Jun 17th, 3:40 PM - 5:20 PM

Using semantic-based spatial reclassification for interoperable data management in Natura 2000 monitoring

Simon Nieland
Technische Universitat Berlin, simon.nieland@tu-berlin.de

Niklas Moran
Technische Universitat Berlin, nilasmoran@mailbox.tu-berlin.de

Birgit Kleinschmit
Technische Universitat Berlin, birgit.kleinschmit@tu-berlin.de

Michael Förster
Technische Universitat Berlin, michael.foerster@tu-berlin.de

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA
Daniel P. Ames, Nigel W. T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Using semantic-based spatial reclassification for interoperable data management in Natura 2000 monitoring

Simon Nieland [a], **Niklas Moran**[a], **Birgit Kleinschmit**[a] **and Michael Förster**[a]

[a] *Department of Landscape Architecture and Environmental Planning, Geoinformation in Environmental Planning Lab, Technische Universität Berlin, Germany (simon.nieland@tu-berlin.de, niklasmoran@mailbox.tu-berlin.de, birgit.kleinschmit@tu-berlin.de, michael.foerster@tu-berlin.de)*

Abstract: In various fields of spatial research, semantic heterogeneity remains an unsolved problem in terms of data comparability. Interoperability of biodiversity monitoring information is especially difficult because of its thematic complexity and the high variability of acquisition methods and national differences in nomenclatures. Each EU member state must compile comprehensive information on protected areas every six years to comply with the reporting obligations of the habitats directive. Since data collection methods and interpretation manuals broadly vary by member state and manual delineation of protected areas can never be neutral, there is a need for automatized, objective methodologies for the generation of comparable datasets. Comparable datasets derived objectively would further support decision making on a European level. Ontology-based applications offer vast opportunities in data management regarding the interoperability of this kind of information. Basis of this study are two datasets of protected heathlands in Germany and Belgium which are derived from remote sensing and semantically formalized in an OWL2 ontology. The proposed methodology uses semantic relations of the two datasets, which are (semi-) automatically derived from remote sensing imagery, to generate objective and comparable information about the status of protected areas by utilizing kernel-based spatial reclassification (methods/algorithms/techniques). The method therefore suggests a generalization approach, which is able to generate delineation of protected areas of the Natura 2000 network in an entirely automatized procedure. Furthermore, it is able to transfer generalization rules between areas surveyed with varying acquisition methods in different countries by taking into account automated inference of the underlying semantics.

Keywords: Spatial reclassification, Ontologies, Generalization, Nature conservation

1 INTRODUCTION

Comparability of environmental management data is a crucial task in Europe's nature conservation policy. Regulations, like the Habitats Directive (Council Directive) 92/43/EEC [1992], encourages member states of the European Union to establish a consistent and comprehensive basis for biodiversity monitoring and nature conservation activities. This information is collected in the so-called Natura 2000 network, which has to be actualized every six years including reporting from each member state. Due to the federal structure of the European Union and the differences in data delivery approaches of the various nature conservation authorities there is a high demand for innovative technical solutions to realize a comparable and comprehensive monitoring program. Since there are already various methods of deriving nature conservation data (semi-) automatically [Thoonen et al., 2010; Bock et al., 2005; Frick and Weyer, 2005; Vanden Borre et al., 2011; Schuster et al., 2011], it is necessary to generate applications that are able to use the produced information to generate interoperable and therefore more valuable outcomes. Decision makers on an international level rely on the comparability of this

kind of information.
Therefore this work proposes a spatial reclassification approach, which is able to extend existing generalization methods [Thoonen et al., 2010; van der Kwast et al., 2011] by using semantic relations and inference to generate comparability of the outcomes of different regions with regard to its content. This procedure is independent from classification approaches and sensors and can therefore be used for data of multiple input sources in order to generate interoperable datasets.

2 METHOD

This section gives an overview on the developed methodology of generalizing remote sensing classification results to Natura 2000 habitat patches. It furthermore highlights the possibility of developing an application, which is able to interact with an OWL2 ontology to produce fully interoperable results. By taking advantage of the underlying semantics, the application is able to use the logic and relations of the given class descriptions to generate comparable Natura 2000 habitats throughout different regions and classification approaches.

2.1 Formalization of remote sensing classification outputs in owl2

The ontological backbone of this work is a hybrid ontology model, which includes the formalization of remote sensing classification outputs of exemplary heathland sites in Belgium and Germany [Nieland et al., submitted]. This ontology has been developed manually and contains an upper level ontology,

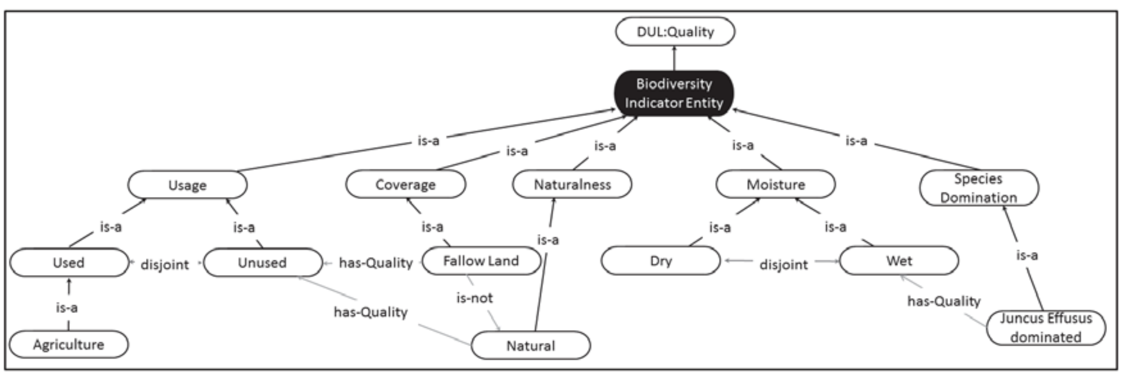

Figure 1. Ontology fragment of the shared vocabulary representing biodiversity indicator entities ([Nieland et al., submitted]

which includes possible attributes of heathland and grassland habitats and their logical relations and several "local" ontologies, which formalize classification outputs of the observed test sites. Additionally, this OWL2/xml ontology includes implicit concept constructors such as relations between classes, (e.g. disjointness), cardinality, equality and characteristics of properties (e.g. symmetry) or richer typing of properties. Figure 1 illustrates an ontology fragment of the upper level ontology. It shows a biodiversity indicator entity with selected, associated concepts. Indicator concepts are illustrated as ellipses, black arrows represent inheritance relationships while grey arrows show concept constructors. The graph structure includes examples of implicit logic relations between indicators. Currently, the shared vocabulary includes 120 concepts for Natura 2000 heathland habitat evaluation.

2.2 Inferring class relations

This work uses Description Logic (DL) for formal knowledge representation stored in an OWL2 ontology (see 2.1). It can be used to automatically infer implicit class relations of the underlying ontology by performing logical reasoning [Donini, 2003]. Therefore, it is possible to achieve matchmaking between classes in different regions by using ontological subsumption, and equality tests in ascending levels of the classification hierarchies ([Nieland et al., submitted]).

2.3 Creation of generalization rules

Since a lot of ecological research and knowledge has lead to the current national and international nomenclatures in the field of biodiversity monitoring and evaluation, the aim of this approach is to use this knowledge to generalize data sources of different origins (observation sensor type, region, methodology) with regard to their content to achieve comparable results. For this study, generalization rules for several heathland habitat types have been created in cooperation with ecological experts [Thoonen et al., 2010] (see table 1). Correct and consistent rules is essential for ensuring high quality results and requires not only profound knowledge of the remote sensing classification procedures but national and international nomenclatures and respective indicators for habitat evaluation as well.

2.4 Generalization algorithm

This section describes a generalization method that is based on a modified spatial reclassification kernel (SPARK) approach [Barnsley and Barr, 1996]. This contextual classification method uses the spatial arrangement and size of pre-classified satellite imagery to define more complex classes. This can be realized by using an adjustable rectangular moving window which analyzes the local characteristics of coverage to assign a generalized class value to its center pixel. In Natura 2000 habitat monitoring, a complex heathland class for example can be a composition of wet or dry heathland, bare sand, ruderal or wet grassland, water bodies, shrubs etc. In contrast to the original SPARK methodology, the described procedure does not contain a pre-classification process, since it uses already existing remote sensing classification results generated by local experts. Another difference to SPARK persists in the fact that the rules for applying a label to a center pixel are already well defined in the national nomenclatures. Therefore these coverage rules, based on expert knowledge, can be used instead of utilizing template kernels that are representative of the habitat classes to be derived (see section 2.1). The proposed method uses a two-step approach to overcome known drawbacks in traditional spatial reclassification kernels [van der Kwast et al., 2011]. [Barnsley and Barr, 1996] define major disadvantages in the traditional SPARK methodology as follows:

- Inability to define window sizes a priori
- Smoothing effects at the edges of the resulting areas
- Rectangular shaped kernels often do not fit to the form of the object of interest

In the first step of the procedure a moving window labels its center pixel according to the predefined rules. The appropriate size can be defined by iterating the process over prevalent window sizes (3x3, 5x5, nxn, 15x15). The percentages of the resulting objects can now be compared to the average coverage of a reference dataset. The kernel size, in which the difference between resulting dataset and reference dataset is at its minimum, can be considered as best fitting. Making sure that the spatial variation of the subject of interest fits to the result and avoiding too big kernel sizes at the same time reduces smoothing effects and leads to more significant results [van der Kwast et al., 2011]. Since conservation areas rarely have strict thematic borders, smoothing effects are not as important as in other fields of object recognition. Since the rules were created for manual fieldwork and adapted to this methodology, areas of multiple labels occur in the result. Mostly transition areas have been affected by this phenomenon. Due to the circumstance, that transitional areas always occur at the borders of habitats, they should not be included in the mapping. Therefore the second processing step uses a simple nearest neighbor interpolation procedure to resolve this problem. Consequently, pixels which have been assigned to multiple labels have been allocated to the nearest unambiguously defined class.

2.5 Using ontological inference for over-regional patch reconstruction

Patch reconstruction with the help of spatial reclassification is a possible solution for the challenge of creating objective Natura 2000 habitat delineation. To additionally address the interoperability in regard to Natura 2000 objects, it is possible to use the relations of components' semantic descriptions from the generalization rules. Therefore it is necessary to know that for example class "Hdc" (heath, dry, calluna dominated) of one region is equal to class "Hzs" (Dry sand heath) of another region. Since we are able to derive equality relations by inferring class relations of the underlying ontology (see 2.2), generalization can be performed in several regions equally, in regard to their content.

2.6 Study sites and data

The input datasets for this study are the results of remote sensing based classification of Natura 2000 heathland areas in the regions Kalmthoutse Heide (Belgium/Flanders) [Thoonen et al., 2013] and Döberitzer Heide (Germany/Brandenburg) [Frick and Weyer, 2005]. The datasets extend over an area of approximately 200 ha in each test site. The focus of this work are selected heathland and grassland habitats (see table 1). The classifications have been performed by using different sensors and classification methodologies [Nieland et al., submitted]. In order to minimize scale effects, the data from Döberitzer Heide (resolution 0.6 m * 0.6 m) was resampled to the resolution of the Kalmthoutse Heide (2.5 m * 2.5 m) using a majority resampling technique.

Table 1. Analyzed Natura 2000 classes. The habitat codes specify natural habitats according to Annex 1 of the European Union council directive on the conservation of natural habitats and wild fauna and flora [(Council Directive) 92/43/EEC, 1992]. The number of rules represent the quantity of rules used to create the delineation of the corresponding habitat type

Habitat code	Habitat name	Number of rules
2310	Dry sand heath with Calluna and Genista	16
2330	Inland dunes with open Corynephorus and Agrostis	16
2310/4030	Dry sand heath with Calluna and Genista or Species-rich Nardus grasslands	16
3100	Standing water	15
4010	Northern Atlantic wet heaths with Erica tetralix	16
4030	Species-rich Nardus grasslands on siliceous substrates in mountain areas	16

2.7 Validation

Comparing the percentage coverage of the remote sensing based input dataset with percentages of the resulting objects will give an indication of the functionality of the generalization technique. In the first step the percentage of coverage will be calculated for each resulting habitat object. In the next step the results will be compared to the percentages of the generalization rule and evaluated as wrong or right. Result of this assessment is a summary on how many objects per habitat class were assigned correctly (see table 2).
Evaluation of the applicability of the methodology is very important since good functional outcomes do

not ensure correctness of the result with respect to nature conservation and biodiversity monitoring. Therefore reference datasets of both test sites have been used to create a pixel based confusion matrix. The matrix is a square array which includes the area of the classified polygons assigned to the area of categories derived from the manual interpretation. The columns represent the reference data, whereas the rows indicate the areas generated in the classification procedure. Therefore the matrix allows conclusions about the accuracy from the producer's perspective (area of a certain correctly classified category divided by the total area of the reference data in the same category) and the user's perspective (area of a certain correctly classified category divided by the total area of the classified data in the same category) [Congalton, 1991] (see table 3).
Since the reference dataset of the destination region (Brandenburg) only includes one complex class ("4030 - Species-rich nardus grasslands on siliceous substrates in mountain areas") which is also present in the origin region (Flanders) an accuracy of delineation in Brandenburg could only be assessed for this class.

3 Experimental results and discussion

The results of the functionality tests are shown in table 2. Two of the result classes have very good results (2330 (84.8%) and 31xx(90%)). In three classes, at least the majority of polygons are correctly classified (4010(53.8%), 2310(60.7%), 6230(75%)). The poor outcome of the class 4030 seems to result from the fact that there is a high amount of very small resulting objects in this class, which are to small to include the respective class percentages. Possible reasons for this are wrong or inconsistent rules or a lack of functionality of the generalization/interpolation procedure. Even with field-based mapping the apperance of Nardus grassland is difficult to delineate, occur often in very small patches and in degraded conservation status.

Table 2. Assessment of the functionality of the generalization algorithm (test site Flanders). The numbers represent the total (percent) amount of resulting habitat patches, which are corresponding (correct) or not corresponding (incorrect) to the defined rules

Habitat code	Habitat polygons correct classified (% of habitat polygons total)	Habitat polygons incorrect classified (% of habitat polygons total)	Habitat polygons total
2310	381(60,7)	247(39,3)	628
2330	89(84,8)	16(15,2)	105
2310/4030	286(61,4)	180(38,6)	466
4010	179(53,8)	154(46,2)	333
31xx	36(90)	4(10)	40
4030	101(17,8)	466(82,1)	567
6230	60(75)	20(25)	80

Table 3 shows the results of the accuracy assessment in the test site Kalmthoutse Heide. Since user accuracies for class 2310 and 4010 are relatively high, whereas producer accuracies are low, it seems likely that these classes tend to be underestimated by the algorithm. However, the classes 2330 and 31xx are highly overestimated. Borders of standing water were very well identified but the reference data did not classify these areas as possible habitats. That indicates that either rule or reference data are not correct for class 31xx.

For the transferred rule (4030) from region Kalmthoutse Heide to Döeberitzer Heide the User Accuracy is 97.8% ,whereas the producer accuracy has a value of 25.5%. Therefore it seems, that the semantic transformation does not necessarily downgrade the quality of the generalization.

Figure 2 visualizes the outcomes of both test sites. It is clearly visible that the results of the Döberitzer

Table 3. Evaluation of the applicability of the methodology (Flanders). Numbers represent the area (m^2) of classification outcomes per classified category assigned to the area of reference data in the same category.

	Reference dataset						
	Habitat code	2310	2330	4010	4030	31xx	Sum
Reclassified dataset (to be evaluated)	2310	301798	12312	19062	2145	0	340202
	2330	105225	140544	20999	1094	0	281007
	4010	36084	693	380023	12099	0	433315
	4030	79942	417	105795	16607	0	206117
	31xx	2955	0	45356	0	5935	109644
	Sum	859697	177691	788388	64338	5935	2138283
	User Acc.	88,71	50,01	87,70	8,06	5,72	
	Prod. Acc.	35,10	79,10	48,20	25,81	100	

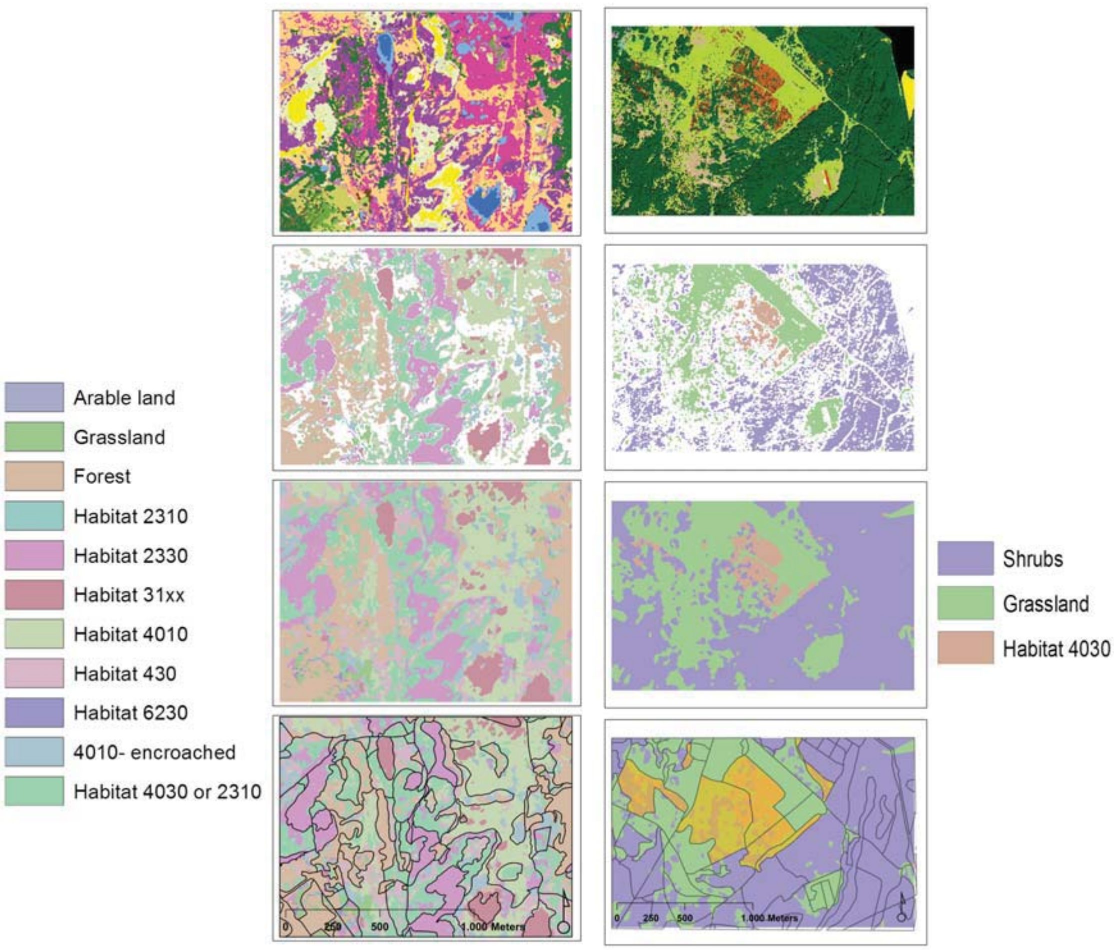

Figure 2. Results of the generalization in Flanders (left) and Brandenburg (right). From top to bottom: 1. The classification outputs. Each color represents one derived class 2. Results of the generalization algorithm (legends shown) 3. Results of the interpolation 4. Interpolation outputs with manual delineations. In Brandenburg the reference for class 4030 is illustrated in yellow

Heide seem to underestimate class 4030, whereas in the Kalmthoutse Heide no clear tendency is distinguishable.

4 Conclusion and outlook

In this work we showed that automated delineation of heathland habitats using spatial reclassification on the basis of remote sensing classification results is technically feasible. Furthermore, generalization rules can be transferred to another region by taking into account the semantics of the respective classification nomenclatures stored in an OWL ontology.
Since the described procedure is dependent on several inputs, the uncertainties in the accuracy assessment are rather high. The quality of the generalization process is always linked to the quality of the remote sensing classification outputs. It also strongly depends on the generalization rules, formulated by ecological experts. Furthermore, uncertainties in the generalization algorithm and respective interpolation can produce lower accuracies. Finally, manual object delineation is always subjective and therefore also includes rather large uncertainties.
Since we have determined that very small resulting objects can lead to poor outcomes an integration of a “minimum mapping unit” in the generalization process is a promising improvement to increase the quality for results of the proposed method. Due to the fact, that each resulting category can have its own ideal kernel size, the kernel sizes can additionally be adapted to categories individually to enhance the quality of the output. Therefore, to reach conclusions about the quality of the described results, the transferability has to be evaluated for a higher number of rules, on larger areas and for more test sites in the future.

References

Barnsley, M. and S. Barr. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric Engineering & Remote Sensing*, 62(8):949–958, 1996.

Bock, M., P. Xofis, J. Mitchley, G. Rossner, and M. Wissen. Object-oriented methods for habitat mapping at multiple scales - Case studies from Northern Germany and Wye Downs, UK. *Journal for Nature Conservation*, 13(2-3):75–89, July 2005.

Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46, July 1991.

(Council Directive) 92/43/EEC, C. The conservation of natural habitats and of wild fauna and flora, 1992.

Donini, F. M. Complexity of Reasoning. In Baader, F., D. Calvanese, D. L. Mcguinness, D. Nardi, and P. Patel-Schneider, editors, *The Description Logic Handbook.Theory, Implementation and Applications*, pages 96–136. Cambridge University Press, 2003.

Frick, A. and G. Weyer. A knowledge-based approach to vegetation monitoring with Quickbird imagery. In *ISPRS Workshop 2005 - High Resolution Earth Imaging for Geospatial Information*, Hannover, 2005.

Nieland, S., B. Kleinschmitt, and M. Förster. Using ontological inference and hierarchical matchmaking to overcome semantic heterogeneity in remote sensing-based biodiversity monitoring. *International Journal of Applied Earth Observation and Geoinformation*, submitted.

Schuster, C., I. Ali, P. Lohmann, A. Frick, M. Förster, and B. Kleinschmit. Towards Detecting Swath Events in TerraSAR-X Time Series to Establish NATURA 2000 Grassland Habitat Swath Management as Monitoring Parameter. *Remote Sensing*, 3(12):1308–1322, June 2011.

Thoonen, G., T. Spanhove, B. Haest, J. V. Borre, and P. Scheunders. Habitat mapping and quality assessment of heathlands using a modified kernel-based reclassification technique. In *Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2707–2710, 2010.

Thoonen, G., T. Spanhove, J. Vanden Borre, and P. Scheunders. Classification of heathland vegetation in a hierarchical contextual framework. *International Journal of Remote Sensing*, 34:96–111, 2013.

van der Kwast, J., T. Van de Voorde, F. Canters, I. Uljee, S. Van Looy, and G. Engelen. Inferring urban land use using the optimised spatial reclassification kernel. *Environmental Modelling & Software*, 26 (11):1279–1288, November 2011.

Vanden Borre, J., D. Paelinckx, C. a. Mücher, L. Kooistra, B. Haest, G. De Blust, and A. M. Schmidt. Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward. *Journal for Nature Conservation*, 19(2):116–125, May 2011.