradicts the specific objectives of regional nature conservation. As one example, the unavailability of a protection measure for *Abies alba*, which has a very important place in regional ecology, is highlighted above. NATURA 2000 protection status is often used to support regional protection status. Therefore, the strict guidelines for defining habitat types should be made more flexible and partly adaptable to a region. Secondly, the protection aims conferred on NATURA 2000 areas are very static. Especially for areas with fast developing processes (as in succession on the heathland sites in Brandenburg), it must be possible to adapt the measures prescribed in the guidelines. Climate change acting as a new driver may lead to the fact that some of the targets of conservation becoming invalid (Neubert et al. 2008a). Therefore, adapting existing management and conservation strategies in protected sites as a pro-active response to likely anthropogenic influences may be an inevitable consequence.

**Acknowledgements:** The development of the knowledge-based approach was part of the research project SARA'04. It was carried out under the responsibility of LUP GmbH with scientific support of the Berlin University of Technology, Prof. Dr. H. Kenneweg. The authors would like to thank all the other partners of SARA'04 for their contribution: Regional Environmental Agency Brandenburg, Natural Park Nuthe-Nieplitz, Ministry for Environmental Protection and Regional Planning Brandenburg. The project was supported by the European Fund for Regional Development and the Brandenburg Ministry of Economics. The authors would like to thank the Bavarian State Institute of Forestry for their assistance in developing the object-based classification approach and its rule-base.

# Chapter V:
# Significance Analysis and Validation of Utilised Additional Geo-Information

"Significance Analysis of different types of ancillary geo-data utilised in a multisource classification process"

*International Journal of Remote Sensing (submitted)*

Michael Förster & Birgit Kleinschmit

*received September 2008*

## Abstract

Ancillary geo-data supply valuable sources of information to enhance classification accuracy for a variety of remote sensing applications. As a contribution to a substantial understanding of the integration of different data into a knowledge-based multisource classification process the presented study evaluates the significance of geo-data for the classification accuracy of a very high spatial resolution satellite image for the identification of forest types in Germany. The approach utilises a Fuzzy Logic classifier for the integration of a knowledge-base, which combines spectral information with ancillary data-layers.

The intention of the study was to find a method to evaluate integration patterns of single geo-data, the effects on different classes, and the impacts of the applied rules. Two approaches were tested: an ISODATA Clustering and a Significance Analysis of Microarrays (SAM). A sequence of classifications with possible combinations of geo-data and rules for the identified classes was derived. The resulting microarray of accuracy percentages of single classes and the overall classification was used for further investigations.

The results from the ISODATA Clustering and the SAM showed a high significance when all additional geo-data and rules were included in the classification process. While the ISODATA clustering showed valuable information about patterns in the accuracy data, the SAM can identify significant influences of different types of additional data to a variety of class accuracies.


## 1 Introduction

Reliable and up-to-date information on different aspects of the land's surface is required for a variety of environmental and planning applications (Foody 2004). Remote sensing data and classification techniques have long been detected as a valuable source of information (Campbell 2002). Recently, a variety of new sensors have been introduced, providing data of a better quality in terms of spatial (e.g. QuickBird) or spectral (e.g. HyMap) resolution as well as high resolution Radar (e.g. TerrasSar-X) or different airborne Laserscan systems (Jacobsen 2005). At the same time, the amount of available geo-data is steadily increasing. Such information can include Digital Terrain Models (DTM) and derived products (such as slope and aspect) as well as terrestrial mapped thematic information (e.g. biotope and soil maps), or weather and climate data. These spatial data contain information on specific aspects relevant to a particular study or problem and have diverse characteristics as different dates of acquisition or different spatial resolution. Hence, they provide additional and complementary information to remote sensing data and can be incorporated into a

classification procedure in various ways. To combine additional geo-data with satellite data, the spectral feature space can be extended by different data layers (e.g. derived remote sensing data layers / ratios, thematic-GIS-data) which can potentially increase the quality of the classification. Therefore, much emphasis has been drawn on the possibility to select or weigh the dimensions of the feature space in order to develop advanced classifiers to achieve highest classification accuracies for the defined classes with lowest computational resources (Duda et al. 2001).

New methods for spatial data processing are developed to handle the greater diversity of the supplied multisource data, and to address the requirements of complex information, especially in environmental research topics (Jensen 2005). Recent review papers examining state of the art classification techniques (Lu and Weng 2007), statistical pattern recognition in remote sensing (Chen and Ho in press), multiple classifier systems (Benediktsson et al. 2007), and object extraction techniques (Baltsavias 2004) highlight the importance of multisource classification from different perspectives. A variety of powerful and robust algorithms were developed, aiming to integrate different data sources to improve the quality of the classification results. It goes far beyond the scope of this paper to evaluate all these techniques. However, as the most common methods non-parametric classifiers mostly combined with knowledge-based systems are identified for the following reasons. For a multisource classification, the conventional parametric statistical classifiers, such as the Maximum Likelihood Classifier, are not appropriate, since a convenient multivariate statistical model does not exist for the data (Benediktsson et al. 2007). With non-parametric classifiers, the assumption of a normal distribution (i.e. Gaussian) of the dataset is not required. Therefore, these classifiers are especially suitable for the incorporation of additional geo-data (Lu and Weng 2007). The most commonly used techniques are Artificial Neural Networks (Kanellopoulos et al. 1997), Expert Systems (Kartikeyan et al. 1995), and Fuzzy Logic approaches (Stolz and Mauser 1996). More recently, multisource methods based on advanced tree-type classifiers, such as Random Forests (Gislason et al. 2004) and Support Vector Machines (Foody and Mathur 2004; Watanachaturaporn et al. 2005) are applied. Because all of these methods have different strengths and limitations (Tso and Mather 2001) it is possible to combine different techniques to Multiple Classifier Systems (Briem et al. 2002). Contrary to traditional methods in remote sensing which use single classifiers to determine which class a given pattern belongs to, Multiple Classifier Systems improve the classification accuracy by using an ensemble of classifiers.

Although non-parametric methods in different combinations can provide an improvement of the classification accuracy, the inclusion of *a priori* knowledge about classes and their relation to data sources is merely considered indirectly by improving the algorithms while

including more data. Therefore non-parametric classifiers without combining geo-data and remote sensing information in a rule-base often tend to compare the power and computational performance of an algorithm. Knowledge-based methods build rules with *a priori* knowledge, to relate the information of the ancillary data to the spatial distribution of certain classes, which are aimed to obtain (Maselli et al. 1995). For example road maps or census data can be often related to certain urban features (Zhang et al. 2002), while a DTM can help to distinguish different vegetation classes (Frick et al. 2005). Knowledge-based methods can be distinguished in (1) a separate analysis of geo-data and remote sensing classification with a post-processing intersection of the results, (2) interactive systems, which combine methodological steps of geoinformatics and digital image analysis, and (3) integrated classification algorithms (Hinton 1999). Especially the latter technique often utilises a known non-parametric method extended with a knowledge-base. Different approaches have been used for the integration of rules into the multisource classification process. One possibility is Evidential Reasoning technique, where each data source is considered as providing a body of evidence with a certain degree of belief. Rules integrate the conditional knowledge about states of nature based on each data source into combined knowledge based on the total evidence (Kim and Swain 1995). In Neural Network approaches the integration can be obtained by using multilayer perceptron Neural Networks for a non-parametric estimation of posterior class probabilities (Bruzzone et al. 1999). Fuzzy Logic approaches can use input and output membership functions of the knowledge (which is available in form of linguistic terms) and include the outcome possibilities as *a priori* instance (Stolz and Mauser 1996).

To derive a quantitative statement about the additional benefit of including geo-data, knowledge-based methods are often compared to standard classification algorithms, such as the Maximum Likelihood Classifier. For these purposes, the accuracy of thematic maps derived by image classification analyses is often utilised (Fitzgerald and Lees 1994; Foody 2004). Amongst others, the Kappa coefficient is the most frequently applied algorithm for remote sensing classification evaluation (Wilkinson 2003).

No matter which of the multisource classification methods is applied, the influence and significance of different geo-data sources and rule-sets on the classification accuracy is still a topic, which is required to be examined. To determine the optimal data input parameters with respect to analytical accuracy different empirical optimisation methods are applied (Peddle and Ferguson 2002). Each method relies on empirical testing with respect to the overall classification accuracy, but supplies no measure of the significance of a single input variable or rule of a knowledge-base. Similarly, cluster analysis of the accuracy assessments can find coherent patterns of valuable geo-data or rule-sets, but provides little information about

statistical significance. An approach of including a measure of significance is the usage of pairwise $t$ tests, which proved to be a good significance measure between different multisource classification settings (Sader et al. 1995). Methods based on the $t$ test provide the probability that the difference between expressions occurs by chance. When the number of possible variables and knowledge-base rules in different combinations is very large in comparison to the number of samples (trainings sets) a few values of statistical significance (e.g. $P$ = 0.01) would be identified by chance. Moreover, there is no possibility of ranking the significance of the results (e.g. for included data sources or rules). Another method for the measurement of significance of single variables of a multisource classification process is a variable importance estimate implemented in the Random Forest method (Breiman 2001). The importance of a variable (e.g. one spectral band) can be estimated by randomly permuting all values of this parameter in the out-of-bag samples for each classifier. An increased out-of-bag error is an indication of the importance of the variable (Gislason et al. 2004). However, as Random Forest multisource classification is not using a rule-base to combine the geo-data within the classification process as for knowledge-based methods, the evaluation of the significance of single rule-sets associated to a data-source would be not possible.

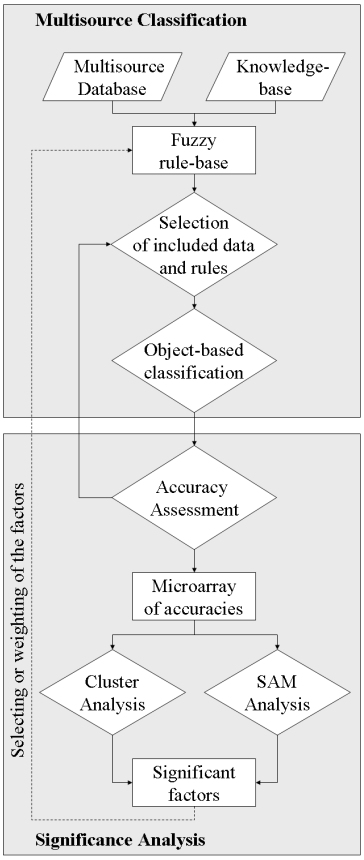



Figure 1: Evaluation scheme of the integration of ancillary data

In this study, a method is introduced, which compares different accuracies for a microarray set of classification results, which include different layers of geo-data as well as knowledge-base rules (see Figure 1 for a general scheme of the processing). The accuracy assessment results are compared with the Significance Analysis of Microarrays (SAM) and an ISODATA clustering. The multisource classification method which was assessed with SAM and ISODATA cluster analysis is a Fuzzy Logic geo-data integration approach. Because this paper is focussed on the significance analysis, the knowledge-based classification is shortly described in the following chapter, explaining preceding steps of multisource classification. For further details on the data and the classification process (Chapters 2.1 and 2.2) see (Förster and Kleinschmit 2006b). More information about the test site and the vegetation types can be found in (Förster et al. in press).

# 2 Preceding multisource classification

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 Definiens Professional (Benz et al. 2004). The segmentation levels of different resolutions were delineated and assigned to hierarchical organised groups of objects, such as forest habitats, crown combinations, and crown types of single tree species and combined with a Fuzzy knowledge-base.

## 2.1 Data and study area

For the presented investigation very high spatial resolution QuickBird data were used. In summer 2005 a scene was acquired from the forested site called “Angelberger Forst”, which is situated in the pre-alpine area of Southern Bavaria, Germany. Within this area, different semi-natural mixed forest types occur, including Beech forest, Alluvial forest-types, such as Black Alder, and additionally a smaller amount of Larch and Afforestation. The scene was acquired on 11.08.2005 and had a cloud-coverage of 10 % and an off-nadir angle of 11.3 degree. The methodological analysis was performed on a subset of 2442 per 1394 pixels (126.7 ha) of the original QuickBird data set. The subset contained no clouds, but a variety of the classes which can be found in the full data-set. The QuickBird image was geometrically corrected and pan-sharpened (Zhang 2002). The scenes were segmented at three landscape scales, but only the single tree / small tree group level (average object size: 10.5 m²) was used for the following methodological approach of investigating the influence of the geo-data, classes, and single rules.

To be integrated in the multisource classification process a Digital Terrain Model 1 : 25.000 (DTM 25), a Conceptual Soil Map 1 : 25.000 (CSM), as well as a Forest Site Map 1 : 10.000 (FSM) were used as geo-data (see Table 1). Especially the Forest Site Map consists of several attributes, which were considered separately. For the DTM, several standard parameters, such as slope, aspect, and curvature (Dirnböck et al. 2003) were derived.

## 2.2 Knowledge-based classification

The classification method is based on the combination of a tree-based (hierarchical) and a k-NN approach. Therefore, 162 training samples of the eight classes (of which only five were combined with rules, shown in Table 1) based on colour infrared aerial Photographs, Forest Organisation Maps, and field surveys were collected based on a random stratified sampling (Förster and Kleinschmit 2006b).

The complementary fuzzy knowledge-base was built-up in cooperation with the Bavarian State Institute of Forestry (Förster and Kleinschmit 2006b; Kleinschmit et al. 2006). A register of 21 rules associated to the selected specific climatic region of the area was developed (see Table 1) based on knowledge about Bavarian woodland types and the types of geo-data introduced above (Walentowski et al. 2004).

Table 1: Fuzzy Membership functions (rule-sets) and included additional data per class (the classes Spruce, Spruce-old, and Sycamore were not included here, because these classes were detected solely by spectral values of QuickBird data).

| **Class** | **Digital Terrain Model** | | | **Conceptual Soil Map** | **Forest Site Map (scaled: 1 = poor to 10 = rich)** | | |
|---|---|---|---|---|---|---|---|
| | Slope [SLO] in Degree | Aspect [AS] scaled 0-3.6 | Curvature [CUR] scaled 0-255 | Soil Type [CSM] | Available Water [FSM1] | Available Nutrients [FSM2] | Soil Substrate [FSM3] |
| Afforestation [AF] | | | | AF-CSM | | | |
| Beech [BE] | BE-SLO | BE-AS | BE-CUR | BE-CSM | BE-FSM1 | BE-FSM2 | BE-FSM3 |
| Beech – young [BEY] | BEY-SLO | BEY-AS | BEY-CUR | BEY-CSM | BEY-FSM1 | BEY-FSM2 | BEY-FSM3 |
| Black Alder [BA] | BA-SLO | | BA-CUR | | BA-FSM1 | BA-FSM2 | BA-FSM3 |
| Larch [LA] | LA-SLO | | | | | | |

For the implementation of additional knowledge via Fuzzy Logic input membership functions (IMF) and output membership functions (OMF) were defined for each rule-set (Openshaw and Openshaw 1997). To give an example, for the case of the class “Black

Alder" and the geo-data "Curvature" (BA-CUR, see Table 1) five IMF were distinguished based on expert knowledge on a scale of 0 (extremely concave curvature – depression) to 255 (extremely convex curvature – hilltop): very concave (0 – 20), concave (>20 – 100), plane (>100 – 150), convex (>150 – 235), and very convex (>235). The OMF are defined by the potential possibility of occurrence of Black Alder: very good, good, medium, poor, very poor. These membership functions were combined with the help of a rule-base. For the example BA-CUR, the rules are defined as:

| | | |
|---|---|---|
| IF Curvature = 0 – 20 | THEN | Black Alder = very good |
| IF Curvature = >20 – 100 | THEN | Black Alder = very good |
| IF Curvature = >100 – 150 | THEN | Black Alder = good |
| IF Curvature = >150 – 235 | THEN | Black Alder = poor |
| IF Curvature = >235 | THEN | Black Alder = very poor |

For each object, the rule-base is defuzzicated by the weighted means method (Tilli 1993). This can be graphically visualised with the help of response curves. Therefore, the defuzzicated value is calculated in constant intervals and assigned to a graph. Response curves are implemented in the fuzzy-classification base of Definiens Professional for all possible rules as shown for the example of BA-CUR (Figure 2).

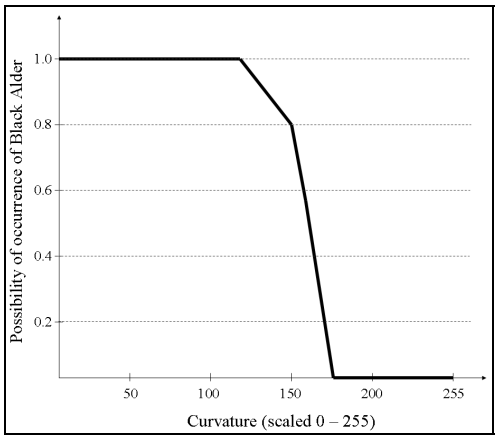



Figure 2: Example of the graphical visualisation of a response curve for BA-CUR

For combining the fuzzy-sets and the remote sensing classification results the approach uses the minimum (AND-) rule, which specifies that the most unacceptable factor is the critical value for the class to occur. Each classified object consists of at least three different classification output probabilities. In a classification, based on training samples, the class with the highest probability will be assigned to the object. The knowledge-based fuzzy-

classification includes the fuzzy-based possibility of the defined rules into this classification process. To combine the classification values with the fuzzy possibilities the spectral classes were normalised (Stolz 1998). Therefore, the spectral probability of a class $\omega_i$ for an object $x$ is divided by the maximum probability $P(x \mid \omega_i)$ of all classes by

$$\Pi_{sp}(\omega_i \mid x) = \frac{P(x \mid \omega_i)}{\max[P(x \mid \omega_i)]} \tag{1}$$

where $\Pi_{sp}(\omega_i \mid x)$ is the resulting possibility of a class. In the example given for the rule BA-CUR a given object could be classified as Beech based on the spectral features, but a lower probability is assigned to the class Black Alder. Since the object contains a very low curvature value, the possibility of occurrence of Black Alder is still high, while the possibility for the class Beech is lowering. When applying the minimum (AND-) rule the possibility of occurrence of the class Black Alder is higher than for Beech. In the following step the minimum possibility of each possible class will be compared. The class with the highest membership will be assigned to the object (Förster and Kleinschmit 2008).

# 3 Data and methods

The main objective of this study was to find a suitable method for measuring the significance of additional information. Therefore, two methods were utilised for detecting significances within a knowledge-based classification. The preceding classification was varied under the inclusion of different parts of the rule-base in the classification process. The interest in significance of additional data can be subdivided in the following research questions:

- Is there a significance or insignificance of certain types of geo-data detectable?
- Is the inclusion of additional data of significance for single classes?
- Can single rules significantly alter the classification accuracy? Does a combination of rules lead to a higher significance than single rules?

To evaluate the significance of different classification processes, variations of included geo-data and rules are subsequently processed and the classification accuracies combined in a microarray (see Chapter 3.1). The various classification accuracies for single classes and the overall accuracy were evaluated in order to find the most significant factors or the combinations of factors which were included. Therefore the derived accuracies were

evaluated by ISODATA clustering (see Chapter 3.2) and by a Significance Analysis of Microarrays (see Chapter 3.3).

## 3.1 Microarray of accuracies as data-basis

To investigate the significance of geo-data, accuracy of classes, and single rules a selection of 50 variations of classifications, including different sets of rules was chosen for classification and a subsequent accuracy assessment (see Table 2). Firstly, all rules for single classes were included according to Table 1. These, so called single rules are all tested separately, so they can be compared with combined rules, which include a combination of at least two rules.

Table 2: Classification setting for significance analysis of the additional geo-data

| **Classification type** | **Abbreviations** |
|---|---|
| All geo-data included | ALL |
| All geo-data excluded (classification with all or single spectral bands) | |
| all spectral Bands included | ONLY-SB |
| only Band 1 included (blue) | ONLY-B1 |
| only Band 2 included (green) | ONLY-B2 |
| only Band 3 included (red) | ONLY-B3 |
| only Band 4 included (near infrared) | ONLY-B4 |
| Single types of geo-data included in all classes and rules | |
| all rules with DTM-data | ALL-DTM |
| all rules with aspect | ALL-AS |
| all rules with curvature | ALL-CUR |
| all rules with slope | ALL-SLO |
| all rules with the CSM | ALL-CSM |
| all rules with the FSM | ALL-FSM |
| Single types of geo-data excluded from all classes and rules | |
| no rules with DTM-data | NO-DTM |
| no rules with aspect | NO-AS |
| no rules with curvature | NO-CUR |
| no rules with slope | NO-SLO |
| no rules with the CSM | NO-CSM |
| no rules with the FSM | NO-FSM |
| All rules included for one class | |
| all rules for Beech | ALL-BE |
| all rules for Beech-young | ALL-BEY |
| all rules for Black Alder | ALL-BA |
| all rules for Afforestation | ALL-AF |
| all rules for Larch | ALL-LA |
| All FSM-data (soil information) or all DTM-data are included for a class | |
| all FSM-rules for Beech | BE-FSM |
| all DTM-rules for Beech | BE-DTM |
| all FSM-rules for Beech-young | BEY-FSM |
| all DTM-rules for Beech-young | BEY-DTM |
| all FSM-rules for Black Alder | BA-FSM |
| all DTM-rules for Black Alder | BA-DTM |
| Only a single rule is included | 21 different rules – see Table 1 for explanation and abbreviations |

To further evaluate the significance of sets of geo-data all rules including an additional data-set (e.g. slope) were applied combined. So it is possible to detect changes occurring when the data were explicitly used. Vice versa, to evaluate significance of the absence of certain geo-data, single types of geo-data were excluded from all rules. Moreover, to evaluate specifically the significance of soil and terrain information, combined rule-sets are tested, when all rules include either the FSM or the DTM.

To estimate the significance of different classes, accuracy assessments were derived, including all rules for a single class. Because different forest types are differently influenced by soil-data or terrain-data, for the three classes with high rule amount of DTM and FSM, the accuracy was assessed separately.

The results of the classification processes including different sets of ancillary information were compared to test samples taken from Forest Organisation Maps, field work, and aerial photographs. The outcomes of the accuracy assessments are accuracy matrices as well as information about user and producer accuracy. Although these results certainly give a better insight into a single classification, for purposes of comparison only the overall Kappa coefficient for all classes and variation was taken into account. As a result a microarray of Kappa coefficient accuracies are produced as an intermediate result (see Table 3 for an example).

Table 3: Example of a microarray for integrated Fuzzy Membership rules for the tree type Beech (Abbreviations are explained in Table 1 and Table 2). Some classes were not affected by this specific rule-set, as it is the case for Afforestation (AF), which achieves a stable Kappa coefficient of 1.00.

| | **BE** | **BEY** | **BA** | **AF** | **LA** | **Overall Kappa** |
|---|---|---|---|---|---|---|
| **BE-ALL** | 0.09 | 0.21 | 0.17 | 1.00 | 0.63 | 0.52 |
| **BE-DTM** | 0.09 | 0.21 | 0.17 | 1.00 | 0.63 | 0.52 |
| **BE-AS** | 0.52 | 0.07 | 0.10 | 1.00 | 0.41 | 0.65 |
| **BE-CUR** | 0.09 | 0.21 | 0.17 | 1.00 | 0.53 | 0.52 |
| **BE-SLO** | 0.46 | 0.06 | 0.09 | 1.00 | 0.63 | 0.64 |
| **BE-CSM** | 0.52 | 0.07 | 0.10 | 1.00 | 0.41 | 0.65 |
| **BE-FSM** | 0.52 | 0.07 | 0.10 | 1.00 | 0.41 | 0.65 |
| **BE-FSM1** | 0.52 | 0.07 | 0.10 | 1.00 | 0.41 | 0.65 |
| **BE-FSM2** | 0.52 | 0.07 | 0.10 | 1.00 | 0.41 | 0.65 |
| **BE-FSM3** | 0.52 | 0.07 | 0.10 | 1.00 | 0.41 | 0.65 |

It would be a very simple procedure just to compare the overall accuracies for the classifications or single classes by ordering the results from highest to lowest accuracy. Although the ordering of the results gives a good overview, the significance of the single accuracy of classes would be neglected. There is still the possibility that the changes occur by chance, when evaluating a microarray. Moreover, the interactions in the classification results between different included rules are not sufficiently examined. The combination of the different rules or geo-data could interfere with each other, obscuring certain accuracy results or indicating only random fluctuations (e.g. single classes are extremely good evaluated while the overall accuracy is comparably low). The most important shortcoming of a simple comparison is the absence of a measure or value for the significance. Therefore, a significance analysis was further developed in Chapters 3.2 and 3.3.

## 3.2 ISODATA cluster analysis

The cluster analysis is suitable for detecting patterns in multidimensional arrays for different applications (Arabie et al. 1998). In remote sensing context it is used as an efficient method of partitioning multispectral image feature spaces and extracting land-cover information (Huang 2002). One of the most efficient and approved methods is the ISODATA clustering, which is still utilised in a variety of applications, mostly in combination with more complex classification algorithms (Stow et al. 2003). Besides others, the algorithm can be affected by the parameters maximum number of clusters, maximum percentage of values which remain unchanged between iterations, and minimum members in a cluster. Unlike most remote sensing approaches, the microarray of 50 accuracies for each of the eight defined classes was classified in order to find clusters of significance.

In order to find the suitable amount of clusters, the Minimum Euclidean Distance ($\varepsilon$) between the clusters of the microarray was evaluated. When the number of clusters ($n$) is increasing, then $\varepsilon$ is subsequently reduced (see right axis in Figure 3) due to the higher cluster density in the feature space. $\varepsilon$ multiplied by the amount of clusters results in a proportion distance measure, normalising the reduction of the standard Euclidian Distance. The output maximum of $n*\varepsilon$ (see left axis in Figure 3) is therefore defined as the optimum cluster amount. Hence, five clusters were chosen for further evaluation of ISODATA clustering.

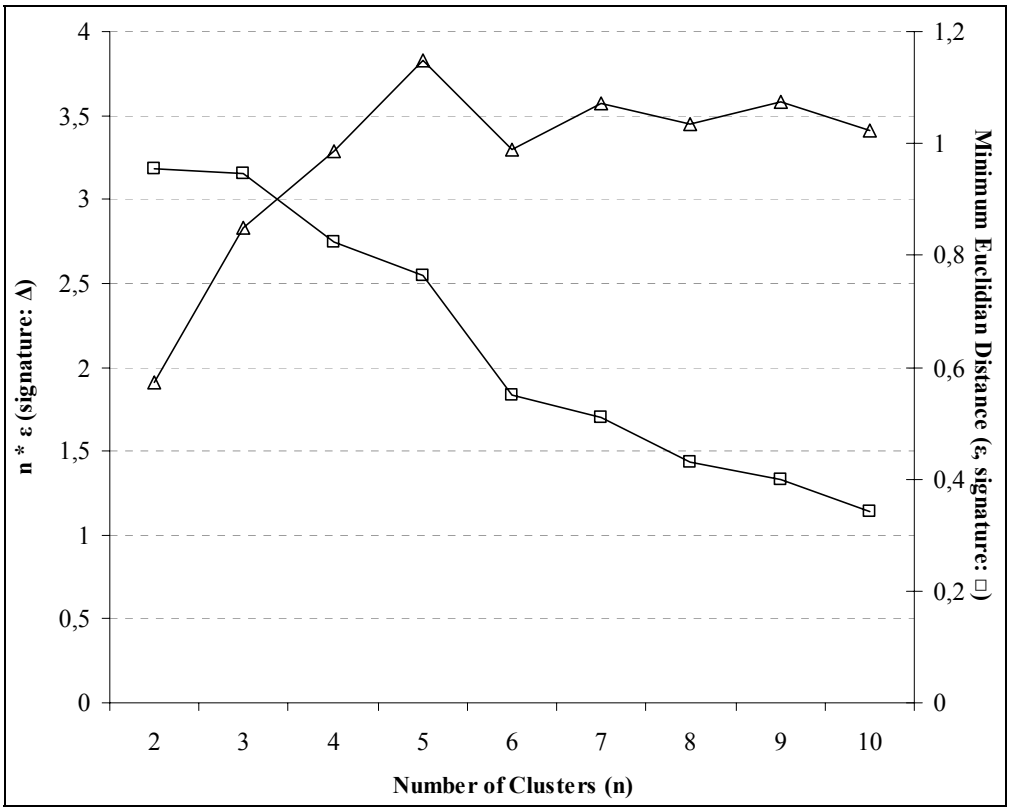



Figure 3: Evaluation of the Euclidian Distance of the microarray for finding the optimal number of clusters.

## 3.3 Significance Analysis with Microarrays (SAM)

Microarrays are frequently used in Bioinformatics and in medical research mainly for pattern research of large gene experiments (Li et al. 2007). Hence, methods were developed to determine the significance of changes within these arrays. One of these techniques is the Significance Analysis of Microarrays. SAM assigns a score to each data set on the basis of change in the results relative to the standard deviation of the measurements. Values with scores greater than a defined threshold are evaluated to be significant (Tusher et al. 2001). The percentages of variables which have no significant influence in the classification accuracy (or even detected by chance) are named False Discovery Rate (FDR). The FDR is estimated by analysing permutations of the measurements (Storey 2002).

Instead of test results on genes, the different variations of the accuracy assessment for test data were used. To account fluctuations which are specific to a data set, a statistic is defined based on the ratio of change in data expression to standard deviation in the data for a certain variable. The ‘‘relative difference’’ $d(i)$ is:

$$d(i) = \frac{r(i)}{s(i) + s_0} \tag{2}$$

where $r(i)$ is called response type, which is defined as the average level of expression for accuracy assessment rule-set ($i$). The term s($i$) is the standard deviation of repeated accuracy measurements under different inclusions of rule-sets and $s_0$ is an exchangeability factor. Since the achieved accuracy assessments are not a part of a repeated measurement (which would result in paired or even multiclass responses), the response type is defined by:

$$r(i) = \sum_j \frac{x_{ij}}{n} \tag{3}$$

which is the average of all rule-sets $(i)$ in all measured classes $j = 1,2,...n$ and

$$s(i) = \sqrt{\sum_j \frac{(x_{ij} - \bar{x}_i)^2}{n(n-1)}} \tag{4}$$

If the measurements have low expression levels (are very similar), the variance of $d(i)$ tends to be very high, because of the small values of $s(i)$. To keep the values of $d(i)$ comparable to other test series, the values should be independent from the expression level. Therefore a small additive factor $s_0$ is introduced, which is computed as a function of $s(i)$ in moving windows across the data (Chu et al. 2008).

The relative difference $d(i)$ is ordered from highest to lowest. Although this result provides a control for random fluctuations, an additional control is required to assign statistical significance to the accuracy assessments. This would be possible by taking more test-samples and performing additional and independent accuracy assessments of the classifications, which would be inefficient and expensive. Therefore, a set of 100 repeated permutations of the data are used to determine if the expression of the data sets are significantly related to the response. To implement the permutations the expression values for each experiment are multiplied by +1 or −1, with equal probability. From the permutations relative differences $d_p(i)$ are also calculated and ranked.

The expected relative difference $d_E(i)$ is defined as the average over the permutations:

$$d_E(i) = \sum_p \frac{d_p(i)}{100} \tag{5}$$

To identify potentially significant changes in expression, a scatter plot of the observed relative difference $d(i)$ vs. the expected relative difference $d_E(i)$ is used (see Figure 5). The classification accuracy data which are represented by points displaced from the $d(i) \cong d_E(i)$ line by a distance greater than a defined threshold $\Delta$ are called significant. This accounts for all values which have a higher $d(i)$ for positively significant values and a lower $d(i)$ for negatively significant values, even if these ranked calculations are below the defined threshold.

To determine the number of falsely significant values generated by SAM, horizontal cutoffs were defined for the smallest $d(i)$ of a value estimated as positively significant and the highest $d(i)$ of a value estimated as negatively significant. The number of falsely significant values was computed by counting the numbers of values that exceed the horizontal cutoffs for each permutation. The average number for all permutation is divided by the number of "genes" called significant. The outcome is called False Discovery Rate (FDR). As $\Delta$ is decreasing, the number of significant values is increasing, but at the same time the probability of a higher FDR is rising. Therefore, FDR gives a control for overestimating $\Delta$. An alternative for this procedure would be a standard $t$ test, which would introduce a symmetric horizontal cutoff. Because there is the possibility that positively and negatively significant values are not evenly distributed, the FDR was applied in this study.

# 4 Results

The microarray of Fuzzy-Logic based classifications was evaluated. The results were assessed with the software R (Venables and Smith 2006) and SAM (Chu et al. 2008). The results of the ISODATA Clustering and the SAM method are described in Chapters 4.1 and 4.2, while the Discussion section gives a comparative overview of the significance of geo-data, different classes, and single rules.

## 4.1 ISODATA clustering of the microarray

The ISODATA cluster analysis of the accuracy results was used to find coherent patterns in the microarray. An exemplary scatterplot of BE Kappa Accuracies and Overall Kappa

Accuracies is shown in Figure 4 for the optimum of five clusters. Parts of the values of this plot are shown in the first and the last row of Table 3.

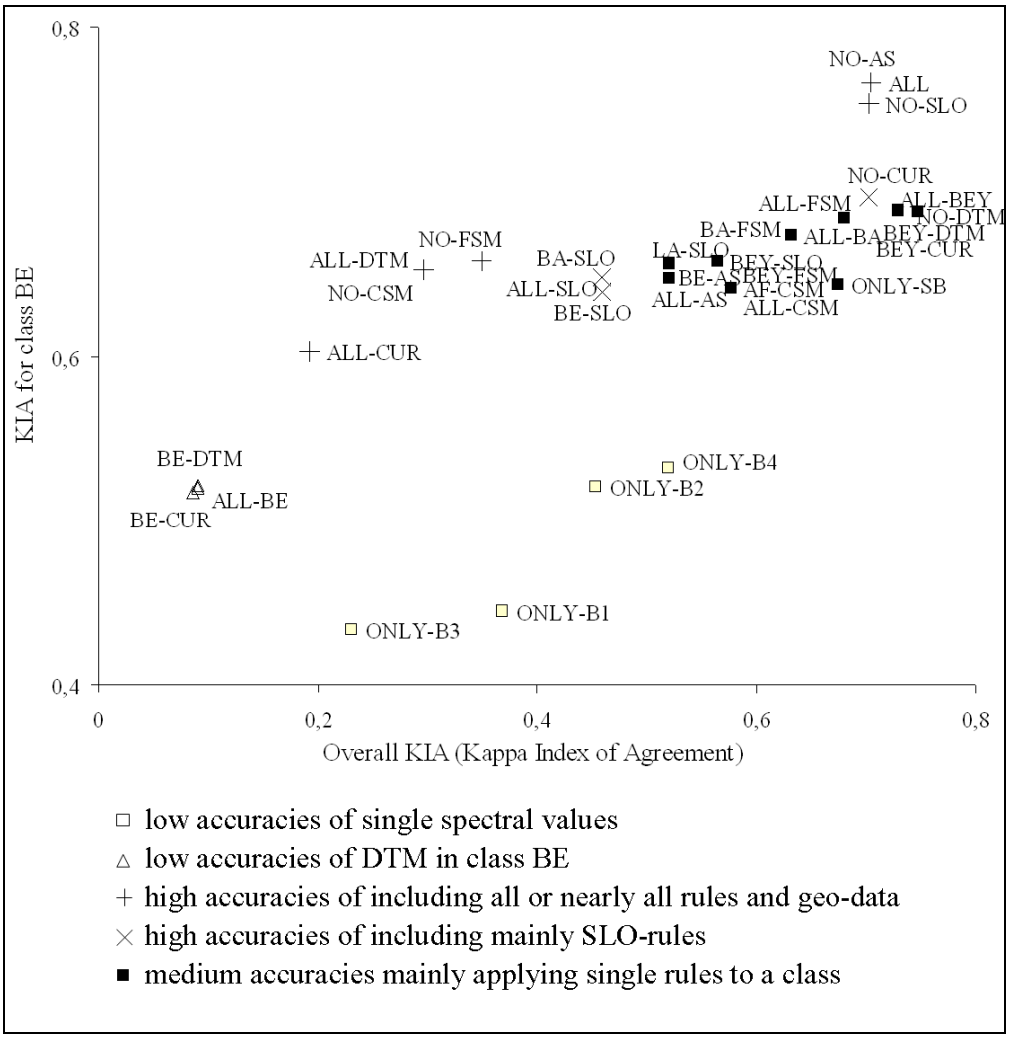



Figure 4: Scatterplot of the ISODATA clustering results for KIA-accuracy of BE with the Overall KIA. The clustering was performed with five classes. For Abbreviations see Table 1 and Table 2.

The first detected pattern is a cluster of low accuracies, when applying only a spectral classification and not using all spectral bands (marked with □). Even the classification calculated with all spectral bands is only assigned to the medium accuracy cluster. Another cluster (marked with +) identified classes with high accuracies, when all or nearly all geo-data were included (e.g. NO-AS or NO-SLO mean that just Slope or Aspect were not integrated in the classification process). A very similar cluster (marked with ×) shows high accuracies, when mainly Slope was included in the classification process. Hence, it can be stated that including the whole set of rules and geo-data improves the classification accuracy. Moreover, the factor Slope seems to be very distinct in classification accuracy response. Another discrete cluster demonstrates very low accuracies when the DTM is included in the

classification of BE (marked with Δ). This feature indicates that the rules might no be well evaluated, which means that the terrain is not influencing the occurrence of BE at all. Therefore, these rules might be excludable from the rule-base. Most of the classification accuracies, however, are within a cluster called medium accuracies which mainly apply a single rule to a class (marked with ■).

Although the ISODATA cluster analysis can detect interpretable patterns of the influence of additional data, it could not give a detailed differentiation of the influence of single classes or types of geo-data (which were accumulated in a medium class). Moreover, this technique provides no information about the statistical significance of single classification approaches.

## 4.2 Significance Analysis of Microarrays (SAM)

In contrast to the ISODATA clustering approach, the SAM technique detects explicitly significant classifications. As explained in Chapter 3.3, the method was applied to the microarray of accuracies described in Chapter 3.1. The relative difference $d(i)$ and the expected relative difference $d_E(i)$ were calculated. To derive the significant classification settings, the threshold of significance ($\Delta$) was applied to different values (0.4. to 1.0 – see Table 4). Thereby, the number of significant values is increasing, while the FDR is reaching simultaneously higher percentages. Taking the steadily rising FDR into account, a threshold $\Delta$ of 0.8 is chosen for further investigation, because the FDR is relatively low in comparison to the overall amount of detected potentially significant variables from the microarray (5.7 %).

Table 4: Different thresholds and FDRs for detecting the optimum set of possibly significant values

| **Threshold of significance ($\Delta$)** | **Number significant (□)** | **FDR (in %)** |
|---|---|---|
| 0.4 / 0.5 / 0.6 | 28 | 28.6 |
| 0.7 | 22 | 12.7 |
| 0.8 | 14 | 5.7 |
| 0.9 / 1.0 | 3 | 0.0 |

Figure 5 visualises the observed score $d(i)$ and the expected score from the permutations of the initial data $d_E(i)$ in a plot. 14 of the classifications were evaluated as significant. While three accuracies had a negative significance (squares on the left side of the plot), 11 of the input data sets were valued as positively significant (squares on the right side of the plot).

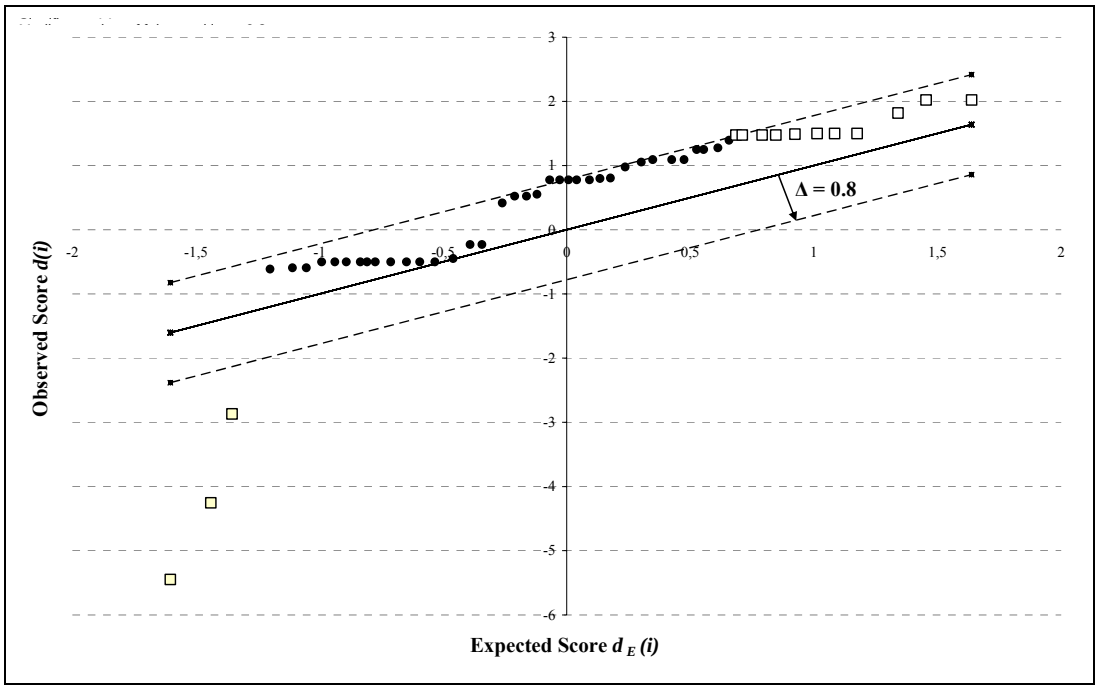



Figure 5: Plotsheet of the SAM method. Significant values for Δ = 0.8 are indicated by the squares (□).

The results of this calculation are listed in a table of significant values, which is summarised in Table 5 beginning with the top entries of the highest relative difference $d(i)$. As the ISODATA clustering was already indicating, SAM shows a high positive significance of the combined utilization of all additional geo-data and rules. The first three ranked rule-sets are containing nearly all rules. It has to be recognised, that excluding terrain-related rules (such as slope and aspect) does not seem to affect high values of relative difference. On the contrary, the factor slope seems to have significant influence when applying to single classes, such as BEY. This might be due to the fact, that the included rules have a particularly high impact on the classes BEY and BA in different rule-combinations. It has to be highlighted that the rule combinations including soil information (especially of the Forest Site Map – FSM) tend to have a high positive significance. Within the FSM data-base the factor of available nutrients to the forest types is calculated most often with a high $d(i)$. Apart from the first three values of relative difference, the values with positive significance have very similar evaluations (1.44 to 1.50).

Concerning the values with a negative significance, classifications without the full spectral range of the sensor are detected. The fact that classification with the optical range (blue = band 1; green = band 2; blue = band 3) have the lowest $d(i)$ reflects the classification experience, that the spectral information in the near infrared range (band 4) contains the most significant information to vegetation classification.

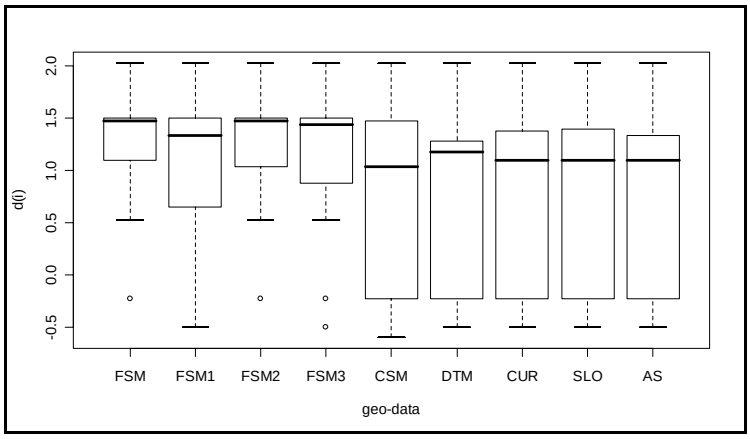
Table 5: List of positively and negatively significant variables of accuracy assessments

| **Positively significant values** | ***d(i)*** | **Negatively significant values** | ***d(i)*** |
|---|---|---|---|
| ALL | 2.02 | ONLY-B3 | -5.44 |
| NO-AS | 2.02 | ONLY-B2 | -4.25 |
| NO-SLO | 1.82 | ONLY-B1 | -2.87 |
| BEY-SLO | 1.50 | | |
| BEY-FSM | 1.49 | | |
| BEY-FSM2 | 1.49 | | |
| ALL-FSM | 1.48 | | |
| ALL-BA | 1.48 | | |
| BA-FSM | 1.48 | | |
| BA-FSM2 | 1.47 | | |
| BA-FSM3 | 1.44 | | |

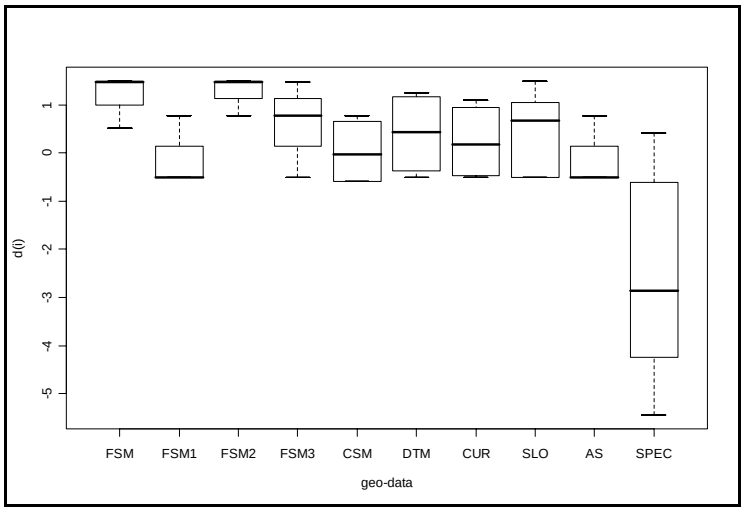
To investigate the relevance of the single types of geo-data all values of the relative difference $d(i)$ were assigned to the geo-data variables, which were included in the classification. In the first case (Figure 6a) all $d(i)$ values of geo-data were selected and averaged, even if the named geo-data are only included as a subset in the classification process (e.g. the classification result of "All geo-data are included" (ALL) is assigned to all variables). Soil data in form of the FSM show highest relative differences, although only to a limited degree for FSM1 (variable: "available water"). All other factors indicate a lower significance. In comparison, the variables CSM and Aspect are of reduced relevance for the classification accuracy. Furthermore, variables which subsume various factors, such as DTM (combined SLO, AS, and CUR) or FSM, showing higher average $d(i)$ than single parts of the geo-database. The results of such combined factors support the assumption that a combination of rules is more significant than single rules.

The relative difference was assigned for the cases when variables were exclusively included in the classification process (e.g. the classification results of ALL are not assigned to the factor Curvature, while the results solely using the rule BE-CUR is chosen; see Figure 6b).

The average relative differences show more distinctly comparable results than the inclusion of subsets. While the variables FSM and FSM2 (available nutrients) indicate still high significance, FSM1 shows one of the lowest values. This factor seems, together with CSM and Aspect, to be not highly relevant for the classification success. The more distinct differentiation of significance of geo-data is extending to the DTM variables, where the factor Slope shows one of the highest average $d(i)$. Again, spectral classification is proving to be negatively significant, which can be seen as supportive measure of working with multisource datasets.




(a)




(b)

Figure 6: Influence of the type of geo-data on the relative difference when variables were included as a subset in the classification process (other geo-data is utilised simultaneously, e.g. ALL or NO-SLO; upper figure) and when exclusively included in the classification process (no other geo-data utilised, e.g. BE-FSM1; lower figure)

# 5 Discussion

The presented study utilised two methods for estimating the significance and the patterns of different classification accuracies derived from a microarray-set including different ancillary information. The ISODATA clustering was useful in order to find coherent patterns in the accuracy values. Especially two clusters give a good basis for interpretation. So have additional terrain information a negative effect on the classification accuracy of the class BE. Moreover, the ISODATA clustering and the SAM agree in giving negative indicators to classifications, where the spectral value is only partly utilised. Hence, the clustering indicates interpretable patterns especially for effects on single classes. Although this information might be helpful for a general evaluation of the influence of additional data to classification processes, the method gives no statistical reliable statement as well as no information about single rules. Moreover, there is a lack of internal differentiation of patterns, which could be reduced by involving a higher amount of clusters. Increasing cluster number leads to separation of patterns which are very similar (Arabie et al. 1998). Therefore, the Euclidian Distances are low compared to the amount of calculated clusters, which makes the interpretation of the clusters more to a task of guessing than of a reasonable conclusion.

As the main objective of this study was to find a suitable method for detecting significant patterns in the microarray of classification accuracies, it can be stated that the SAM method was more appropriate. SAM is not only showing trends or patterns but merely reliable values of significance. In the following chapters (5.1 to 5.3) the results of the SAM are discussed related to the influence of different types of geo-data and rule-sets on the defined classes.

## 5.1 Significance of types of geo-data

To estimate the significance of certain types of geo-data was one of the major interests of the study. Different factors of soil and terrain data were utilised in this special case, but the SAM method is capable of detecting significance of other additional data and application cases (Efron and Tibshirani in press). When calculating the positively significant values, the highest results are reached by an accumulation of all or nearly all additional information (see Table 5). Of the seven included variables (see Table 2) especially the rules which include the factors Slope and FSM occur above average. One reason for this result is that some variables (e.g. Aspect) are less relevant to the classification process. As an example, Aspect (southern and northern exposition) has only a very limited influence into the growing behaviour of the forest types in this region.

Another reason for low significances can be found in inappropriate data quality. The CSM soil-map contains similar information to the FSM soil map. However, the scale of the CSM

is not as fine. Moreover, the aggregation of the classes of the CSM is not as useful for relating them to vegetation information as the FSM information. Therefore, not only type of variable (e.g. soil) but quality of data is verifiably relevant for the significance of variables.

### 5.2 Class dependency on rule-base

Moreover, the significance of the classification accuracy is influenced by the type of classes. The most significant changes are to recognise for the classes BEY and BA. Of the 11 rules detected as significant are four for BA and three for BEY. If the relative difference from all rules is taken into account, $d(i)$ of the classes BA and BEY is much higher in comparison to e.g. BE. Therefore it can be assumed that the classification accuracy of classes with smaller ecological niches depend more on the application of ancillary information than other forest types. However, because some classes have applied only one or no rule at all it is difficult to draw further conclusions about the significance of types of classes. For further research a study area with a greater variety of different classes and more connected rules could give more insights into class dependency of a rule-base.

### 5.3 Significance of single rules

Of the 11 cases detected as significant four are based solely on one rule. This indicates that single rules can be as important as combined rules included into the classification process. In observing the relative differences for all classification results, the variance of the significance of single rules is very high. On average the $d(i)$ for combined rules (0.86) is higher than for single rules (0.42).

Generally, if there is a high significance of single rules then the relation between class and geo-data is very high. However, this argument even more accounts for combined rules. Nevertheless, single rules of high significance for the classification process could be used for weighting the rule-based or further investigation of the relations between rule and class-related information.

## 6 Conclusion and outlook

This article presented a series of experiments of classifying QuickBird data with different types of additional information in a multisource classification process. Different studies proved the importance of multisource data for classifications (Lu and Weng 2007; Tso and Mather 2001). The added value of included geo-data was often demonstrated by comparing the classification accuracies with and without such information (Sader et al. 1995).

Nevertheless, a detailed investigation of the significance of different types of geo-data or knowledge-based rule-sets was still a research topic to be investigated. Therefore, two methods of evaluating the significance of multiple classification accuracies were presented in this study. While the ISODATA clustering showed valuable information about patterns in the accuracy data, the SAM technique can identify significant influences of different types of additional data to different class accuracies. This contributed to a better evaluation of classification results, especially when using multiple spatial data-types.

In conclusion it can be stated that the integration of ancillary data and knowledge was valuable to the classification success for the particular example of forest types. This accounts especially for forest types with a narrow ecological niche, such as BA. Such forest types prefer more clearly defined locations, which makes it easier to formulate and implement precise rules (Schulz et al. 2002). The influence and significance of different types of geo-data was even more comprehensible for this classification example. In the SAM-order the accuracies using the Conceptual Soil Map (CSM) and the parameter Aspect of the DTM were not detected as negatively significant. However, all accuracies which included only these variables indicated significance below average. Hence, the SAM method gives the possibility to either weigh the more significant geo-data or rules or omit insignificant parameters.

The presented significance measures can be utilised in other application cases and classification methods. Further studies could involve SAM for testing the significance of hyperspectral bands (Camps-Valls et al. 2004) or for the evaluating of the significance of scales when segmenting objects. The integration in other classification methods dealing with multisource data, such as Support Vector Machines (Foody and Mathur 2004) would be an interesting and challenging task. Moreover, the process can be transferred to other regions with other data-types and qualities (Velazquez et al. 2008) to detect differences and similarities of geographic regions.

**Acknowledgement:** This research was supported and partially funded by the Bavarian State Institute of Forestry and the Bavarian Forest Management. Dr. Walentowski, Mr. Seitz and Mr. Fischer are thanked for their knowledge in forest ecology while developing a rule-base.

# Chapter VI:
# Synthesis

# 1 Conclusions

In this thesis the integration of geo-data in classification processes was investigated in the context of biodiversity monitoring using approaches of different scales. Based on the research questions posed at the introduction of the thesis, this chapter discusses the research results and gives recommendations for future research.

## 1.1 Suitable techniques for NATURA 2000 monitoring

*Research Question: Which type of remote sensing data and which method is especially suitable for the monitoring biodiversity in the framework of NATURA 2000? Which classification technique is best to be utilised for the different monitoring tasks at different scales of the Habitats Directive?*

The results of Chapter III and Chapter IV clearly showed that a support of monitoring tasks in the context of NATURA 2000 is possible with different multisource remote sensing methods. For a satellite-based measuring of biodiversity often two or multi-scale approaches including modelling as well as direct measurement of habitats are suggested (Gillespie et al. 2008; Rocchini 2007; Turner et al. 2003). Therefore, depending on pre-defined scales of the Habitats Directive two techniques were applied in this thesis to cover the requirements of the European Commission.

The utilised two-scale process shown in Table 1 is concerned with different requirements of satellite data, implementation costs as well as possible monitoring tasks. Because the biogeographic level requires remote sensing data covering large areas an approach using high spatial resolution (HSR) data instead of very high spatial resolution (VHSR) data was considered to be appropriate. This is even more the case when considering the costs for data and for classification processing. At the proposed spatial resolution of approximately 2.5 to 15 meters it is possible to classify major land-cover and/or habitat types and therefore derive information about covered range and area of habitats. The SAC-Level requires more detailed information about species and habitats. Therefore, it is assumed that the spatial resolution of the utilised remote sensing data should be higher. Although (as demonstrated especially in Chapter IV) the classification success at the SAC-level depends strongly on the type of habitats and quality parameters to monitor, it is supposed that the on-site conservation status of an area and its changes can be assessed.

Table 1: Outline of different data, costs, monitoring tasks, and scale used for the two-scale NATURA 2000 monitoring approach

| **Criteria for monitoring** | **Biogeographic Level** | **SAC - Level** |
|---|---|---|
| Possible remote sensing data | high spatial resolution (e.g., SPOT5, ASTER) | very high spatial resolution (e.g., IKONOS, QuickBird) |
| Estimated implementation costs[21] | | |
| - for data | 2.5 – 5 € per km² | 15 – 30 € per km² |
| - for method development | high | high |
| - for method application and transfer (recurring tasks) | low | medium |
| Possible monitoring tasks | Potential habitat types<br>Biogeographic indicators (range and area of habitats) | Potential habitat types<br>Quality parameters and conservation status |
| Working Scale | 1 : 25,000 | 1 : 5,000 |

Although this thesis is mainly concerned with forested NATURA 2000 areas, the two scale approach was successfully applied to other ecosystems (Bock et al. 2005; Lang and Langanke 2005). SAC-scale studies dominate the literature and show operational success (Hajek 2008; James et al. 2007; Korpela In Press; Maltamo et al. 2005; Nuske et al. 2007). However, the application of remote sensing methods on the biogeographic level (suggested by the European Commission) is still at the beginning of the research activities. As Table 1 indicates, monitoring tasks can be distinguished in

- the detection of **habitat types** at the SAC and the biogeographic level,
- the detection of **quality parameters** and the **conservation status** to be monitored at the SAC-level,
- the detection of the parameters range and area as **biogeographic indicators** (European Commission 2006).

As Chapter III shows, detection and monitoring of forested **habitat types** is possible when using HSR satellite data and modelling approaches with varying accuracy, depending on the habitat type. Based on the knowledge that a classification of forest cover (forest / non-forest / clear cut) is possible for large areas (Oehmichen 2007; Tiehl et al. 2008) and that a classification of single trees is possible with VHSR data and additional LIDAR information for small areas (Hirschmugl et al. 2007; Tiede et al. 2008), Chapter III recommends an

[21] The costs for method development and transfer are difficult to estimate. As shown in Chapter III (Table 3) it is possible to record the working hours for the implementation of a classification. However, these recorded numbers are closely related to classification quality. Moreover, the costs depend on the application case and the required quality parameters or classes. Because of these difficulties a qualitative estimation based on the experiences of the author is given for method development and application.

intermediate way. The method combines a high accuracy classification of main forest types (mixed, coniferous, and deciduous) with a modelling of potential natural vegetation.

For forested NATURA 2000 areas at the SAC-scale, national **quality parameters** for monitoring the conservation status are defined by habitat structures, availability of typical species in the habitat and disturbances of the habitat type (see Table 2). Within this thesis the parameter "Percentage of natural forest type" was exemplarily investigated in more depth (see Chapters IVa and IVb). Moreover, the parameters "Disturbance of forest structure" and "Disturbance caused by forest fragmentation" can be detected by the method proposed in this thesis. Besides the availability of VHSR satellite data the utilisation of LIDAR data seems crucial to derive some of the parameters. Especially for the detection of surface changes from erosion and different forest development stages this type of data is clearly required. Even for the detection of understory vegetation there are approaches using LIDAR data. Hence, it can be stated that (apart from the parameters "Number of biotope trees" and "Faunal quality") remote sensing approaches for the monitoring of quality parameters are suitable. This was tested in a variety of studies (see Table 2). However, more effort should be put into developing a coherent and operational method which includes most of or all parameters to assess the conservation status of a SAC.

According to the European Commission (2006) biogeographic indicators are defined by range, area, and population. Apart from a short description the terms are not explained in detail. Because a full coverage for the biogeographic region is required, VHSR satellites are regarded less suitable due to availability and cost constraints (see Table 2). The parameter "population", which can be specified for single species by indirect approaches (Jobin et al. 2008), is not detectable for large areas in sufficient accuracy with remote sensing methods. For the parameters "area" and "range" a combination of the modelling approaches of Chapter III with an analysis of landscape metrics can be suggested. Landscape metrics have proved to be helpful when analysing NATURA 2000 areas (Klug et al. 2003). For forested areas the term "range" could be defined by core area metrics, such as Core Area Distribution. The term "area" is more likely to be analysed by isolation/proximity indices (Miethke 2006). However, since patches or objects derived by remote sensing are scale-dependent when analysing landscape metrics (Förster and Kleinschmit 2006a), the results have to be carefully evaluated.

These monitoring tasks are described in further detail in Table 2. In addition to the findings of this thesis, references for alternative approaches utilising other data-types (e.g. LIDAR) and / or methods are provided, based on exemplary literature. However, the table represents examples of possible approaches and is not exhaustive.

Table 2: Possible semi-automatic Classification Methods for forest – parameters according to European and national NATURA 2000 guidelines (++ = method very good applicable; + = method good applicable; — = method not applicable)

| Policy-based Assessment Parameters | Parameters to Ascertain with Remote sensing | Conditions | | | | Method | Classification scale dependent feasibility | | Reference studies |
|---|---|---|---|---|---|---|---|---|---|
| | | spectral | temporal | spatial | ancillary data & LIDAR | | Biogeographic Level (e.g. Spot5) | SAC-Level (e.g. QuickBird) | |
| *Detection of Habitat extent and change of Habitats* | Main forest types or tree species | NIR-band | vegetation period | ≤ 15m | DEM<br>Forestry Site Map<br>Further soil maps | GIS-based modelling in combination with classification | ++ | + | see Chapter III |
| *National Quality Parameters* | | | | | | | | | |
| *Habitat Structures* | | | | | | | | | |
| Number of Forest Development Stages (e.g. for 9110: >3; 2; less than 2) | Tree heights<br>Structural characteristics | | vegetation period | < 5m | LIDAR | Analysis of height related to age classes<br>Object-based analysis | — | ++ | Heurich 2006<br>Maltamo et al. 2005 |
| Number of Biotope Trees (e.g. for 9110: ≥6; ≥3; < 3 per ha) | — | | | | | | — | — | |
| Number of Deadwood (e.g. for 9110: >3; >1; ≤ 1 per ha) | Canopy gap mapping<br>Deadwood detection | NIR-band | winter image | < 1m | LIDAR | Support Vector Machines<br>Manual delineation | — | + | Nuske et al. 2007 |
| *Availability of typical Species in the Habitat* | | | | | | | | | |
| Percentage of Natural Forest Types (e.g. 9110: ≥90%; ≥80%; ≥70%) | Tree species | NIR-band | vegetation period | ≤ 1m | DEM<br>Forestry Site Map<br>LIDAR | Integrated classification approach (e.g. Fuzzy Logic) | — | ++ | see Chapter IV |
| Quality of understory herb layers | Tests for single species (e.g. Lichen) | | | ≤ 1m | LIDAR | LIDAR intensities | — | (+) | Korpela In Press |
| Faunal quality | — | | | | | | — | — | |
| *Disturbances of the Habitat Type* | | | | | | | | | |
| Disturbances of Soil and Hydrology (e.g. Erosion) | Gully mapping | | | | LIDAR | LIDAR last pulse | — | + | James et al. 2007 |
| Disturbances of forest structure (e.g. changes in relief, influence of neighbouring vegetation) | Tree species<br>Successional stages of forest | NIR-band | vegetation period | ≤ 1m | DEM<br>Forestry Site Map<br>LIDAR | Object-based and knowledge-based classification | — | ++ | Hajek 2008 |
| Disturbances caused by forest fragmentation (e.g. road building)) | Forest/non-forest mask (Average Patch Size, fragmentation indices) | NIR-band | vegetation period | ≤ 15m | | Knowledge-based classification landscape metrics | + | ++ | Oehmichen 2007 |
| *Biogeographic Level* | | | | | | | | | |
| Range | Main forest types or tree species<br>Minimum patch size | NIR-band | vegetation period | ≤ 15m | DEM<br>Forestry Site Map<br>Further soil maps | GIS-based modelling in combination with classification | ++ | — | Similar Chapter III |
| Area | Main forest types or tree species<br>Proximity index | NIR-band | vegetation period | ≤ 15m | DEM<br>Forestry Site Map<br>Further soil maps | GIS-based modelling in combination with classification | ++ | — | Similar Chapter III |
| Population | — | | | | | | — | — | |

## 1.2 Integration of geo-data

*Research Question: How can multisource data be included in a classification process to improve the classification results?*

### 1.2.1 Classification of forest types

The thesis shows that the classification accuracy of the investigated forest types can be enhanced by including ancillary data and is influenced by the type of geo-data and the applied integration technique. When comparing the results with and without additional data an average increase of 13 % (from 64 % to 77 %) of the overall accuracy was observed for the Angelberger Forst (see Chapter IVa), while the average increase for the Taubenberg was lower (6 % - from 65 % to 71 %). This is comparable to other studies which investigated the classification accuracy when including ancillary geo-data for different applications for Landsat TM data (Bolstad and Lillesand 1992; Janssen et al. 1990). Especially the identification of forest types with narrow ecological niches, such as Black Alder (from 17 % to 98 %) or Sycamore (from 68 % to 88 %), has been distinctly increased. However, classification accuracy for forest types with wide ecological niches, such as Spruce or Beech, is similar for both methods (difference less than 10 %).

It can be concluded that the increase in classification accuracy depends on the **type of geo-data** which are included in the classification process (see Table 3). The classification accuracy for the Angelberger Forst was derived including the Forest Site Map (FSM), while the Taubenberg lacks this data source. If the research area Angelberger Forst was classified without including the FSM (as it was calculated in Chapter V – classification case "NO-FSM"), the accuracy decreases to 68 %. Therefore, the relief information (DTM) as well as soil and hydrology information (FSM) both contribute to increased classification accuracy. The best classification result, however, can be achieved using all available sources of geo-data.

Table 3: Comparison of the overall accuracies for the research areas including different types of additional data and different integration methods

| | **Pure satellite-based classification** | **Fuzzy classification incl. all available data sources** | **Fuzzy classification incl. comparable data sources (no FSM)** | **Threshold-based Classification** |
|---|---|---|---|---|
| Angelberger Forst | 0.64 | 0.77 | 0.68 | 0.70 |
| Taubenberg | 0.65 | 0.71 | 0.71 | - |

Moreover, classification accuracy depends on the **integration technique** of ancillary data. As in Chapter V discussed geo-data are more valuable, when applying them together with a knowledge-base. The successful implementation of additional geo-data with different types of a knowledge-base and for different vegetation types was shown in Chapter IV.

To directly compare different data integration methods the fuzzy implementation (see Chapter IVa) was repeated with a crisp rule-base for the same research area (Förster and Kleinschmit 2007). This classification worked with an identical rule-set for the additional data. However, the rules were integrated via thresholds instead of fuzzy membership functions. As threshold, the weighted mean of the fuzzy rules were used (Tilli 1993). The classification with additional data of crisp thresholds obtained better results than the pure spectral classification (70 % instead of 64 %). However, the major improvement of including additional data and knowledge is only seen when the cognitive uncertainties and the imprecise formulation of the information are acknowledged via fuzzy logic (see Table 3). Fuzzy logic models are specified and built from linguistic statements, based on common sense. Therefore, a greater potential of available information can be utilised (Openshaw and Openshaw 1997). Hence, for applications dealing with target classes which are closely linked to environmental resources fuzzy logic approaches can be recommended for the integration of geo-data into a classification process.

### 1.2.2 Classification of habitat types

Contrary to the identification of forest types, validating the classification of habitat types is more difficult to implement. This is due to the fact that a habitat consists of a variety of vegetation classes, which have to be combined in a specific composition. For forested habitats this composition is often related to a main tree species such as sycamore or black alder. Nevertheless, some of the main tree species can account for different habitat types, as it is the case for beech (e.g. *Luzulo-Fagetum* – habitat type 9110 and *Asperulo-Fagetum* – habitat type 9130), due to the fact that other species form the specific characteristics of the habitat type. Moreover, the allocation of a NATURA 2000 habitat type does not require a full coverage of a main tree species. Therefore, the spatial extent of a delineated habitat is interrelated to its internal classes. This so-called modifiable areal unit problem (Openshaw 1984) makes it difficult to validate the results derived by remote sensing methods, because different combinations of a more detailed scale (forest type objects) to a larger scale (habitat type objects) are possible.

One method to validate and compare the results is the subsumation of the classes to generalised categories. This was shown in Chapter III in combining single forest classes to deciduous, mixed, and coniferous forest. However, only very limited conclusions could be

drawn concerning the method which performed best in integrating additional data. The multivariate clustering delivers best results for habitat types which are strongly influenced by the relief. The fuzzy logic-based and the rule-based modelling supplied similar results when compared to the validation basis. The findings are comparable to another study, which is using a rule-based technique for deriving forested habitat types (Langar 2008). However, due to the methodological difficulties to find a definition of the spatial extent of a habitat type, all modelling techniques can equally be seen as valuable concepts of supporting and complementing terrestrial mapping.

Similarly, quality parameters for habitat types can be derived by the integrated classification technique (see Chapter IV). The results of these parameters are difficult to validate, for the same reasons as the modelling approaches. Exemplarily, two segmentation scales were used in Chapter IVa. It is shown that for selected parameters the share of habitat qualities varies with an altering scale. With the given quality parameter monitoring standards the questions of which is the more correct result or which influence have additional information to these scales is not answered finally, up to now.

Obviously a variety of other factors influence the habitat quality. Nevertheless, the factor scale has been rather neglected in the mapping guidelines or monitoring results. The challenge with validating habitat types could be minimised by defining a standard spatial reference size (e.g. minimum mapping units). This question should be addressed to ecologists and included in monitoring guidelines. If a certain habitat requires a coherent large area, a larger segmentation scale should be applied, while small-sized habitats should be classified with a finer object size. Additionally, validation methods which are not completely depending on the defined spatial extend of an object, such as fuzzy accuracy assessments, could be applied to reduce the modifiable areal unit problem (Schiewe and Gähler 2008).

## 1.3 Significance analysis

*Research Question: By what technique can the significance of different types of geo-data be quantified? Which of the available geo-data concerning this case-study are significant for the classification success?*

To investigate the significance of geo-data, the accuracy of the forest type classes was assessed for the fuzzy logic based classification (Chapter IVa). 50 variations of classifications, including different sets of rules and geo-data were therefore classified and subsequently validated. The resulting microarray of accuracy percentages of single classes and the overall classification accuracy were used for further investigations.

Two methods of evaluating the significance of multiple classification accuracies were presented in this thesis (see Chapter V). While the ISODATA clustering provided valuable information about patterns in the accuracy data, the Significance Analysis of Microarrays (SAM) can identify significant influences of different types of additional data to a variety of class accuracies. This contributes to a better evaluation of classification results, especially when using multiple spatial data-types. The SAM method provided reliable values of significance. Different to the ISODATA clustering, significant geo-data or rules can be detected and interpreted. The analysis provides information about separate interactions within the classification results of the included rules. Therefore, SAM is a suitable technique for evaluating knowledge-based classifications using ancillary geo-data.

In the particular case of forest type classification, soil factors of the Forest Site Map represent the most significant data-type, while the factors derived from the Digital Terrain Model are somewhat less relevant. One reason for this result is that some variables are not sufficiently important for the vegetation classes in this specific region. As an example, the rules related to aspect are probably not significant because the influence of a more shadowed or sunny hillside on the occurrence of a forest type is not of major relevance in this region. Moreover, a large dependency on the included data quality was observed. Data that are more adapted to a specific application are more important for the classification success. For example, soil information from the Forest Site Map yielded highly significant results, while soil data from the Conceptual Soil Map (which is not specifically produced for forest applications) resulted in lower values.

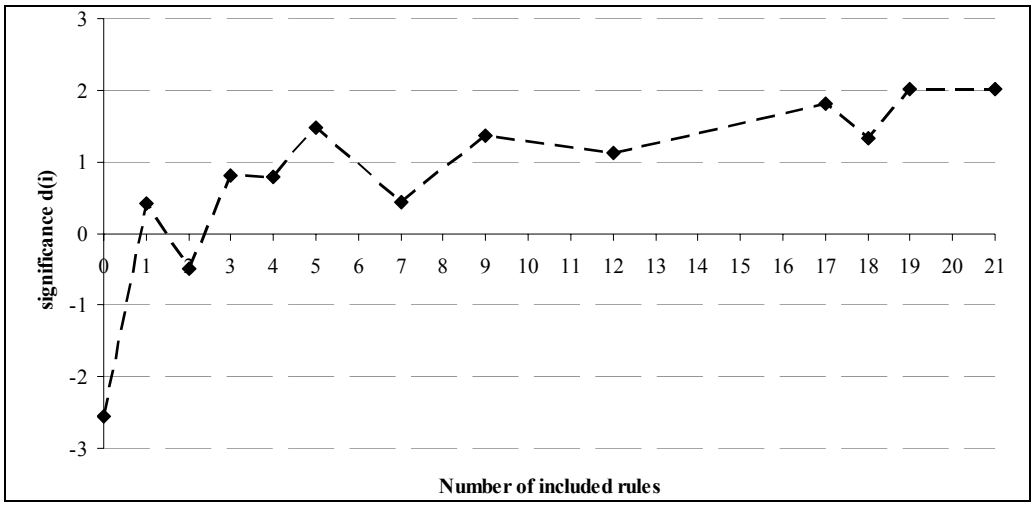



Figure 1: Significance *d(i)* of included rules for the classification accuracy per included rules as obtained when applying the SAM-method to the study area "Angelberger Forest" (see Chapter V).

Generally, for the tested areas it can be stated that more geo-data and more rules of a knowledge-base influence the classification process positively (see Figure 1). With an

increasing number of included rules the average measurement of the SAM parameter "relative difference" – *d(i)*, which is a value of significance, was increasing. This leads to the conclusion that inefficient rules or geo-data do not obstruct a successful classification, as long as they are not contradicting the results of more significant geo-data and rules. The reduction of *d(i)* when including 2, 7, or 18 rules is related to the different composition of the rule-sets contributing to the *relative difference*. In the appointed composition, the absence of rules which include the Forest Site map (as one of the main driver for high significance) explains the decrease of the results.

# 2 Future research

In this thesis it was demonstrated that the analysis of data integration into a remote sensing classification process for NATURA 2000 monitoring requires comprehensive interdisciplinary approaches, including methods from different research fields. Several interesting issues for follow-up research beyond the scope of this work evolved during the course of this thesis. Amongst them, transferability of the methods, adapted processes for NATURA 2000 monitoring under changing climate, and the utilisation of further data-integration techniques will be discussed as potential fields for future research.

## 2.1 Transferability

Since the NATURA 2000 network covers large areas of Europe, the possible transferability of the classification methods (see Chapters III and IV) is suggested to be tested and analysed in more detail. As the presented techniques in this thesis were applied to two research areas a transfer of the methods can be assumed for deciduous forest habitats of the continental biogeographic region. However, these results were limited to forested areas in the pre-alpine region of the continental biogeographic region.

For classifications at the **SAC-level** different studies proved the applicability for diverse habitat types (Bock et al. 2005; Weiers et al. 2004). However, the large variety of ecosystems with different sets of additional specific data makes it impossible to find a single classification method that is applicable for all habitats. Even for the quality parameters of forested habitats a huge selection of approaches exists (see a selection of examples in Table 2). However, joint European actions, such as Global Monitoring for Environment and Security (GMES) with their Global Service Element – GSE-*nature* aim at combining different research outputs to an overall monitoring strategy.

At the **biogeographic level** the modelling approach presented in Chapter III can be recommended for extension to larger areas for defined habitats. This was already shown for the federal state of Bavaria (Förster and Kleinschmit 2006c). However, the approach was applied only to two very similar areas in a single biogeographic region. Because of the small variability of the "Taubenberg" and the "Angelberger Forst" the method required other application cases of different preconditions concerning data availability, natural resources, and habitat types.

Therefore, the rule-based approach of Chapter III was successfully transferred to forested areas in the Mediterranean biogeographic region (Förster et al. 2007; Velazquez et al. 2008). The forest test area "Ávila" (Castilla y León Region) is located in Central Spain. The medium altitude of the area is 1308 m. The region is characterised by a fresh mild Mediterranean weather, with hydric stress during the summer. Ávila is dominated by a natural Oak Forest stand of *Luzulo foresteri-Quercetum pyrenaicae* (habitat type 9230) and patches of *Junipero oxycedri-Quercetum rotundifoliae* (9340).

The general approach of rule-based modelling of the potential natural vegetation combined with a remote sensing classification proved to be transferable between the regions. However, the rule-base had to be adapted to the given geo-data and habitat conditions in the Ávila region in Spain. Once implemented, strategies of building rule-sets were identified. While for the Ávila region preferably exclusive rules[22] were used due to the local conditions (which are defined by shortage of available water), in the Taubenberg region more additive rules[23] were applied (see Figure 2). The share of exclusive rules at the Taubenberg was 12.7 %, while in the Ávila region 34.9 % of such rule-sets were applied. Therefore it is concluded that, if different strategies in building the rule-base are considered, the modelling method (see Chapter III) can be suggested for transferring to other areas.

However, there are large differences within the biogeographic regions. The continental region includes areas of such different characteristics as the pre-alpine zones, the Baltic Sea coastal area, and the Po Plain. To be able to transfer the presented approach, sub-regions and adapted rule-sets have to be defined.

---

[22] Exclusive rules are defined for locations with high and excludable suitability for a habitat type. As an example, the possibility of the existence of bog woodland on brown soils will be excluded, while the existence on peaty soils is highly possible.

[23] Additive rules are defined for locations with possible occurrences of a habitat type. As an example, a habitat type can grow on different relief types as well as on different soil types. Therefore a site can be chosen as a possible habitat type due to several parameters. In this case, the number of possible occurrences was summed up.

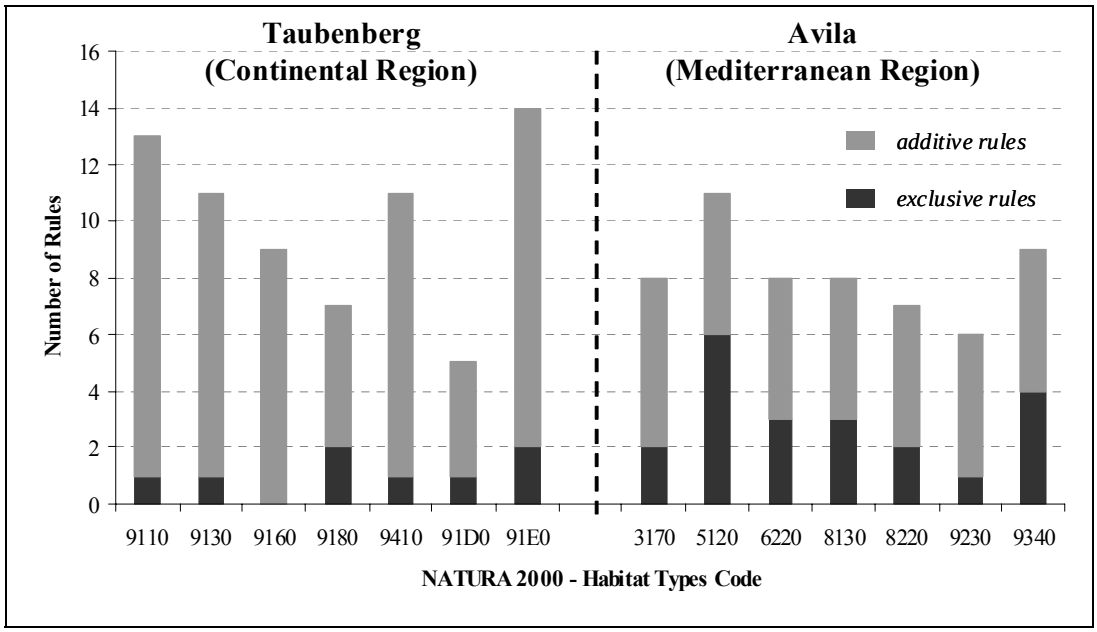



Figure 2: Comparison of the rule-bases for two classifications in the Continental and Mediterranean biogeographic region. The share of exclusive rules is higher in Ávila than in Taubenberg.

## 2.2 Challenge of climate change

The NATURA 2000 network, although originally not initiated with climate change motivation, can significantly support the aim of reducing the loss of biodiversity under the circumstances of altering natural conditions. However, depending on the specific purpose of the protected areas, the conservation targets might be missed under climate change. Therefore it is necessary to face new challenges when observing and managing habitat diversity (Neubert et al. 2008a).

Climate change and its impact on habitats with respect to shifting temperature and precipitation patterns, increasing probabilities and intensities of natural hazards as well as the interrelation with increasing $CO_2$ and changed nutrient dynamics (e.g. N) altogether require adaptive management strategies. Only a limited number of the interdependent effects of climate change on habitats in protected areas can be clearly addressed in advance. Most of the changes start gradually without large impacts; they can be detected and managed early enough only if an effective monitoring system is available (Hannah and Salm 2005). Therefore, monitoring linked to management adaptations is perhaps the most important part of managing climate change effects (Hannah et al. 2002). For that purpose, enhanced monitoring methods are required, with clear attention on vulnerable habitats which represent specific ecological niches. Because the effects of climate change on habitats will vary in extent and intensity the monitoring cycles have to be adapted. The restriction of a static six-

year interval for the monitoring of NATURA 2000 habitats should be reconsidered especially in the context of climate change.

To detect the spatial extent of the most vulnerable part of habitats, a combined remote sensing and modelling approach is recommended. It could link the detection of possible actual changes with scenario-based models to a map of habitat locations which are especially vulnerable to climate change. One possibility to achieve current information about the situation of habitats is the preparation of phenologic profiles, which can be compared to multi-temporal satellite imagery. In contrast to long-term multi-temporal studies which detect inter-annual changes, phenologic profiles rely on several acquired images during the vegetation period. Similar to the method developed for agricultural applications (Itzerott and Kaden 2007), each image could be compared to a series of terrestrial spectral measurements. Because of the broadly varying phenologic patterns of different habitats, a high level of differentiation can be expected. At a second stage, changes of habitats or even their damages or stresses could be distinguished from undisturbed habitats.

With a combined remote sensing and modelling technique it might be possible to develop regional management strategies for a climate change adapted management. A possible methodology is summarised in Table 4 for two exemplary habitat types, including forest.

Table 4: Possible climate change impacts, indicators based on modelling and remote sensing, as well as adapted management strategies for the ecosystems forest and wetland

| | **Forest** | **Wetland** |
|---|---|---|
| Climate change implications | Water stress<br>Insect pests<br>Increased storm risk and intensity<br>Altered competition of tree types | Eutrophication (and subsequently mineralisation)<br>Increased evaporation<br>Decreased groundwater capacity |
| Remote sensing-based indicators | Forest composition<br>Forest health (phenology based)<br>Identification of storm affected areas | Adjacent land use<br>Vegetation change (e.g. grasses)<br>Multi-temporal moisture index |
| Climate change model-based indicators | Shifts in the composition of the natural vegetation<br>Changes in productivity | Changes in abiotic conditions, especially soil moisture and run off |
| Adapted management measures for vulnerable areas | Changed silviculture (drought resisting forest types)<br>Forest according to soil and wetness conditions of the site<br>Mixed deciduous forest without monocultural stand structures | Management of agriculture in adjancent areas<br>Change in water/groundwater management |

## 2.3 Data integration and availability

This thesis introduced methods of integrating geo-data and additional information into remote sensing classification processes. Although these methods proved to have a positive effect on the resulting products, there is a variety of further multisource classification approaches as well as ancillary geo-data available to be tested, developed, and compared.

Since the presented **classification methods** were utilised with object-based procedures, a further analysis could compare different methods. Recent advances were made with Support Vector Machines (Tzotsos and Argialas 2008) as well as domain specific class modelling (Tiede et al. 2008). A classification for the same research area and target classes with a successive significance analysis could give insights into the mechanisms of data-integration. Furthermore, implications of different scales when using segmentation algorithms could be focussed on. As shown in Chapter IV, altered segmentation parameters change the extent of habitats as well as their quality description. An estimation of optimal image object sizes for forest stands (Kim et al. 2008) could improve the reliability of classification and could give a recommended minimum mapping unit for further NATURA 2000 monitoring.

The available **geo-data** in this thesis were restricted to a specific country (Germany) or even federal state (Bavaria) and a specific ecosystem (forest). It would be a challenging task to compare the integration of geo-data adapted to specific applications and more abstracted, general sources, such as CORINE land cover or GSE products. Moreover, the influence of the mapping scale of geo-data on the integration process needs to be assessed. Firstly, it is relevant to know, whether a small-scale data-set correlates to smaller object sizes. Secondly, the scale influences the detail of the attributes of data-sets. Therefore, findings about the influence of more generalised attributes on data integration would advance the insights into the value and significance of multisource classifications.

The development of new sensors of high spectral (e.g. EnMap – proposed launch 2011), spatial (e.g. GeoEye-1 – launched September 2008), and temporal (e.g. RapidEye – launched September 2008) resolution will improve the possibilities of classification methods as well as data availability and costs.

Anyhow, the step from purely scientific applications to operational systems which integrate different sources of information depends on the harmonisation of geo-data. If there is a future with standardised spatial information, there should be a future with remote sensing monitoring programs including these different sources of additional information.

# References


AFL (2003). *Luftbildinterpretation: Bestimmungsschlüssel für die Beschreibung von strukturreichen Waldbeständen im Color-Infrarot-Luftbild*. (pp. 48). Pirna: Arbeitsgruppe Forstlicher Luftbildinterpreten

Ahrens, W., Brockamp, U., & Pisoke, T. (2004). *Zur Erfassung von Waldstrukturen im Luftbild*. (pp. 54). Freiburg: Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg

AK Standortkartierung (1996). *Forstliche Standortsaufnahme - Begriffe, Definitionen, Einteilungen, Kennzeichnungen, Erläuterungen*. Eching b. München: IHW-Verlag

Arabie, P., Hubert, L.J., & De Soete, G. (Eds.) (1998). *Clustering and Classification*. (pp. 508). New York: World Scientific

Baatz, M., & Schäpe, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: J. Strobl, T. Blaschke & G. Griesebner (Eds.), *AGIT* (pp. 12-23). Salzburg: Wichmann

Baban, S., & Yusof, K. (2001). Mapping land use/cover distribution on a mountainous tropical island using remote sensing and GIS. *International Journal of Remote Sensing, 22*, 1909-1918

Baltsavias, E.P. (2004). Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry & Remote Sensing, 58*, 129-151

Barnsley, M.J., & Barr, S.L. (1996). Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric Engineering & Remote Sensing, 62*, 949-958

Benediktsson, J.A., Chanussot, J., & Fauvel, M. (2007). Multiple Classifier Systems in Remote Sensing: From Basics to Recent Developments. In: M. Haindl, J. Kittler & F. Roli (Eds.), *Multiple Classifier Systems* (pp. 501-512). Berlin: Springer

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

BfN (2008). FFH-Gebietsmeldungen von Deutschland. Bonn: Bundesamt für Naturschutz (p. 1). Report

Blaschke, T. (Ed.) (2001). *Fernerkundung und GIS. Neue Sensoren - innovative Methoden*. (pp. 271). Heidelberg: Wichmann

Blaschke, T., & Felbermeier, B. (2003). Satellitendaten für die Forstwirtschaft - Nutzeranforderungen der forstlichen Praxis und Potential der optischen Satellitenfernerkundung. *Forstliche Forschungsberichte, 196*, 93-108

Blaschke, T., & Strobl, J. (2001). What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS*, 12-17

Bobbink, R., Beltman, B., Verhoeven, J., & Whigham, D. (2008). *Wetlands: Functioning, Biodiversity Conservation, and Restoration*. (pp. 316). Berlin: Springer

Bock, M., Xofis, P., Mitchely, J., Rossner, G., & Wissen, M. (2005). Object-oriented methods for habitat mapping at multiple scales – Case studies from Northern Germany and Wye Downs, UK. *Journal for Nature Conservation, 13*, 75-89

Bolstad, P., & Lillesand, T.M. (1992). Rule-Based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data. *Photogrammetric Engineering & Remote Sensing, 58*, 965-971

Boteva, D., Griffiths, G., & Dimopoulos, P. (2004). Evaluation and mapping of the conservation significance of habitats using GIS: an example from Crete, Greece. *Journal for Nature Conservation, 12*, 237-250

Breiman, L. (2001). Random Forests. *Machine Learning, 40*, 5-32

Briem, G.J., Benediktsson, J.A., & Sveinsson, J.R. (2002). Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing, 40*, 2291-2299

Bronge, L. (1999). Mapping boreal vegetation using Landsat TM and topographic map data in a stratified approach. *Canadian Journal of Remote Sensing, 25*, 460-474

Bruzzone, L., Prieto, D.F., & Serpico, S.B. (1999). A Neural-Statistical Approach to Multitemporal and Multisource Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing, 37*, 1350-1359

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

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

Campbell, J.B. (2002). *Introduction to Remote Sensing*. (pp. 620). New York: Taylor & Francis

Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J., Martin-Guerrero, J.D., Soria-Olivas, E., Alonso-Chorda, L., & Moreno, J. (2004). Robust Support Vector Method for Hyperspectral Data Classification and Knowledge Discovery. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 1530-1542

Chen, C.H., & Ho, P.G. (in press). Statistical pattern recognition in remote sensing. *Pattern Recognition*, 1-40

Chu, G., Narasimhan, B., Tibshirani, R., & Tusher, V.G. (2008). SAM-Significance Analysis of Microarrays-Users guide and technical document. Stanford (pp. 1-41). Software Documentation

Coenradie, B. (2003). Waldzustandserfassung und -monitoring mit hochauflösenden Satellitenbilddaten. Berlin: Technische Universität Berlin (p. 281). Dissertation

Cohen, Y., & Shoshany, M. (2002). A national knowledge-based crop recognition in Mediterranean environment. *International Journal of Applied Earth Observation and Geoinformation, 4*, 75-84

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

de Kok, R. (2001). Objektorientierte Bildanalyse - Ein Lösungsansatz für den automatisierten Einsatz sehr hoch auslösender Satellitendaten für forstliche Fragestellungen. Weihenstephan: TU München (p. 112). Dissertation

de Lange, N. (2002). *Geoinformatik in Theorie und Praxis*. (pp. 436). Berlin: Springer

Dees, M., Koch, B., & Pelz, D.R. (1998). Designalternativen zur Integration von Satellitendaten in das Design der Bundeswaldinventur. In: D.R. Pelz, O. Rau & J. Saborowski (Eds.), *11. Jahrestagung der Sektion Forstliche Biometrie und Informatik des DVFF* (pp. 97-117). Freiburg: Universität Freiburg

Diedershagen, O., Koch, B., & 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* (pp. 208-212). Freiburg

Digital Globe (2005). QuickBird Imagery Products, FAQ. http://www.digitalglobe.com /product/product_docs.shtml. March 2006

Digital Globe (2006a). Product information, imaging systems, ordering and pricing - standard imagery. http://www.digitalglobe.com/product/standard_imagery.shtml. March 2006

Digital Globe (2006b). Product information, imaging systems, ordering and pricing. http:// www.digitalglobe.com /product/. March 2006

Digital Globe (2006c). QuickBird Imagery - FAQ. http://www.digitalglobe.com/ press/FAQ_press.shtml#t11. June 2006

Dimitrakopoulos, P., Memtsas, D., & Troumbis, A. (2004). Questioning the effectiveness of the Natura 2000 Special Areas of Conservation strategy: the case of Crete. *Global Ecology and Biogeography, 13*, 199-207

Dirnböck, T., Dullinger, S., Gottfried, M., Ginzler, C., & Grabherr, G. (2003). Mapping alpine vegetation based on image analysis, topographic variables and Canonical Correspondence Analysis. *Applied Vegetation Science, 6*, 85-96

Dirzo, R., & Raven, P. (2003). Global State of Biodiversity and Loss. *Annual Review of Environmental Ressources, 28*, 137-167

Döring, C., & Jansen, M. (2005). Mixed or mixed? - Spatial simulation of multifunctional forest stand development in the Solling area. *Multifunctionality of Landscapes*. Gießen

Duda, R., Hart, P., & Stork, D. (2001). *Pattern Classification*. (pp. 637). New York: John Wiley

Duro, D., Coops, N.C., Wulder, M., & Han, T. (2007). Development of a large area biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography, 31*, 235-260

Efron, B., & Tibshirani, R. (in press). On testing the significance of sets of genes. *Annals of Applied Statistics, 1*, 1-31

Ehlers, M. (2002). Fernerkundung für GIS-Anwender - Sensoren und Methoden zwischen Anspruch und Wirklichkeit. In: T. Blaschke (Ed.), *Fernerkundung und GIS - Neue Sensoren - innovative Methoden* (pp. 10-23). Heidelberg: Wichmann

Ehlers, M., & Klonus, S. (2004). Erhalt der spektralen Charakteristika bei der Bildfusion durch FFT basierte Filterung. *PFG*, 495-506

Eisenbeiss, H., Baltsavias, E., Pateraki, M., & Zhang, L. (2004). Potential of Ikonos and Quickbird imagery for accurate 3D point positioning, orthoimage and DSM generation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part 3*, 522-528

Eiumnoh, A., & Shresta, R.P. (2000). Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand. *Photogrammetric Engineering & Remote Sensing, 66*, 297-304

Ellenberg, H. (1996). *Vegetation Mitteleuropas mit den Alpen in ökologischer, dynamischer und historischer Sicht*. (pp. 1096). Stuttgart: Ulmer

Eurimage (2006a). Eurimage Reseller Germany. http://www.eurimage.com/reseller/area/germany.html. March 2006

Eurimage (2006b). Products, services, and prices. http://www.eurimage.com. March 2006

European Commission (1979). Council Directive 79/409/EEC. *Official Journal, L103*, 1-18

European Commission (1992). Council Directive 92/43/EEC. *Official Journal, L206*, 1-66

European Commission (2003a). Bericht der Kommission über die Umsetzung der Richtlinie 92/43/EWG zur Erhaltung der natürlichen Lebensräume sowie der wild lebenden Tiere und Pflanzen. Brüssel (pp. 44-49)

European Commission (2003b). Interpretation Manual of European Union Habitats - EUR 25. Brüssel (pp. 89-118)

European Commission (2005). Assessment, monitoring and reporting of conservation status - Preparing the 2001-2007 report under Article 17 of the Habitats Directive. Brussels (pp. 1-10)

European Commission (2006). Assessment, monitoring and reporting under Article 17 of the Habitats Directive - explanatory notes & guidelines (Draft 4). Brussels (pp. 1-66)

European Commission (2007a). Directive 2007/2/EC of the European Parliament and the Council establishing an Infrastructur for Spatial Information in the European Community (INSPIRE). *Offical Journal of the European Union, 50*, 1-14

European Commission (2007b). Interpretation Manual of European Union Habitats: European Commission, DG Environment (pp. 1-144)

European Commission, & European Environment Agency (2002). Continental Region - Referene List of habitat types and species present in the region. Paris: European Environment Agency (pp. 1-12)

Fitzgerald, R.W., & Lees, B.G. (1994). Assessing the Classification Accuracy of Multisource Remote Sensing Data. *Remote Sensing of Environment, 47*, 326-368

Foody, G.M. (2004). Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing, 70*, 627-633

Foody, G.M., & Mathur, A. (2004). A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 1335-1343

Forgy, E.W. (1965). Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classifications. *Biometrics, 21*, 755-768

Förster, M., Frick, A., Walentowski, H., & Kleinschmit, B. (in press). Approaches of utilising Quickbird-Data for the Monitoring of NATURA 2000 habitats. *Community Ecology*

Förster, M., & Kleinschmit, B. (2006a). Einfluss von Segmentierungsskalen auf Strukturparameter zur wissensbasierten Klassifikation am Beispiel von NATURA 2000 Lebensraumtypen. In: B. Kleinschmit & U. Walz (Eds.), *Landschaftsstrukturmaße in der Umweltplanung - Beiträge zum Workshop der IALE-AG Landschaftsstruktur* (pp. 109-119). Berlin: Technische Universität Berlin

Förster, M., & Kleinschmit, B. (2006b). Integration of Ancillary Information into Object-based Classification for Detection of Forest Structures and Habitats. In: S. Lang, T. Blaschke & E. Schöpfer (Eds.), *1st International Conference on Object-based Image Analysis* (pp. 1-6). Salzburg, Austria: ISPRS

Förster, M., & Kleinschmit, B. (2006c). Remote Sensing and GIS-Modelling for the Monitoring of NATURA 2000 Habitats. In: E. Csaplovics (Ed.), *International Symposium on Geoinformatics in European Nature Protection Regions - NatureProtection:GIS* (pp. 63-74). Dresden: TU Dresden - Eigenverlag

Förster, M., & Kleinschmit, B. (2007). Comparison of Different Integration Techniques of Ancillary Information into an Object-Based Classification Process. In: C. Heipke, K. Jacobsen & M. Gerke (Eds.), *ISPRS Workshop on High Resolution Earth Imaging for Geospatial Information* (pp. 1-5). Hannover

Förster, M., & Kleinschmit, B. (2008). Object-based classification of QuickBird data using ancillary information for the detection of forest types and NATURA 2000 habitats. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 275-290). Berlin: Springer

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

Förster, M., Kleinschmit, B., & 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 (Ed.), *ForestSat* (pp. 21-25). Borås, Sweden: Swedish National Board of Forestry

Förster, M., Velazquez, J., & Kleinschmit, B. (2007). Modelling of NATURA 2000 Habitat Types in Different Biogeographic Regions - Experiences from Spain and Germany. *ForestSat 2007*. Montpellier, France: Cemagref

Frick, A. (2007). Beiträge höchstauflösender Satellitenfernerkundung zum FFH-Monitoring - Entwicklung eines wissensbasierten Klassifikationsverfahrens und Anwendung in Brandenburg. Berlin: Technische Universität Berlin. Dissertation

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

Friedl, M.A., & Brodley, C.E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment, 61*, 399-409

Gähler, M., Janowsky, R., & Schiewe, J. (2004). Automatisierte Ableitung von Biotoptypen aus räumlich höchstaufgelösten, multispektralen Bild- und Laserscanning-Daten. *PFG*, 475-484

Gerke, M. (2002). Scene Analysis in Urban Areas Using a Knowledge-Based Interpretation System. *ISPRS Com. III Symposium* (pp. 63-66). Graz

Gillespie, T.W., Chu, J., Frankenberg, E., & Thomas, D. (2007). Assessment and predictions of natural hazards from satellite imagery. *Progress in Physical Geography, 31*, 459-470

Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., & Saatchi, S. (2008). Measuring and modelling biodiversity from space. *Progress in Physical Geography, 32*, 203-221

Gislason, P.O., Benediktsson, J.A., & Sveinsson, J.R. (2004). Random Forest Classification of Multisource Remote Sensing and Geographic Data. *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International* (pp. 1049-1052): ISPRS

Glavac, V. (1996). *Vegetationsökologie*. Jena: Fischer

Golden, M., & Lackey, L. (1992). Using Ancillary Data in Post-Classification Modeling to Increase the Accuracy of Conifer Species Classification from Landsat Thematic Mapper Data. *GIS `92 Symposium* (pp. 23-30). Vancouver, British Columbia, Canada

Granke, O., Schmiedinger, A., & Walentowski, H. (2004). Konzept und Schlüsselkriterien für die Bewertung der Biodiversität von Wald-Lebensräumen in Deutschland. *waldökologie online, 1*, 25-28

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

Hahn, M., & Baltsavias, E. (1998). Cooperative Algorithms and Techniques of Image Analysis and GIS. *IAPRS, 32*, 1-5

Hajek, F. (2008). Process-based approach to automated classification of forest structures using medium format digital aerial photos and ancillary GIS information. *European Journal of Forest Research, 127*, 115-124

Hannah, L., Midgley, G., & Millar, D. (2002). Climate change-integrated conservation strategies. *Global Ecology and Biogeography, 11*, 485-495

Hannah, L., & Salm, R. (2005). Protected Areas Management in a Changing Climate. In: T.E. Lovejoy & L. Hannah (Eds.), *Climate Change and Biodiversity* (pp. 363-371). New Haven: Yale University Press

Harris, P., & Ventura, S. (1995). The Integration of Geographic Data with Remotely Sensed Imagery to Impove Classification in Urban Area. *Photogrammetric Engineering & Remote Sensing, 61*, 993-998

Helmer, E., Brown, S., & Cohen, W. (2000). Mapping montane tropical forest successional stage and land use with multi-date Landsat imagery. *International Journal of Remote Sensing, 21*, 2163-2183

Hese, S. (2001). Entwicklung und Anwendung geostatistischer Verfahren zur strukturellen Klassifizierung von Forstbeständen in hochauflösenden Daten der HRSC-A. Berlin: Technische Universität Berlin (p. 94). Dissertation

Heurich, M. (2006). *Evaluierung und Entwicklung von Methoden zur automatisierten Erfassung von Waldstrukturen aus Daten flugzeuggetragener Fernerkundungssensoren*. (pp. 328). München: TU München

Hildebrandt, G. (1969). *Bibliographie des Schrifttums auf dem Gebiet der forstlichen Luftbildauswertung: 1887 : 1968*. (pp. 307). Freiburg: Institut für Forsteinrichtung und forstliche Betriebswirtschaft der Universität Freiburg

Hill, R.A., Granica, A., Smith, G.M., & Schardt, M. (2005). Characterization of alpine treeline ecotones: an operational approach? In: H. Olsson (Ed.), *ForestSat 2005* (pp. 74-78). Boras, Sweden: Swedish National Board of Forestry

Hinton, J. (1999). Image classification and analysis using integrated GIS. In: P. Atkinson & N. Tate (Eds.), *Advances in Remote Sensing and GIS Analysis* (pp. 207-218). New York: Wiley

Hirschmugl, M., Ofner, M., Raggam, J., & Schardt, M. (2007). Single tree detection in very high resolution remote sensing data. *Remote Sensing of Environment, 110*, 533-544

Hoffmann, B. (2001). Untersuchungen zum Monitoring von Naturwäldern mit Hilfe von Geo-Informationssystemen, modernen Luftbildauswertungsverfahren und Geostatistik - dargestellt am Beispiel des nordrhein-westfälischen Naturwaldreservats Hellerberg. Göttingen: Georg-August-Universität Göttingen. Dissertation

Huang, K. (2002). A Synergistic Automatic Clustering Technique for Multispectral Image Analysis. *Photogrammetric Engineering & Remote Sensing, 68*, 33-40

Hutchinson, C. (1982). Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering & Remote Sensing, 48*, 123-130

Itzerott, S., & Kaden, K. (2007). Gütebewertung für die Klassifizierung landwirtschaftlicher Fruchtarten aus spektralen Normkurven. *PFG*, 109-120

Jacobsen, K. (2005). High Resolution Satellite Imaging Systems - an Overview. *PFG*, 487-496

James, L.A., Watson, D.G., & Hansen, W.F. (2007). Using LiDAR data to map gullies and headwater streams under forest canopy: South Carolina, USA. *CATENA, 71*, 132-144

Jansen, M., Judas, M., & Saborowski, J. (Eds.) (2002a). *Spatial Modelling in Forest Ecology and Management*. (pp. 225). Berlin: Springer

Jansen, M., Schmidt, W., Stüber, V., Wachter, H., Naeder, C., Weckesser, M., & Knauft, F.J. (2002b). Modelling of natural woodland communities in the Harz mountains. In: M. Jansen, M. Judas & J. Saborowski (Eds.), *Spatial Modelling in Forest Ecology and Management* (pp. 162-175). Berlin: Springer

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

Janssen, L., Jaarsma, M., & van der Linden, E. (1990). Integrating Topographic Data with Remote Sensing for Land-Cover Classification. *Photogrammetric Engineering & Remote Sensing, 56*, 1503-1506

Jensen, J.R. (2005). *Introductory Digital Image Processing: A Remote Sensing Perspective*. (pp. 526): Pearson Education

Jobin, B., Labrecque, S., Grenier, M., & Falardeau, G. (2008). Object-Based Classification as an Alternative Approach to the Traditional Pixel-Based Classification to Identify Potential Habitat of the Grasshopper Sparrow. *Environmental Management, 41*, 20-31

Kanellopoulos, I., Wilkinson, G.G., Roli, F., & Austin, J. (1997). *Neuro-computation in Remote Sensing Data Analysis*. (pp. 284). Berlin: Springer

Kartikeyan, B., Majumder, K.L., & Dasgupta, A.R. (1995). An Expert System for Land Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing, 33*, 58-66

Kenneweg, H., Schardt, M., & Sagischewski, H. (1996). *Beobachtung von Waldschäden im Gesamtharz mit Methoden der Fernerkundung*. (pp. 229). Berlin: TU Berlin

Kim, H., & Swain, P.H. (1995). Evidential Reasoning Approach to Multisource-Data Classification in Remote Sensing. *IEEE Transactions on Systems, Man, and Cybernetics, 25*, 1257-1265

Kim, M., Madden, M., & Warner, T. (2008). Estimation of optimal object image size for the segmentation of forest stands with multispectral IKONOS imagery. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 291-308). Berlin: Springer

Kleinschmit, B., & Coenradie, B. (2005a). Entwicklung von Methoden zum Monitoring von Trockenheitsfolgeschäden unter Nutzung von hochauflösenden Fernerkundungsdaten. Berlin: TU-Berlin (pp. 9-10)

Kleinschmit, B., & Coenradie, B. (2005b). Kartierung von Trockenfolgeschäden mit hochauflösenden Satellitenbilddaten. In: E. Seyfert (Ed.), *DGPF-Tagung* (pp. 120-128). Rostock

Kleinschmit, B., Förster, M., Frick, A., & Oehmichen, K. (2007). QuickBird Data – experiences with ordering, quality and pan sharpening. *PFG*, 73-84

Kleinschmit, B., Walentowski, H., Förster, M., Fischer, M., Seitz, R., Kraft, P., Schlutow, A., & Bock, M. (2006). Erfassung von Wald-Lebensraumtypen in FFH-Gebieten - Fernerkundung am Taubenberg und im Angelberger Forst. *LWF Wissen, 51*, 1-39

Klug, H., Langanke, T., & Lang, S. (2003). IDEFIX - Integration einer Indikatorendatenbank für landscape metrics in ArcGIS 8.x. In: J. Strobl, T. Blaschke & G. Griesebner (Eds.), *Angewandte Geographische Informationsverarbeitung - AGIT* (pp. 224-233). Salzburg: Wichmann

Korpela, I.S. (In Press). Mapping of understory lichens with airborne discrete-return LiDAR data. *Remote Sensing of Environment, In Press, Corrected Proof*

Küchler, M., Ecker, K., Feldmeyer-Christe, E., Graf, U., Küchler, H., & Waser, L. (2004). Combining remotely sensed spectral data and digital surface models for fine-scale modelling of mire ecosystems. *Community Ecology, 5*, 55-68

Laben, E.A. (2000). Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening (Gram-Schmidt). US Patent. Eastman Kodak Company

Lang, S., & Blaschke, T. (2007). *Landschaftsanalyse mit GIS*. (pp. 404). Stuttgart: Ulmer

Lang, S., & Langanke, T. (2005). Multiscale GIS tool for site management. *Journal of Nature Conservation, 13*, 185-196

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

Langanke, T., Blaschke, T., & Lang, S. (2004). An Object-Based GIS / Remote Sensing Approach Supporting Monitoring Tasks in European-Wide Nature Conservation. *First Mediterranean Conference on Earth Observation (Remote Sensing)* (pp. 245-252). Belgrade

Langanke, T., Burnett, C., & Lang, S. (2007). Assessing the mire conservation status of a raised bog site in Salzburg using object-based monitoring and structural analysis. *Landscape and Urban Planning, 79*, 160-169

Langar, F. (2008). Ausweisung und Ansprache von Waldlebensraumtypen mit Hilfe von Fernerkundung und GIS. Freiburg: Albert-Ludwigs-Universität Freiburg (p. 135). Dissertation

Li, J., Tang, X., Zhao, W., & Huang, J. (2007). A new framework for identifying differentially expressed genes. *Pattern Recognition, 40*

Lister, N.E. (1998). A Systems Approach to Biodiversity Conservation Planning. *Environmental Monitoring and Assessment, 49*, 123-155

Long, J.N., Dean, T.J., & Roberts, S.D. (2004). Linkages between silviculture and ecology: examination of several important conceptual models. *Forest Ecology Management, 200*, 249-261

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing, 28*, 823-870

Lux, F. (2007). Inspire INSPIRE! Beseelt INSPIRE! In: J. Strobl, T. Blaschke & G. Griesebner (Eds.), *AGIT* (pp. 457-462). Salzburg: Wichmann

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., & Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management, 216*, 41-50

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

Maselli, F., Rodolfi, A., Bottai, L., Romanelli, S., & Conese, C. (2000). Classification of Mediterranean vegetation by TM and ancillary data for the evaluation of fire risk. *International Journal of Remote Sensing, 21*, 3303-3313

McIver, D.K., & Friedl, M.A. (2002). Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment, 81*, 253-261

Meinel, G., Neubert, M., & Reder, J. (2001). The Potential Use of Very High Resolution Satellite Data for Urban Areas - First Experiences with IKONOS Data, their Classification and Application in Urban Planning and Environmental Monitoring. In: C. Jürgens (Ed.), *2nd International Symposium of Remote Sensing in Urban Areas* (pp. 196-205). Regensburg: University Regensburg

Mesev, V. (1998). The use of census data in urban image classification. *Photogrammetric Engineering & Remote Sensing, 64*, 431-438

Miethke, A. (2006). Die Veränderung der Landschaftsstruktur am Beispiel der Waldflächen des Freistaates Sachen. In: B. Kleinschmit & U. Walz (Eds.), *Landschaftsstrukturmaße in der Umweltplanung - Beiträge zum Workshop der IALE-AG Landschaftsstruktur* (pp. 120-140). Berlin: Technische Universität Berlin

Möller, M. (2002). Untersuchung zur Modellierung und Analyse höchstauflösender Scannerdaten. In: T. Blaschke (Ed.), *GIS und Fernerkundung - Neue Sensoren - Innovative Methoden* (pp. 10-23). Heidelberg: Wichmann

Nagendra, H. (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing, 22*, 2377-2400

Narumalani, S., Zhou, Y., & Jelinski, D. (1998). Utilizing geometric attributes of spatial information to improve digital image classification. *Remote Sensing Reviews, 16*, 233-253

Neubert, M., Förster, M., Lang, S., Vohland, K., & Kleinschmit, B. (2008a). A concept for analysing the impact of climate-induced changes on habitat diversity. In: A. Car, G. Griesebner & J. Strobl (Eds.), *GI-Forum* (pp. 215-220). Salzburg: Wichmann

Neubert, M., Herold, H., & Meinel, G. (2008b). Assessing image segmentation quality - concepts, methods and applications. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 215-230). Berlin: Springer

Neukirchen, M. (2005). Was ist erforderlich?Bewertungs- und Überwachungsvorgaben seitens der EU. *Workshop - Kombinierte Methoden aus terrestrischer Erhebung, Modellierung und Fernerkundung bei Natura 2000*. Freising

Nuske, R., Ronneberger, O., Burkhardt, H., & Saborowski, J. (2007). Self-learning Canopy Gap Mapping for Aerial Images using Photogrammetric Height, Color, and Texture Information. In: Cemagref (Ed.), *ForestSAT 2007* (p. 133). Montpellier: Cemagref

Oberdorfer, E. (1992). *Süddeutsche Pflanzengesellschaften (Tabellenband)*. (pp. 580). Jena / Stuttgart / New York

Oehmichen, K. (2007). Satellitengestützte Waldflächenkartierung für die Bundeswaldinventur. Hamburg: Universität Hamburg (p. 105). Dissertation

Openshaw, S. (1984). Ecological fallacies and the analysis of aereal census data. *Environmental Planning, A 16*, 17-31

Openshaw, S., & Openshaw, C. (1997). Fuzzy Logic, Fuzzy Systems and Soft Computing. In: S. Openshaw & C. Openshaw (Eds.), *Artificial Intelligence in Geography* (pp. 268-308). Chichester: Wiley

Pakzad, K. (2001). Wissensbasierte Interpretation von Vegetationsflächen aus multitemporalen Fernerkundungsdaten. Hannover: Universität Hannover

Peddle, D.R. (1995). Knowledge Formulation for Supervised Evidential Classification. *Photogrammetric Engineering & Remote Sensing, 61*, 409-417

Peddle, D.R., & Ferguson, D.T. (2002). Optimisation of multisource data analysis: an example using evidential reasoning for GIS data classification. *Computers & Geociences, 28*, 45-52

Petersen, B., Ellwanger, G., Biewald, G., Hauke, U., Ludwig, G., Pretscher, P., Schröder, E., & Ssymank, A. (Eds.) (2003). *Das europäische Schutzgebietssystem Natura 2000 - Ökologie und Verbreitung der Arten der FFH-Richtline in Deutschland*. (pp. 743). Bonn: BfN

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

Rocchini, D. (2007). Effects of spatial and spectral resolution in estimating ecosystem alpha-diversity by satellite imagery. *Remote Sensing of Environment, 111*, 423-434

Rubin, J. (1967). Optimal Classification into Groups: An Approach for Solving the Taxonomy Problem. *Journal of Theoretical Biology, 15*, 103-144

Sader, S.A., Ahl, D., & Liou, W.S. (1995). Accuracy of Landsat-TM and GIS Rule-Based Methods for Forest Wetland Classification in Maine. *Remote Sensing of Environment, 53*, 133-144

Schiewe, J., & Gähler, M. (2008). Modelling uncertainty in high resolution remotely sensed scenes using a fuzzy logic approach. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 755-768). Berlin: Springer

Schlutow, A., & Hübener, P. (2004). The BERN Model: Bioindication for Ecosystem Regeneration towards Natural Conditions. Berlin: Federal Environmental Agency (pp. 1-50)

Schlutow, A., & Hübener, P. (2005). Ermittlung von aktuellen Regenerierungspotenzialen naturnaher Ökosysteme mittels Bioindikation. *Archiv für Naturschutz und Landschaftsforschung, 44*, 27-79

Schulz, R., Mues, V., Jansen, M., Judas, M., & Saborowski, J. (2002). Spatial prediction of climate, soil, and macrofauna. In: M. Jansen, M. Judas & J. Saborowski (Eds.), *Spatial Modelling in Forest Ecology and Management* (pp. 27-118). Berlin: Springer

Schulze, E.D., & Mooney, H.A. (Eds.) (1993). *Biodiversity and Ecosystem Function*. (pp. 525). Berlin: Springer

Seibert, P. (1968). Übersichtskarte der natürlichen Vegetationsgebiete von Bayern 1:500.000 mit Erläuterungen. *Schriftenreihe für Vegetationskunde, 3*

Seitz, R., & Fischer, M. (2005). Praxistest Fernerkundung: Soll-Ist Vergleich. *Workshop - Kombinierte Methoden aus terrestrischer Erhebung, Modellierung und Fernerkundung bei Natura 2000*. Freising: 19.07.2005

Seitz, R., & Kessler, J. (2004). Managementplan zum FFH-Gebiet 7829-301 „Angelberger Forst“. Augsburg: Bayerische Staatsforstverwaltung (pp. 1-82). Report

Sell, E., Grubinger, J., & Eisl, M.M. (2004). Earth Observation for Natura 2000+ - Eine Bottom-up-Studie - Ergebnisse aus Deutschland und Österreich. In: J. Strobl, T. Blaschke & G. Griesebner (Eds.), *AGIT* (pp. 626-631). Salzburg: Wichmann

Siddiqui, Y. (2003). The modified IHS method for fusing satellite imagery. *ASPRS 2003 Annual Conference*. Anchorage, Alaska

Spooner, J.D. (1991). Automated Urban Change Detection Using Scanned Cartographic and Satellite Image Data. *ACSM-ASPRS Fall Convention* (pp. B118-B126). Atlanta, Geogia: ASPRS

Stohlgren, T., Coughenour, M., Chong, G., Binkley, D., Kalkhan, M., Schell, L., Buckley, D., & Berry, J. (1997). Landscape analysis of plant diversity. *Landscape Ecology, 12*, 155-170

Stolz, R. (1998). Die Verwendung der Fuzzy Logic Theorie zur wissensbasierten Klassifikation von Fernerkundungsdaten. *Münchner Geographische Abhandlungen, B 26*, 1-177

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

Storey, J.D. (2002). A direct approach to false discovery rates. *J. Roy. Stat. Soc., B*, 479-498

Stow, D., Coulter, L., Kaiser, J., Hope, A., Service, D., Schutte, K., & Walters, A. (2003). Irrigated Vegetating Assessment for Urban Environments. *Photogrammetric Engineering & Remote Sensing, 69*, 381-390

Strahler, A.H. (1980). The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data. *Remote Sensing of Environment, 10*, 135.163

Su, Y., Huang, P.S., Lin, C.F., Lin, C.F., & Tu, T.M. (2004). Approach to maximize increased details and minimize color distortion for IKONOS and QuickBird image fusion. *Optical Engineering, 43*, 3029-3037

Sundseth, K. (2005). *Natura 2000 in the Continental region*. (pp. 12). Luxembourg: European Commission,

Terhalle, U. (2005). Sensor-Merge in ERDAS IMAGINE "unplugged". *4. Leica Geosystems Geospatial Imaging User Group Meeting*. Fürstenfeldbruck: Geosystems

Tiede, D., Lang, S., & Hoffmann, C. (2008). Domain-specific class modelling for one-level representation of single trees. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 133-151). Berlin: Springer

Tiehl, C., Tiehl, C., Riedel, T., & Schmulius, C. (2008). Object-based classification of SAR data for the delineation of forest cover maps and the detection of deforestation - A viable procedure and its applications in GSE Forest Monitoring. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 327-343). Berlin: Springer

Tilli, T. (1993). *Mustererkennung mit Fuzzy-Logik*. (pp. 336). München: Franzis

Tso, B., & Mather, P.M. (2001). *Classification Methods for Remotely Sensed Data*. New York: Taylor and Francis

Tu, T.M., Lee, Y.C., Chang, C.P., & Huang, P.S. (2005). Adjustable intensity-hue-saturation and Brovey transform fusion technique for IKONOS/QuickBird imagery. *Optical Engineering, 44*

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote Sensing for biodiversity science and conservation. *Trends in Ecology & Evolution, 18*, 306-314

Tusher, V.G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *PNAS, 98*, 5116-5121

Tzotsos, A., & Argialas, D. (2008). Support Vector Machine classification for Object-Based Image Analysis. In: T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 663-678). Berlin: Springer

UN (2000). Cartagena Protocol on Biosafety to the Convention on Biological Diversity - Text and Annexes. Montreal: United Nations (pp. 1-30). Report

UNEP (1992). Convention on Biological Diversity. Montreal: United Nations Environmental Programme. Report

UNEP World Conservation Monitoring Centre (2000). European Forests and Protected Areas: Gap Analysis. Cambridge: UNEP World Conservation Monitoring Centre (pp. 1-27). Report

Velazquez, J., Förster, M., & Kleinschmit, B. (2008). GIS and remote sensing for Natura 2000 monitoring in Mediterranean biogeographic region. In: A. Car, G. Griesebner & J. Strobl (Eds.), *GI Forum* (pp. 268-275). Salzburg: Wichmann

Venables, W.N., & Smith, D.M. (2006). An Introduction to R (pp. 1-99). Report

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

Volpe, F. (2003). Geometrical Processing of QuickBird High Resolution Satellite Data. *Workshop on High Resolution Mapping from Space 2003*. Hannover: ISPRS/EARSeL

Vrabel, J. (1996). Multispectral Imagery Band Sharpening Study. *Photogrammetric Engineering & Remote Sensing, 62*, 1075-1083

Walentowski, H. (2001). Waldland Bayern - ein Buchenmeer? *LWF-Aktuell, 30*, 1-3

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

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

Wang, L., & Howarth, P. (1994). Multisource spatial data integration: problems and some solutions. *Canadian Journal of Remote Sensing, 20*, 360-367

Watanachaturaporn, P., Varshney, P.K., & Arora, M.K. (2005). Multisource Fusion for Land Cover Classification using Support Vector Machines. In: IEEE (Ed.), *7th International Conference on Information Fusion (FUSION)* (pp. 614-621): IEEE

Weiers, S., Bock, M., Wissen, M., & Rossner, G. (2004). Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS methods. *Landscape and Urban Planning, 67*, 43-65

Whittaker, R., Araujo, M., Jepson, P., Ladle, R., Watson, J., & Willis, K. (2005). Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, 3-23

Wilkinson, G.G. (2003). Are Remotely Sensed Image Classification Techniques Improving? Results of a Long Term Trend Analysis. In: IEEE (Ed.), *Workshop on Advances in Techniques for Analysis of Remotely Sensed Data* (pp. 30-34): IEEE

Wulder, M., Han, T., White, J., Sweda, T., & Tsuzuki, H. (2007). Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterisation. *Remote Sensing of Environment, 110*, 123-137

Zadeh, L.A. (1978). Fuzzy Sets as a basis for a theory of possibility. *Fuzzy sets and fuzzy systems, 1*, 3-28

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

Zebisch, M., Wechsung, F., & Kenneweg, H. (2004). Landscape response functions for biodiversity - assessing the impact of land-use changes at the county level. *Landscape and Urban Planning, 67*, 157-172

Zhang, Q., Wang, J., Peng, X., Gong, P., & Shi, P. (2002). Urban built-up land change detection with road density and spectral information from multitemporal Landsat TM data. *International Journal of Remote Sensing, 23*

Zhang, Y. (1999). Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS Journal of Photogrammetry & Remote Sensing, 54*, 50-60

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*

Zhang, Y., & Hong, G. (2005). An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. *Information Fusion, 6*, 225-234

Zubrow, A. (2003). Mapping Tension: Remote Sensing and the Production of a Statewide Land Cover Map. *Human Ecology, 31*, 281-307

# Acknowledgement

It was a long way from starting this thesis and finishing it was not possible without help. I would like to thank the colleagues and friends who supported me throughout this time.

First of all I thank Birgit Kleinschmit, which gave me the possibility to work at the Department of Geoinformation Processing for Landscape and Environmental Planning. Birgit, you gave me the freedom to develop my ideas and concepts, but reminded me to keep deadlines as well. Prof. Dr. Hartmut Kenneweg was a real help in letting me share all his experience and giving me in lots of small conversations the impression that he fully understood my concerns and improvements.

I am especially grateful to my contributing co-authors, which shared their ideas and work with me during the intense process of writing manuscripts. Annett Frick supplied valuable information about image quality issues of QuickBird data and supplied information about her methodology as well as results in NATURA 2000 heathland sites for comparing our studies. Helge Walentowski provided contributions about forest ecology, which were important for implementing the rules into classification processes and to understand the natural forest succession. Katja Oehmichen helped with her supply of information about ordering QuickBird data. Michael Bock and Philipp Kraft processed the data and gave additional information about the model techniques of multivariate clustering and fuzzy integration of BERN. Michael Fischer and Rudolf Seitz are thanked for evaluating the modelling results with aerial photographs and their regional knowledge of the test sites.

Thank goes to my colleagues of the Department of Geoinformation Processing for Landscape and Environmental Planning. The were fruitful discussions with our remote sensing experts Stefan Lang, Tobia Lakes, and Bodo Coenradie, but my colleagues Britta, Antje, Lutz, Helge, Isabella, Elke and many others always supported me and gave me time when I needed it to research and write.

While writing a thesis I had the possibility to meet and discuss with many colleagues, which all are somehow contributed to this work. Amongst many others I had helpful discussions with Patrick Hostert, Manuela Hirschmugl, Marco Neubert, Roeland de Kok, Friecke van Coillie, and Lieven Verbeke. Sibylle, Erika, and Sandra are thanked for their friendly and continuing support.

Invaluable is the help from family & friends accompanying Saskia and me during these years. Thank you, Kathrin & Philipp (for endless phone chats), Claudia & Sascha, Annekathrin & Falk, Anja, Uwe, Kerstin & Pfeffi (without you our flat would not be finished until now), Thesi & Stefan, Franki, Frank Glaßer, Frank Junggeburth, and Klaudia & Matthias (for long hours of babysitting).

Saskia – what a way from seeing each other for the first time in the University library! You know me and gave me a lot of strength, confidence and (even more important) practical advises during the time often stressful with lecturing, proposal writing, house building, Matilda caring – simply living.

Matilda, one day you will be able to read this, so I send all best wishes for you into the future. Thank you for showing me the things that are really important in life – which is certainly not writing of scientific publications.

Liebe Eltern – danke für das grenzenlose Vertrauen, dass ihr immer in mich und meine Entscheidungen hattet. Nur durch die Mischung aus gesunder Skepsis und Begeisterungsfähigkeit, die ihr mir mitgegeben habt, konnte ich diese Aufgabe bewältigen.

# Curriculum Vitae and List of Publications

Michael Förster (né Winkler) was born on 11th December 1975 in Burgstädt (Saxony). After completing his secondary school and civil service in 1996, he studied Geoecology at the University of Potsdam including a two semester stay at the University of Southampton, UK. During his studies he specialised in GIS and remote sensing techniques. After graduating he worked as a consultant and GIS-coordinator at an Environmental Consulting and Planning Agency. In 2003 he started his PhD work at the Institute of Landscape Architecture and Environmental Planning (Department of Geoinformation Processing for Landscape and Environmental Planning) of the Berlin Institute of Technology. He was involved in different research project ranging from bioenergy studies to data homogenisation. His main interest is the integration of additional data in remote sensing classification processes. Apart from his PhD work he is lecturer for GIS and remote sensing courses and supervised students in their study projects and Diploma theses.

**List of Publications of the author related to the scientific research of this thesis:**


*Peer-reviewed articles in journals or as book chapters*

**Förster, M.**, & Kleinschmit, B. (submitted). Significance Analysis of Different Types of Ancillary Geoinformation into an Object-Based Classification Process. *International Journal of Remote Sensing*

**Förster, M.**, Frick, A., Walentowski, H., & Kleinschmit, B. (in press). Approaches to utilising Quickbird-Data for the Monitoring of NATURA 2000 habitats. *Community Ecology*

**Förster, M.**, & Kleinschmit, B. (2008). Object-based classification of QuickBird data using ancillary information for the detection of forest types and NATURA 2000 habitats. In T. Blaschke, S. Lang & G. Hay (Eds.), *Object-Based Image Analysis* (pp. 275-290). Berlin: Springer

Kleinschmit, B., **Förster, M.**, Frick, A., & Oehmichen, K. (2007). QuickBird Data – experiences with ordering, quality and pan sharpening. *PFG*, 73-84

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

*Non-reviewed articles*

**Förster, M.**, & Frick, A. (in press). Einsatzmöglichkeiten von geometrisch, spektral und zeitlich sehr hoch aufgelösten Fernerkundungsdaten für das Biodiversitätsmonitoring. *Naturschutz und Biologische Vielfalt*

Kleinschmit, B., Walentowski, H., **Förster, M.**, Fischer, M., Seitz, R., Kraft, P., Schlutow, A., & Bock, M. (2006). Erfassung von Wald-Lebensraumtypen in FFH-Gebieten - Fernerkundung am Taubenberg und im Angelberger Forst. *LWF Wissen, 51*, 1-39

*Contributions to conference proceedings (peer reviewed extended abstracts)*

Neubert, M., **Förster, M.**, Lang, S., Vohland, K., & Kleinschmit, B. (2008). A concept for analysing the impact of climate-induced changes on habitat diversity. In A. Car, G. Griesebner & J. Strobl (Eds.), *GI-Forum* (pp. 215-220). Salzburg: Wichmann

Velazquez, J., **Förster, M.**, & Kleinschmit, B. (2008). GIS and remote sensing for Natura 2000 monitoring in Mediterranean biogeographic region. In A. Car, G. Griesebner & J. Strobl (Eds.), *GI Forum* (pp. 268-275). Salzburg: Wichmann

**Förster, M.**, Velazquez, J., & Kleinschmit, B. (2007). Modelling of NATURA 2000 Habitat Types in Different Biogeographic Regions - Experiences from Spain and Germany. In Cemagref (Ed.) *ForestSat 2007* (pp. 1-5). Montpellier, France: Cemagref

**Förster, M.**, & Kleinschmit, B. (2007). Comparison of Different Integration Techniques of Ancillary Information into an Object-Based Classification Process. In C. Heipke, K. Jacobsen & M. Gerke (Eds.), *ISPRS Workshop on High Resolution Earth Imaging for Geospatial Information* (pp. 1-5). Hannover

**Förster, M.**, & Kleinschmit, B. (2006). Einfluss von Segmentierungsskalen auf Strukturparameter zur wissensbasierten Klassifikation am Beispiel von NATURA 2000 Lebensraumtypen. In B. Kleinschmit & U. Walz (Eds.), *Landschaftsstrukturmaße in der Umweltplanung -Beiträge zum Workshop der IALE-AG Strukturmaße* (pp. 109-119). Berlin: TU Berlin

**Förster, M.**, & Kleinschmit, B. (2006). Integration of Ancillary Information into Object-based Classification for Detection of Forest Structures and Habitats. In S. Lang, T. Blaschke & E. Schöpfer (Eds.), *1st International Conference on Object-based Image Analysis* (pp. 1-6). Salzburg, Austria: ISPRS

**Förster, M.**, & Kleinschmit, B. (2006). Remote Sensing and GIS-Modelling for the Monitoring of NATURA 2000 Habitats. In E. Csaplovics (Ed.), *International Symposium on Geoinformatics in European Nature Protection Regions - NatureProtection:GIS* (pp. 63-74). Dresden: TU Dresden - Eigenverlag

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

**Förster, M.**, Kleinschmit, B., & Walentowski, H. (2005). Nutzung unterschiedlicher FE-Sensoren zum Monitoring von Biodiversität in NATURA 2000 - Waldgebieten des Alpenvorlandes. In J. Strobl, T. Blaschke & G. Griesebner (Eds.), *AGIT - Angewandte Geoinformatik* (pp. 165-170). Salzburg, Österreich: Wichmann

*Oral presentations or poster presentations (abstracts)*

**Förster, M.**, Frick, A., & Kleinschmit, B. (2007). Approaches of Utilising QuickBird Data for the Monitoring of NATURA 2000 habitats. In E. Feldmeyer-Christe (Ed.), *Monitoring the effectiveness of nature conservation. Abstracts. International Conference* (p. 59). Birmesdorf: Swiss Federal Institute for Forest, Snow and Landscape Research WSL

Fischer, M., **Förster, M.**, Kleinschmit, B., Seitz, R., & Walentowski, H. (2005). Comparison of three modelling approaches of potential natural forest habitats for monitoring NATURA 2000 areas in Bavaria. In, *35th Annual Conference of the Ecological Society of Germany, Switzerland and Austria* (p. 335). Regensburg

Kleinschmit, B., **Förster, M.**, Walentowski, H., Seitz, R., & Fischer, M. (2005). Detection of NATURA 2000 habitat types by the use of modelled potential natural vegetation in combination with satellite based classification. In, *Geosystems User Group Meeting 2005*. Fürstenfeldbruck

**Förster, M.**, Kleinschmit, B., Kenneweg, H., Walentowski, H., Seitz, R., & Fischer, M. (2004). NATURA 2000 und Fernerkundung – Vorstellung der Pilotstudie Taubenberg und Angelberger Forst. In, *Workshop N2K-Express*. Freising

# Author's Declaration

I prepared this dissertation without illegal assistance. This work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. This dissertation has not been presented to any other University for examination, neither in Germany nor in another country.

Michael Förster

Berlin, September 2008