DEVELOPMENT OF AN ADVANCED UNCERTAINTY MEASURE FOR CLASSIFIED REMOTELY SENSED SCENES

J. Schiewe, C. Kinkeldey

HafenCity University Hamburg, Lab for Geoinformatics and Geovisualization (g²lab)
Hebebrandstr. 1, 22297 Hamburg, Germany
{jochen.schiewe, christoph.kinkeldey}@hcu-hamburg.de

KEY WORDS: classification, uncertainty, quality assessment, fuzzy borders, transition zones

ABSTRACT:

Classified remotely sensed data serves as the basis for various types of city models. Since the requirements concerning the correctness of these models are rapidly growing, the demands for a significant assurance of their quality increase as well. Standard methods for the a posteriori evaluation of classified data have successfully been applied but they do not fully meet the requirements resulting from recent developments, primarily the higher geometric and thematic accuracies of modern sensor systems. One consequence is that the uncertainties inherent in all kinds of data cannot be ignored anymore – not even those in the so-called *ground truth* data which is used as reference in the quality assessment process. Hence, we propose an integrated approach that considers uncertainties in both the classification and the reference data. The phenomenon of indeterminate boundaries – another effect of more accurate remote sensing data – is treated using a border model based on fuzzy logic.
This paper describes the overall concept (section 1) as well as its key steps, the generation of transition zones including the fuzzification process (section 2) and the derivation of the advanced uncertainty measure (section 3). In section 4 we present an example application of the concept dealing with the evaluation of a classified orthophoto scene.

1. CONCEPTUAL OVERVIEW

Basically, the conventional evaluation procedures for classified remote sensing data (especially in connection with a comparison between classification result and reference) take the reference data to be error-free (*ground truth*). This assumption is certainly acceptable for lower resolution remote sensing data due to a larger probability of the existence of reference data of higher order. But for high resolution data uncertainties in the reference data should not be neglected because of the worse 'relative resolution' between reference and sensor data. Consequently, we propose an integrated evaluation method which considers uncertainties in the reference data as well.
Due to the higher spatial resolution the effect of indeterminate boundaries between certain object classes is increased in terms of the absolute number of pixels of the respective transition zones. Although adapted methods (like fuzzy logic approaches) have been developed for the actual thematic classification, analogous fuzzy logic methods for evaluation purposes are hardly applied.
We propose an integrated fuzzy approach which deals with the issue of indeterminate boundaries by the definition of buffer zones around the object boundaries. This allows for the computation of the *class-specific fuzzy certainty measure* (*CFCM)* which implicates uncertainty information. Figure 1 outlines the resulting overall process whose key procedures, i.e. the generation of transition zones and the derivation of an integrated uncertainty measure, will be treated in detail in the next sections.

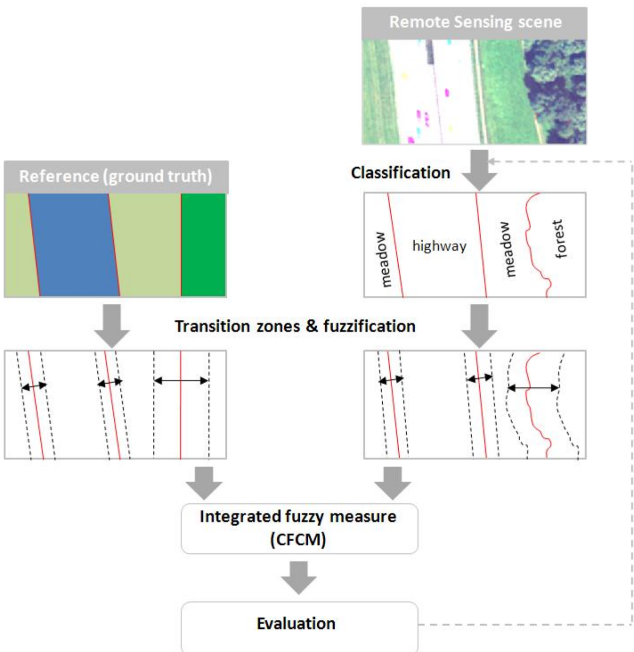

Figure 1. Overall workflow for the derivation of the advanced uncertainty measure CFCM

2. GENERATION OF TRANSITION ZONES

2.1 Idea and previous work

During an interpretation process rules for the allocation of an image object to a given (topographic) class normally assume determinate, discrete boundaries between these objects. However, in high resolution remotely sensed data larger regions (i.e. a larger number of pixels) evolve between classes (e.g. along the boundary of a forest) which make a unique allocation impossible or at least very subjective. Such indeterminate transition zones originate from limited positional accuracies or insufficient semantic definition of objects and their boundaries. Using modern sensors this fuzziness effect is even more severe due to the smaller ground pixel sizes. The spectral variance within regions representing a single topographical object increases and this leads to a larger number of mixed elements (e.g. forest consists of trees, bare ground, etc.).

For modeling indeterminate regions within a classification process the application of *(ε-) bands* (refer to Chrisman, 1992) and the fuzzy logic theory have been proposed. With respect to the latter the concept of varying memberships to a class (from 'no membership at all' to 'perfect membership') along with its application for classification tasks, have been demonstrated by Fisher (2000). Also Wang (1990) warrants the application and proposes the derivation of a fuzzy partition matrix which summarizes the membership values of a feature to every possible class as defined in the object catalogue. Edwards & Lowell (1996) define a membership function for the description of spatial uncertainties. Here for all pairs of objects classes (so-called *twains*) fuzzy widths are introduced based on the mean deviations derived from repeatedly digitizing boundaries from aerial photos.

2.2 Geometric aspects

The *transition zones* serve as a model of the boundary area between two classified geographical objects. Their geometry is constructed depending on the kind of object pair. Basically it is assumed that the transition areas are symmetric, i.e. two adjacent objects share the same transition zone geometry.

In order to create the geometries for these zones the boundaries are buffered on both sides (figure 2). The boundary width depends on the classes of the respective objects and is determined in advance on the basis of semantic aspects for each occurring pair of object classes (see section 2.3). Inside of the transition zone a *fuzzy membership function* is defined perpendicular to the object's boundary. The result is a function that provides a value of 1.0 (full membership to an object class) on the inner boundary of the transition zone and a value of 0.0 (no membership) on the outer boundary.

Currently, the concept is limited to linear fuzzy functions but future research will consider non-linear membership functions as well. It appears reasonable to apply different kinds of fuzzy functions in order to consider the shape of different transitions between certain objects.

In the case of an object with multiple neighbouring objects, its boundary is being split up and the partial boundaries are buffered separately. After this, all single zones along each boundary are aggregated to an overall zone using a standard union operation (see the example application in chapter 4).

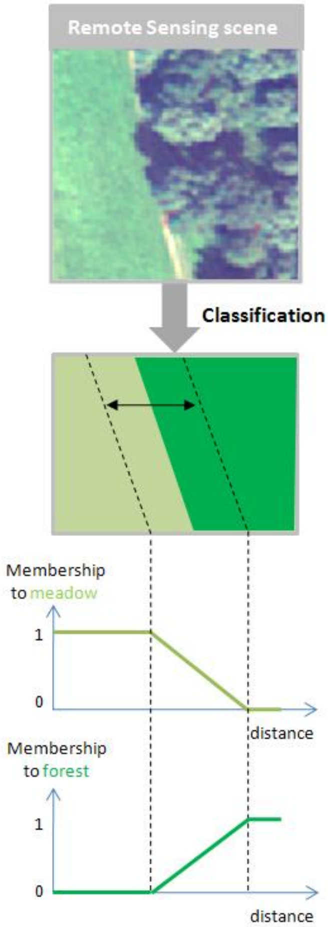

Figure 2. Generation of transition zones and fuzzification per class

2.3 Semantic aspects

As already pointed out, the thematic membership of the neighbouring objects have a significant influence on the fuzziness of the boundary and the width of the transition zone (Edwards & Lowell, 1996). A generally accepted specification for the width of border regions in terms of absolute numbers is virtually not possible due to a couple of factors like different ground sampling distances, seasonal influences or up-to-dateness. Alternatively, a qualitative approach for the definition of the width of boundary regions can be transferred from ecology. Jalas (1955) and Sukopp (1972) developed a system that describes the intensity of human influences, distinguishing between 'natural habitats' and 'artificial habitats'. Based on that, regions under consideration are classified on a scale from *ahemerob* (natural) to *polyhemerob* (artificial) – which is a measure for the influence of mankind on landscape. First approaches for the combination of the *degree of naturalness* with remote sensing data have been developed in the Austrian SINUS-project (refer to Wrbka et al. 2003). In this project a statistical correlation has been determined between landscape metrices calculated from remote sensing data and the degree of naturalness.

Based on these results and our own test series we have started with a qualitative definition of boundary region widths. The result is a ranking that includes typical object class pairs and their boundary width in relation to each other (on a scale from '0' to '+++++').

This guideline helps us to assign absolute, quantitative zone widths to the object classes of a specific classification dataset. This must be done with respect to the characteristics of the classification data and the used object class catalogue (minimum mapping unit, quality of data sources etc.). Obviously, expert knowledge is needed for this initial step of the process so our aim is to build up a collection of parameters for common landcover / landuse catalogues (e.g. CORINE Landcover) being transferable to different datasets of the same classification product.

3. DERIVATION OF THE CHARACTERISTIC VALUE

3.1 Definition

The class memberships of a pixel or a region within the reference (μ_{REF}) and in the classification result (μ_{CLASS}) as described in the previous section form the basis for the derivation of the *class-specific fuzzy certainty measure (CFCM)*. Outside of the transition zones the class memberships can either be 1.0 (if the point lies within an object of the class) or 0.0 (if it does not). Inside of the transition zones the class memberships are determined by the fuzzy function leading to a value within the interval [0.0, 1.0].

Computing the difference between the class memberships μ_{REF} and μ_{CLASS} yields the overall certainty measure for each class as follows:

$$CFCM(c) = 1 - \frac{1}{n}\sum_{i=1}^{n} | \mu_{i,REF}(c) - \mu_{i,CLASS}(c) |$$

$$\forall i \mid \mu_{i,REF} > 0 \vee \mu_{i,CLASS} > 0$$

$\mu_{REF}(c)$: class membership value of a pixel / area for class c in reference data

$\mu_{CLASS}(c)$: class membership value of a pixel / area for class c in classification data

n: number of pixels / areas under examination

A high resulting CFCM value for a certain class expresses a high accordance between the objects of the class in the reference dataset and the classification. The CFCM can be computed for specific topographical classes either in an entire scene or in arbitrary areas such as single objects.

Figures 3a and 3b demonstrate the effect of the advanced uncertainty measure CFCM (in a qualitative manner) in comparison to a conventional approach. Concerning the latter the evaluation results in a binary decision (in the presented case a classification as 'false'). In contrast to this, the transition zone definition (figure 3b) considers the effect of indeterminate boundaries – the degree of coincidence is expressed by a CFCM value being larger than 0 (but still clearly below 1.0).

The benefit from the advanced method is that the statement about the (un-)certainty of the classification allows for a much more differentiated conclusion about the quality of the dataset. One intended effect is that geometric differences in the object boundaries resulting from class definition characteristics and interpretation tolerance are weighted less compared to 'real' differences (e.g. divergent allocation of an object to a class).

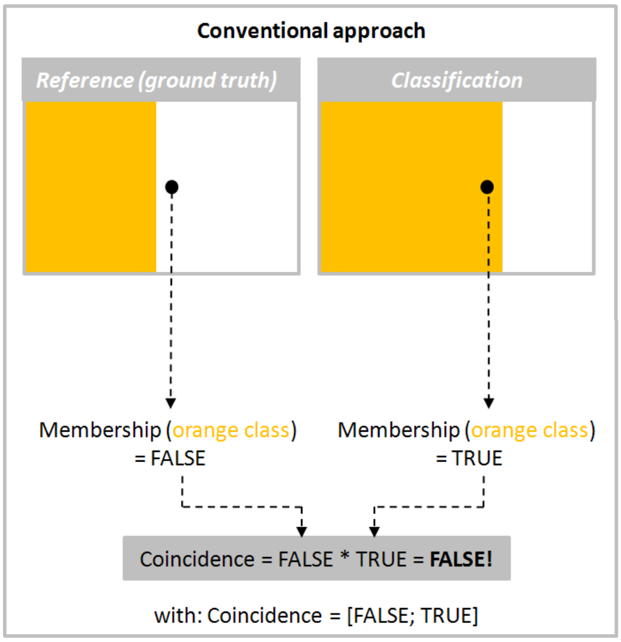

Figure 3a. Comparison of ground truth and classification. Without fuzzy borders the coincidence at the marked position is characterized as 'false'.

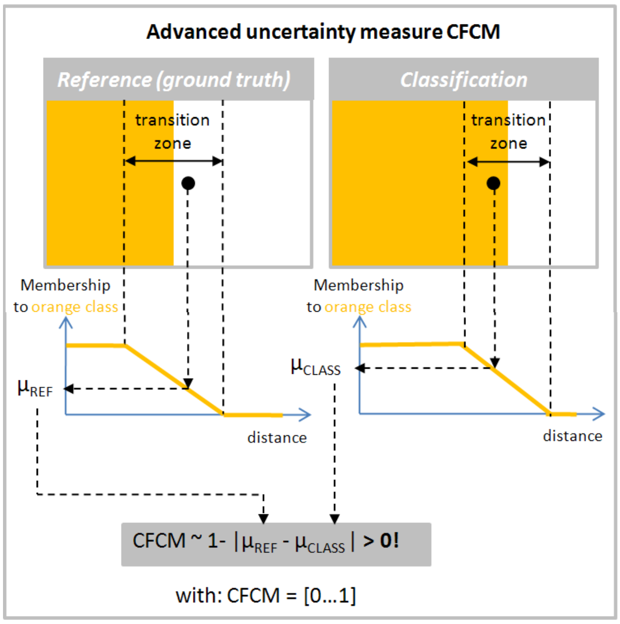

Figure 3b. The advanced uncertainty measure CFCM uses fuzzy borders. The coincidence at the marked position is characterized as 'neither completely false nor true' – a value between 0 and 1.

4. EXAMPLE OF USE

Apart from a range of synthetic test scenarios we apply our concept on first datasets based on real high resolution remote sensing scenes. The most important aim is to gather experience with the transition zone model, especially regarding the definition of the fuzzy border geometry (zone widths) on the basis of the existing semantic information. One of the examples is presented in this chapter.

We consider two classification datasets of a research area in the region around Hamburg, Germany – a residential area next to a forest and an agricultural area. From a student project we have obtained multiple manual classifications of this area on the basis of orthophoto material (pixel-resolution of 10 cm). We have selected two of them in order to demonstrate the process by means of the forest object contained in this dataset (see figure 5).

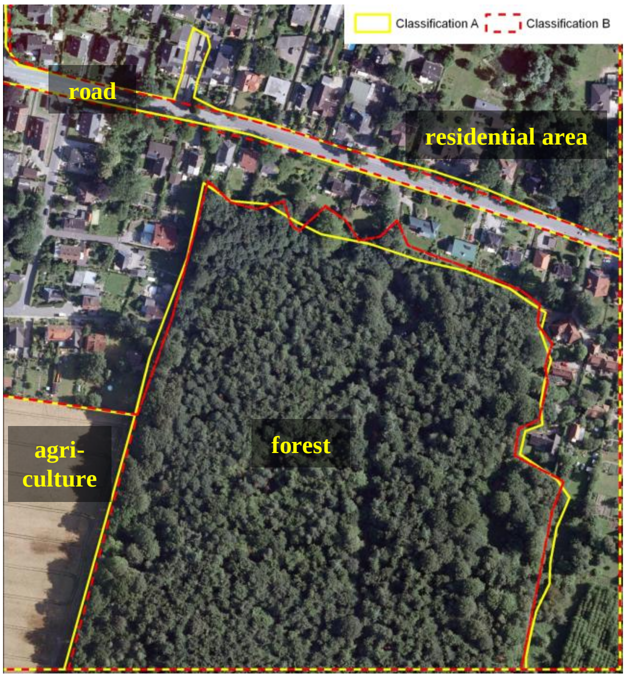

Figure 5. Two different manual classifications based on an orthophoto scene.

The first step is to compile the existing landcover classes and define the degree of uncertainty for the occurring object class pairs. Table 1 shows the assignment of fuzzy zone widths to the object classes whereas a value of 1 m results in a symmetric buffer zone with 2x1 m (1 m to both sides of the border). The choice of the values has been done according to their degree of naturalness (see 2.3) and our experience from multiple classifications. The largest zone width is assigned to the boundary between 'forest' and 'residential area' in contrast to the 'road' class which separates from 'residential area' quite clearly. Figure 6 shows the transition zone of the forest object completed by a detail view of the buffer between 'forest' and 'residential area' in figure 7.

	residential area	**road**	**agriculture**	**forest**
residential area		1 m	4 m	10 m
road			-	-
agriculture				5 m
forest				

Table 1. Object classes and pairwise zone widths.

Figure 6. Definition of the transition zones for the forest object with different zone widths for agriculture (2x5 m) and residential area (2x10 m).

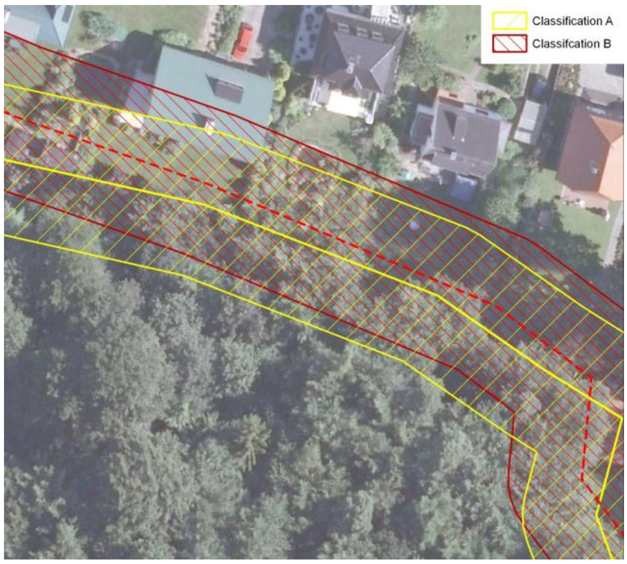

Figure 7. Detail view of transition zones in the boundary area between 'forest' and 'residential area'.

According to our concept the following steps have been performed:

1. Splitting up the boundaries according to adjacency → partial boundaries
2. Buffering of each partial boundary with the respective width according to its object pair
3. Merging the partial buffers of each object to one ring buffer → *transition zone*
4. Application of (here: linear) fuzzy function → fuzzy memberships

After defining the transition zones we are able to determine the membership to a class for each point in the dataset. Figure 8 presents the membership values for class ‘forest’ ranging from 1.0 (inside the object) to 0.0 (outside) including values in between within the transition zones.

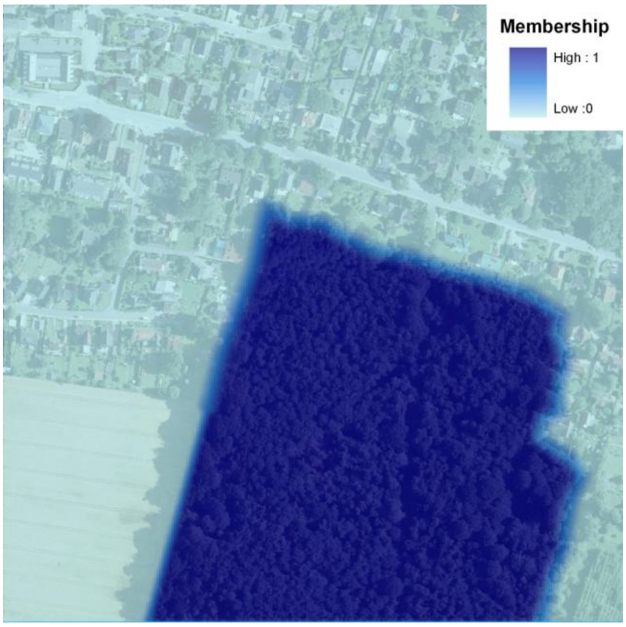

Figure 8. Membership values for ‘forest’ (Classification A).

From the membership values, the quality measure can be obtained:

1. Computation of CFCM measure for each object class (for the whole area)
2. Computation of mean CFCM value for each class

The software implementation performs the whole process automatically so that the manual effort for the computation is limited to the provision of the parameters. The following outputs are created by the software:

- The buffer zones
- The membership values
- The CFCM distribution for each class
- One overall CFCM value for each class

Figure 9 exemplarily shows the distribution of the CFCM values for class ‘forest’. It depicts the certainty of the forest object which turns out to be relatively low at the northern boundary adjacent to the residential area whereas the border area between the forest and the agriculture object is evaluated as more certain.

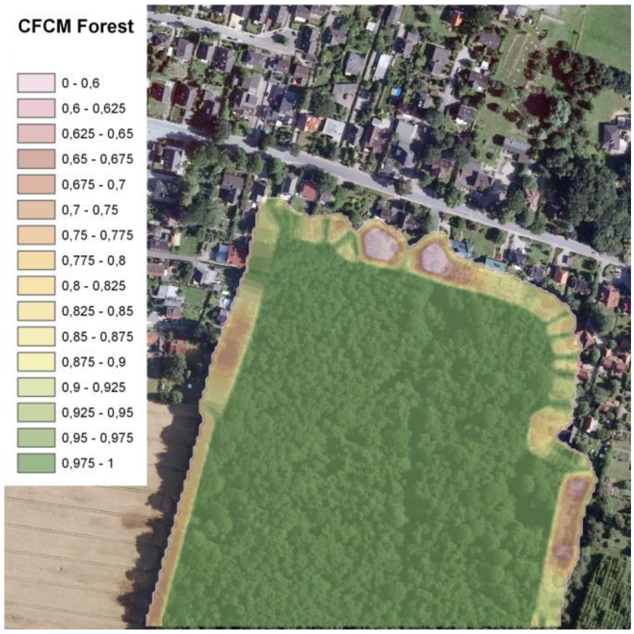

Figure 9. CFCM distribution for class ‘forest’.

Apart from this distribution data that can be used for an uncertainty examination we compute the mean value of all CFCM values for each class which yields in one single value. The outcome for this example is compiled in table 2. The CFCM figures express the agreement between the two classified datasets considering the fuzziness of the contained objects. One can see that the values are close to 100% aside from the road object which shows an accordance below 80% which is due to the different interpretation of the two editors in respect to a dead-end street (see figure 5).

This example points out how the *CLAIM* concept supports the assessment of classification quality with the help of an uncertainty model. A number of applications will follow in the near future to support the further development of the concept.

	residential area	**road**	**agriculture**	**forest**
CFCM	96,09%	78,37%	98,45%	96,77%

Table 2. Mean CFCM value for object classes

5. STATUS AND OUTLOOK

The presented concept of modeling indeterminate transition zones in both the reference and the classified data allows for a well founded description of uncertainty in classified remotely sensed data. With that the derivation of fuzzy logic membership values becomes possible which leads to an advanced certainty measure named *class-specific fuzzy certainty measure CFCM*, giving the desired quantitative, integrated evaluation of the classification accuracy.

The discussed model has already been implemented as Java software that allows for automated test series based on synthetic and real data. This and the integration of additional expert knowledge will help to further develop and quantify the needed control parameters (e.g. for determining the width and the symmetry properties of transition zones). The choice of zone widths plays an important role; that is why the definition of parameters sets for standard landcover / landuse catalogues is intended.

The assignment of the zone widths is currently done under consideration of class-specific attributes. The extension of this concept by the provision for object-specific criteria (size, shape etc.) seems promising and will be one of our next steps of further development leading to an object-specific certainty measure.

Connected to this we will continue with the successive expansion of our *CLAIM* software library that will be available under open source license in order to gain as much user feedback as possible.

6. ACKNOWLEDGEMENTS

The presented work is part of the project *CLassification Assessment using an Integrated Method (CLAIM)* which is currently funded by the German Research Foundation (DFG).

Many thanks to Anne Lübcke for the illustrations in chapter 4.

7. REFERENCES

Chrisman, N.R., 1982. A Theory of Cartographic Error and Its Measurement in Digital Databases. *AUTOCARTO*, (5), pp. 159-168.

Edwards, G. & Lowell, K.E., 1996. Modeling Uncertainty in Photointerpreted Boundaries. *Photogrammetric Engineering & Remote Sensing*. 62(4), pp. 337-391.

Fisher, P., 2000. Sorites paradox and vague geographies. *Fuzzy Sets and Systems*, (113), pp. 7-18.

Jalas, J., 1955. Hemerobe und hemerochore Pflanzenarten. Ein terminologischer Reformversuch. *Acta Soc. Flora Fauna Fennicia*, (72), pp. 1-15.

Schiewe, J., Ehlers, M., Kinkeldey, C. & Tomowski, D., 2009. Modellierung von Unsicherheiten in klassifizierten, räumlich hoch aufgelösten Fernerkundungsszenen. *Tagungsband der Jahrestagung der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF)*, Jena.

Sukopp, H., 1972. Wandel von Flora und Vegetation in Mitteleuropa unter dem Einfluß des Menschen. *Berichte über Landwirtschaft*, 50, pp. 112-139.

Wang, F., 1990. Improving Remote Sensing Image Analysis through Fuzzy Information Representation. *Photogrammetric Engineering & Remote Sensing*. 56(8), pp. 1163-1169.

Wrbka, T. et al., 2003. Landschaftsökologische Strukturmerkmale als Indikatoren der Nachhaltigkeit, *final report of "SINUS" research project*, Federal Ministry of Education and Research, Vienna, Austria (CDROM).