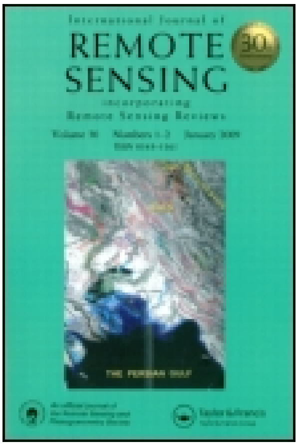


International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Classification of heathland vegetation in a hierarchical contextual framework

G. Thoonen [a] , T. Spanhove [b] , J. Vanden Borre [b] & P. Scheunders [a]
[a] IBBT, Vision Lab, University of Antwerp, B-2610, Wilrijk, Belgium
[b] Biodiversity & Natural Environment, Research Institute for Nature and Forest (INBO), B-1070, Brussels, Belgium

Version of record first published: 10 Sep 2012.

International Journal of Remote Sensing
Vol. 34, No. 1, 10 January 2013, 96–111

Classification of heathland vegetation in a hierarchical contextual framework

G. Thoonen[a]*, T. Spanhove[b], J. Vanden Borre[b], and P. Scheunders[a]

[a]*IBBT, Vision Lab, University of Antwerp, B-2610 Wilrijk, Belgium;* [b]*Biodiversity & Natural Environment, Research Institute for Nature and Forest (INBO), B-1070 Brussels, Belgium*

(*Received 12 October 2010; accepted 19 November 2011*)

Heathlands in Western Europe have shown dramatic declines over the last century and therefore have been given a high conservation priority in the Habitats Directive of the European Union (EU). Accurate surveying and monitoring of heathland habitats is essential for appropriate conservation management, but the large heterogeneity of vegetation types within habitats as well as the occurrence of similar vegetation across habitat types hinders a straightforward, automated mapping based on aerial images. In such a case, a context-dependent classification algorithm is expected to be superior to traditional classification techniques. This article presents a novel approach to map the conservation status of heathland vegetation by using a hierarchical classification scheme that describes the structural dependencies in the field between the basic vegetation and the land-cover types that habitats are composed of. These dependency relationships are included as contextual information in the classification process, using a tree-structured Markov random field (TS-MRF) technique with a tree that reflects the hierarchy of the classification scheme. Results of this approach for a heathland area in Belgium were compared with results from more conventional classification approaches. Validation of the results showed that the structure of the scheme contained important spatial relationships, which were further reinforced by using the contextual classification strategy, especially for the most detailed level of the classification scheme. Accuracy increased and the classification results were more suitable for visual interpretation.

1. Introduction

Atlantic lowland heathlands in Europe have dramatically declined since their maximal extent in the late eighteenth century (De Blust 2007). Not surprisingly, many of the remaining heathland sites in Western Europe have therefore been included in the Natura 2000 network, a Europe-wide ecological network of protected sites that harbour important remnants of rare or endangered habitats and species (European Commission 2005). Following the Habitats Directive of the European Union (EU), member states are committed to adequately conserve these Natura 2000 sites, undertake surveillance of the conservation status of the targeted habitats and species, and report regularly on the implementation of the measures taken under this directive (Vanden Borre et al. 2011). As a result, authorities across the EU are faced with a need for data that cannot be fulfilled by fieldwork alone.

*Corresponding author. Email: guy.thoonen@ua.ac.be

ISSN 0143-1161 print/ISSN 1366-5901 online

http://dx.doi.org/10.1080/01431161.2012.708061
http://www.tandfonline.com

Remote sensing has been recognized as a powerful, innovative tool to aid in Natura 2000 habitat monitoring, but very few readily applicable procedures have been developed so far (Alexandridis et al. 2009; Bock et al. 2005; Langanke et al. 2005). This is probably related to the complex nature of the Natura 2000 habitats: most habitat types should not be thought of as homogeneous vegetation patches of a single or a few dominant species. Instead, they show a high variety of facies at different scale levels. At a large scale, the facies of the same habitat may differ between regions as a result of climatic or soil conditions. But also at a very fine scale, most habitats are in fact intricate mixtures of different land-cover types.

Earlier, we developed a framework for heathland habitat mapping that deals with the complete trajectory, from breaking down habitats into land-cover types, to reconstructing the habitat types and their conservation status from classification results of these land-cover types (Haest et al. 2011). That study covers the transition from habitat definitions and habitat quality aspects to land-cover types, and discusses the translation of these land-cover types back into habitat quality maps. Classification of the land-cover types is a key step in this procedure. This article specifically deals with the classification of land-cover types and, rather than considering only one pixel at time, investigates the use of contextual information, derived from a dedicated classification scheme, in order to increase the accuracy, to facilitate the visual interpretation of the classification result, and to achieve a better habitat characterization in the end.

In order to adequately describe the composition of the various heathland habitat types, the typical components of each habitat type were arranged in a four-level hierarchical land-cover classification scheme. Aside from the information necessary to identify each habitat type, the scheme also comprises relevant aspects of the habitat quality. These include quality indicators specific for each habitat type (e.g. age structure of *Calluna vulgaris*) as well as factors related to more general pressures acting on several habitat types at a time (grass and tree encroachment and occurrence of invasive species).

The choice for the hierarchical scheme opens up new possibilities for classification. Although it is possible to classify on each detail level separately, disregarding the results from the higher levels of the scheme, it is also possible to classify using a hierarchical tree that resembles the structure of the classification scheme. Classifying hierarchically means that any pixel, classified as a certain class on a higher, that is, more general, level of the tree, is restricted to one of its child classes on the lower, i.e. more detailed, levels. The hierarchy itself originates from expert knowledge and represents groups of classes that stem from the same class on a lower level and that are therefore often spatially interwoven on a smaller scale. As a consequence, the scheme has the potential to reinforce the use of contextual information during classification.

However, although contextual information is present in the structure of the classification scheme, even hierarchical classification still processes each pixel separately, disregarding its neighbourhood. Several strategies for including spatial information in the classification process are available (Driesen, Thoonen, and Scheunders 2009; Fauvel et al. 2008). One strategy is to incorporate relationships between neighbouring pixels as prior information in a Bayesian framework. A popular implementation involves modelling these dependencies as Markov random fields (MRFs) (Berthod et al. 1996). In the simplest form, the dependencies can be modelled by a single set of fixed parameters (referred to as ‘flat’ MRF models), describing the transitions between classes. With this approach, however, the parameters do not adapt to the local statistics of the image, nor do they adapt to specific class transitions. Although adaptive methods exist, their computational complexity is very high (Smits and Dellepiane 1999). The tree-structured Markov random field (TS-MRF) technique, introduced by Poggi, Scarpa, and Zerubia (2005), approaches this problem by modelling the hierarchy in the image using a binary tree. As a result, the technique reduces a general MRF to multiple binary MRFs, one for each node in the tree. The reasoning behind this approach is the flexibility of the model parameters. Rather than the common flat approach, a distinct (set of) model parameter(s) can now be used for each node of the tree, fine-tuned to the separation of its two child classes and, at the same time, keeping the problem tractable by reducing the number of parameter sets to the order of the total number of classes. The tree structure is by no means restricted to the binary case, but is selected in the experiments of Poggi, Scarpa, and Zerubia (2005) and Gaetano, Poggi, and Scarpa (2005) because of its relative simplicity, as it reduces the computational burden.

In this work, we aim to combine the strengths of the hierarchical classification approach with those of the TS-MRF technique by forcing the algorithm to follow the classification scheme, rather than selecting a tree at classification time. The resulting tree is not binary, but has a structure that reflects the biological situation in a more suitable way. Moreover, contrary to Poggi, Scarpa, and Zerubia (2005) and Gaetano, Poggi, and Scarpa (2005), the internal nodes of the tree (i.e. any nodes that are not end nodes) are no longer 'virtual' classes obtained by merging their descendant classes, but are 'real' classes in the classification scheme.

In the experiments, the two aforementioned aspects of including contextual information in the classification process are studied: first, the effect of performing hierarchical classification in comparison with straightforward classification on a single level of the classification scheme and, second, the use of MRF to exploit the dependencies between neighbouring pixels. Both effects are studied separately and combined with each other (the TS-MRF approach).

For validation, ground reference data are collected following a randomly stratified approach. In order to compare the results of the described methods with each other, each of the validation samples consists of a small area in which the percentage occurrence of every class is given, rather than just a single class label for an isolated pixel. Consequently, this validation set provides information on the spatial distribution of the classes and therefore facilitates the validation of the contextual classification results.

2. The hierarchical classification scheme

When trying to classify heterogeneous habitats such as heathlands, the classification can be expected to perform better if narrowly defined land-cover classes, rather than complex habitat types, are used as end members for the algorithm, especially if these classes are delimited based on attributes that strongly influence the spectral signature. By doing so, not only the land cover but also the inherent complexity of such habitats can be documented from the remotely sensed data. Therefore, the list of habitats present in the study area was translated into a list of 'land-cover classes', which can be interpreted as spatial units of homogeneous structure and plant species dominance.

In a first step, a provisional list of expected 'land-cover classes' was drawn up for the study area from the list of habitats present. Habitat definitions (European Commission 2007) and ecological knowledge on quality indicators served as inputs for this translation, and the resulting land-cover classes can be interpreted as spatial units of homogeneous structure and plant species dominance. The focus was on quality indicators that relate to vegetation patterns and processes in the habitat, and not to species composition, because the latter is concentrated on typical but usually rare plant species that hardly influence the spectral signature. This list was completed with other land-cover types in the study area that were not covered by the Natura 2000 habitats. Field visits and additional terrain knowledge from local experts ensured that all relevant land-cover classes were considered, while very rare classes were omitted. All classes were subsequently arranged in a hierarchical scheme (see Table 1), based on the (dis)similarity of plant life forms or dominant species present.

The scheme consists of four hierarchical levels. In the first level, the six classes represent broad land-cover classes: heathland, grassland, forest, sand dunes, water, and arable land. These six classes are gradually refined into 23 classes at the fourth level, which are crucial for the assessment of habitat quality. Multivariate analysis (TWINSPAN (Hill 1979)) and Ward's clustering (McCune and Mefford 2006)) of a set of vegetation relevees supported the classification scheme for this study area. The classification scheme can easily be adjusted for application in other heathland areas, by adding site-specific classes or altering threshold values for classes to better correspond with the site's field situation.

Noteworthy is that the classification scheme in Table 1, by itself, already contains 'contextual information', in the sense that it shows which lower level classes occur together in the field or, in other words, which class transitions can be considered very important when doing contextual classification. For instance, 'water' and 'non-water' pixels are easily separated on level 1 of the classification scheme by considering their reflectances and classifying one pixel at a time. These waterbodies, however, depending on their size, usually contain a mixture of shallow and deep waters (banks and centres). Consequently, level 3 classes 'Wov' and 'Wou' occur together and separating them is a much more difficult task, which is expected to benefit from considering contextual information between the two. In the next paragraph, we will introduce a way of incorporating contextual information in the classification process, while, at the same time, using the hierarchy of the classification scheme.

3. Classification methodology

This section first briefly reviews MRF, followed by an overview of the particularities of the TS-MRF approach. To conclude, the approach of this article, which is heavily based on the TS-MRF method, as well as the link to the classification scheme, is described.

3.1. Markov random field

In a Bayesian maximum *a posteriori* (MAP) classification framework, a configuration of class label ω is assigned by maximizing the posterior probability, given by

$$\arg\max_{\omega}\left(P(\omega|x)\right) = \arg\max_{\omega}\left(P(\omega)P(x|\omega)\right), \tag{1}$$

where x is the measured data. Whereas $P(x|\omega)$, the likelihood, can be estimated from the training data, the prior probability $P(\omega)$ may be obtained by modelling the dependencies between neighbouring pixels as MRF (Geman and Geman 1984). The distribution of the prior probability then corresponds to a Gibbs distribution (Li 2001),

$$P(\omega) = \frac{1}{Z}\exp\left[-U(\omega)\right], \tag{2}$$

where $U(\omega)$ is called the energy and Z the normalizing partition function,

$$Z = \sum_{\omega}\exp\left[-U(\omega)\right]. \tag{3}$$

Table 1. Four-level classification scheme.

Level 1		Level 2		Level 3		Level 4	
H	Heathland	Hd	Dry heathland	Hdc	*Calluna*-dominated heathland	Hdcy	*Calluna*-stand of predominantly young age
						Hdca	*Calluna*-stand of predominantly adult age
						Hdcm	*Calluna*-stand of two or three mixed age classes
		Hw	Wet heathland	Hwe	*Erica*-dominated heathland	Hwe-	*Erica*-dominated heathland
		Hg	Grass-encroached heathland	Hgm	*Molinia*-dominated heathland	Hgmd	*Molinia*-stand on dry soil
						Hgmw	*Molinia*-stand on moist soil
G	Grassland	Gt	Temporary grassland	Gt-	Temporary grassland	Gt–	Temporary grassland
		Gp	Permanent grassland	Gpa	Permanent grassland in intensive agricultural use	Gpap	Species-poor permanent agricultural grassland
						Gpar	Species-rich permanent agricultural grassland
				Gpn	Permanent grassland with semi-natural vegetation	Gpnd	Dry semi-natural permanent grassland
				Gpj	*Juncus effusus*-dominated grassland	Gpj-	*Juncus effusus*-dominated grassland
F	Forest	Fc	Coniferous forest	Fcp	Pine (*Pinus* sp.) forest	Fcpc	Corsican pine (*Pinus nigra laricio*)
						Fcps	Scots pine (*Pinus sylvestris*)
		Fd	Deciduous forest	Fdb	Birch (*Betula* sp.) forest	Fdb-	Birch (*Betula pendula*/*pubescens*)
				Fdq	Oak (*Quercus* sp.) forest	Fdqz	Peduculate oak (*Quercus robur*)
S	Sand dune	Sb	Bare sand	Sb-	Bare sand	Sb–	Bare sand
		Sf	Fixated sand dune	Sfg	Sand dune with grasses as important fixators	Sfgm	Sand dune fixated by grasses and mosses
				Sfm	Sand dune with mosses as dominating fixators	Sfmc	Fixated sand dune with predominantly *Campylopus introflexus*
						Sfmp	Fixated sand dune with predominantly *Polytrichum piliferum*
W	Waterbody	Wo	Oligotrophic waterbody	Wov	Shallow, vegetated oligotrophic waterbody	Wov-	Shallow, vegetated oligotrophic waterbody (banks of pools)
				Wou	Unvegetated (deep) oligotrophic water	Wou-	Unvegetated (deep) oligotrophic water (centre of pools)
A	Arable fields	Ac	Arable field with crop	Acm	Arable field – maize	Acm-	Arable field – maize
				Aco	Arable field – other crops	Aco-	Arable field – other crops

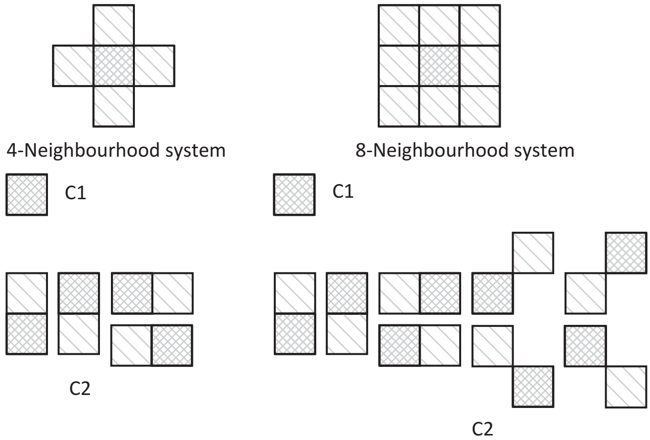


Figure 1. Single-site (C1) and two-site (C2) cliques, related to the central pixel in a 4-neighbourhood and an 8-neighbourhood system in a regular lattice, respectively.

Given a certain neighbourhood relation, for example, a 4-neighbourhood (also referred to as first-order neighbourhood) or an 8-neighbourhood (also referred to as second-order neighbourhood) on a regular lattice, a subset of neighbouring sites can be defined, referred to as cliques. This concept is illustrated in Figure 1. The energy can then be expressed as a sum of potential functions $V_c(\omega)$ over the collection of cliques c,

$$U(\omega) = \sum_{c \in \mathrm{c}} V_c(\omega). \tag{4}$$

The form of the potential functions depends on the choice of the model and of the types of cliques used. A simple and popular model, considering only two-site cliques, is the generalized Ising/Potts model, defined by

$$V_c(\omega) = \begin{cases} \beta, & \text{if } \omega_r \neq \omega_{r'}, \{r, r'\} \in \mathrm{c} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Here, β is a penalty parameter that disfavours heterogeneous cliques by contributing to a higher energy and r, r' are elements of c. However, using just one single penalty parameter does not account for the difference in transitions between different classes. Consequently, a unique β parameter, one for each pair of classes, is desirable, to describe class transitions in a flexible manner. Unfortunately, apart from significantly increasing the computational complexity, this also involves modelling penalty parameters for classes that are not spatially interwoven. A solution for this problem is provided by the introduction of TS-MRF (Poggi, Scarpa, and Zerubia 2005).

3.2. *Tree-structured Markov random field model*

For the details of the TS-MRF technique, the reader is referred to Poggi, Scarpa, and Zerubia (2005). This section provides a brief description of the method in order to highlight the differences with the approach used in our work.

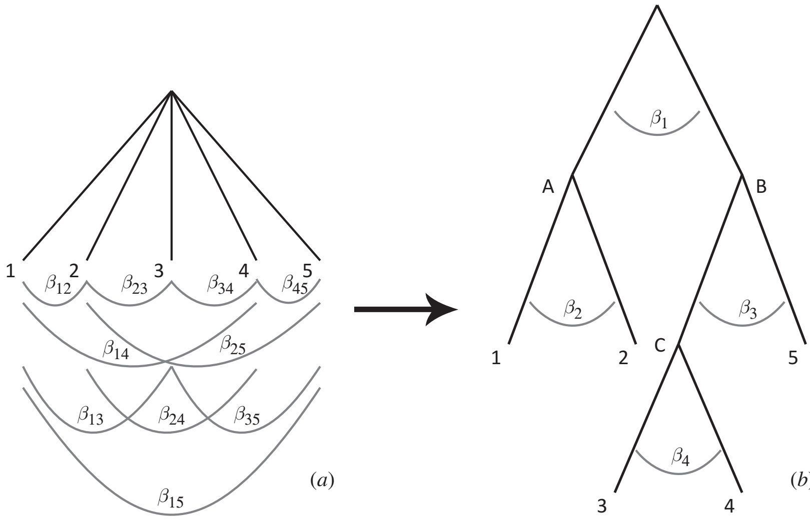


Figure 2. Example of the TS-MRF approach for five classes. (*a*) The non-hierarchical approach with a penalty parameter for each pair of classes. (*b*) The TS-MRF approach with a single-edge penalty parameter for each node.

As illustrated in Figure 2, the TS-MRF approach reduces a complex MRF to multiple binary MRFs, by introducing a binary tree for hierarchical classification. In this binary tree, nodes labelled 1–5 are the end nodes, whereas nodes A, B, and C are the internal nodes. Hierarchical classification implies that any pixels, previously classified as, for example, A, can only go to class 1 or 2. As a consequence, the number of edge penalty parameters reduces significantly, because only one parameter is required for each node of the tree, fine-tuned to the separation of only two classes. More specifically, in the non-hierarchical approach, if there are N classes in total, the total number of edge penalty parameters equals $\frac{1}{2}N(N-1)$. By contrast, in the TS-MRF approach, the number of edge penalty parameters reduces to $N - 1$. In short, the TS-MRF technique offers more flexibility in comparison with the flat MRF approach and, at the same time, reduces the complexity.

The algorithm is largely based on iteratively estimating the penalty parameter β from the previous classification ω^k and constructing a new realization ω^{k+1} based on the estimated parameters. This iterative process is sequentially performed for each group of pixels that are associated with an internal node of the tree (i.e. a node that has child nodes) while moving through the hierarchy, and is illustrated for the pixels associated with one single internal node by the flow chart in Figure 3. Essentially, the pixels associated with a node are those pixels that are labelled as the class associated with this particular node in a previous step. For example, for the binary tree in Figure 2, the pixels associated with the root node are all the pixels in the image. After a preliminary classification (with $\beta_1 = 0$), in which the pixels are labelled as either 'A' or 'B', β_1 and the classification map are iteratively updated until convergence. The process is then repeated for all of the pixels associated with node 'A' (i.e. the pixels classified as 'A' in the previous run of the process) and subsequently for the pixels associated with nodes 'B' and 'C'.

A way to estimate the penalty parameters from the previous classification ω^k is maximum pseudo-likelihood estimation (MPLE Besag 1974). This technique estimates the β parameters that maximize Equation (1), given $\omega = \omega^k$. The prior $P(\omega^k)$ is calculated by Equation (2). Unfortunately, the global partition function Z is intractable. Instead, in MPLE,

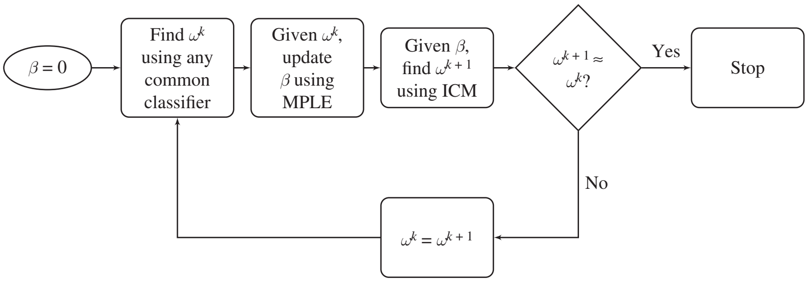


Figure 3. Flow chart of the TS-MRF algorithm for a single internal node of the tree.

a local partition function is calculated for each pixel neighbourhood by summing the energies over all possible values that the central pixel can assume, while keeping the labels of the neighbourhood pixels fixed. The global partition function is then replaced by the product of the local partition functions.

A way to construct a new realization ω^{k+1}, based on the estimated parameters, is iterated conditional modes (ICM Besag 1986); ICM estimates ω^{k+1} that, given the new β parameters, maximizes Equation (1) locally, that is, by assuming that the posterior probability from Equation (1) is proportional to

$$P(\omega|x) \propto \prod_r P(\omega_r)P(x_r|\omega_r). \tag{6}$$

Finding the realization ω^{k+1} for which the posterior probability is maximized reduces to finding ω_r^{k+1} for which Equation (6) is maximized at each pixel location r. This process is iterative, because every new realization has an effect on the prior probabilities given by Equation (2), which, in turn, may or may not change class labels.

3.3. The extended TS-MRF

Rather than just using a random binary tree, determined at classification time, this work implements an extended tree that follows directly from the hierarchy of the classification scheme. This tree is depicted in Figure 4. Moreover, the hierarchical nature of this scheme

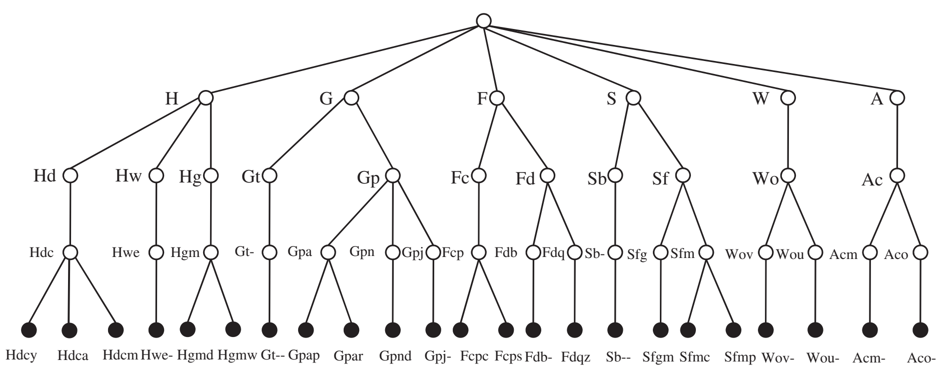


Figure 4. The hierarchical tree that follows from the classification scheme.

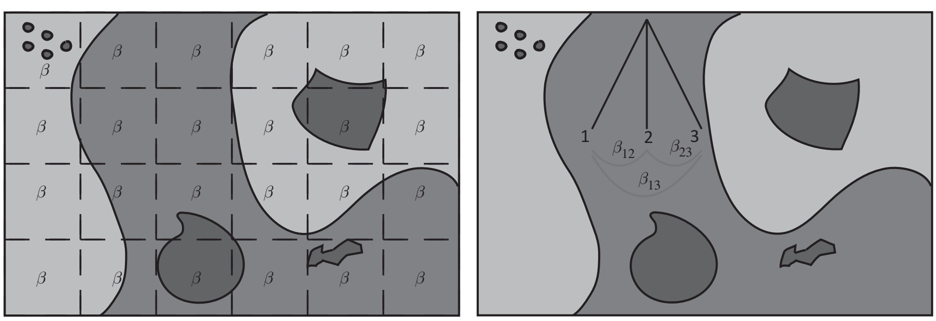
reflects, by itself, already the spatial relationships between vegetation types in the field. The goal of the extended TS-MRF method is to reinforce these spatial relationships even further.

The implication is a trade-off between freedom in tree structure and a larger number of model parameters to be estimated. If one penalty parameter is selected for the description of the transition between each pair of classes, then, for each node, the number of parameters would be

$$\frac{1}{2}N(N-1), \tag{7}$$

where N now equals the number of child nodes instead of the number of classes. For example, for the top node of the tree in Figure 4, there are 6 child nodes and, therefore, 15 penalty parameters are needed. According to this formula, the required number of parameters is 35 for the complete tree. For the original, binary TS-MRF approach, the number of parameters is only 22. However, in the original TS-MRF approach, the model parameters are estimated over the whole image. In general, this approach is too coarse to capture local structural variations.

For the purpose of better modelling these local variations, a raster is used to divide the image into several smaller areas or sub-images. For each of these sub-images, parameters are estimated. Additionally, the number of parameters per node and per sub-image can be reduced to 1 without significant loss of flexibility. Indeed, estimating one single parameter per node for each of the sub-images offers more flexibility than estimating multiple parameters per node over the entire image. For example, while the top node of the tree in Figure 4 has six child nodes, these six classes are very broad land-cover classes. Even for a reasonably large sub-image, just a few of these classes will occur together, rendering most parameters useless. On the other hand, for different locations, textural patterns can be quite different, even for the transitions between the same pair of classes. This is also illustrated in Figure 5 for the case of three classes. Although three classes imply three pairwise parameters according to Equation (7), a single parameter suffices for each of the sub-images because in almost all the sub-images only two classes are co-occurring. On the other hand, depending on the location in the image, different spatial patterns are occurring,



(a) (b)

Figure 5. Illustration of how parameter estimation can be improved by dividing the image in a raster of small sub-images and estimating one single parameter per sub-image in order to incorporate local variations (a) rather than estimating one parameter between each pair of classes for the whole image (b).

which can be modelled more accurately by estimating a unique penalty parameter for each sub-image. A practical measure for the window size is a size proportional to the percentage of the classes, related to a certain node, that are present in the training set or in the pixel-based classification result.

4. Experiments and results

4.1. Experimental set-up

The performance of the hierarchical classification, the flat MRF and the TS-MRF techniques, was tested in the 'Kalmthoutse Heide' and 'Klein schietveld', two protected Natura 2000 sites in the north of Belgium (shown in Figure 6). The central heathland area of approximately 10 km^2 contains a mixture of wet and dry heath, inland sand dunes, and waterbodies. Airborne hyperspectral data were obtained in 2007 with an Airborne Hyperspectral Scanner (AHS) sensor with a ground resolution of approximately 2.5 m. The range of 450–2550 nm is covered by 63 spectral bands. Ground reference data were collected at the time of flight, but were not selected in a random stratified fashion. However, to validate the classification results, random sampling is desirable, because it takes into account the true variation of land-cover classes that are present in the field (Gruijter et al. 2006). This issue was corrected in 2009, when new ground reference data were collected in 586 plots, this time selected in a random stratified way. The study revealed no drastic changes between newly collected ground reference data and the image data from 2007. The few plots that showed changes were removed from the data set. The plots were 10 m in diameter, meaning that each plot contained multiple pixels. This way, minor issues with the georeferencing resulted in relatively smaller errors.

Each plot was labelled according to the classification scheme described in Section 2. Because of the random way of collecting, the plots were often situated on class edges and therefore contained multiple classes. Apart from the obvious workload related to separately labelling all the fine-scaled pixels in each plot at the very high detail of the classification scheme, the assumption that the land-cover classes fit well into a grid of square pixels can

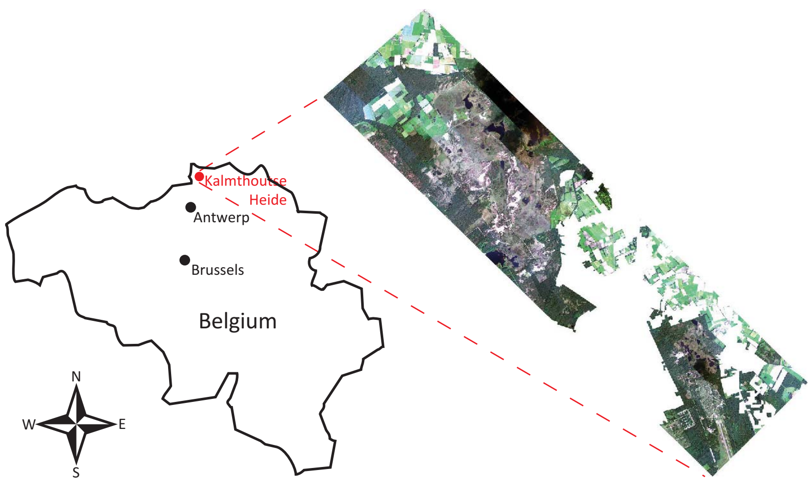


Figure 6. Geographical situation of the study area and true colour AHS mosaic image of the study area used in this study. Urban areas were masked out prior to classification.

lead to a large disagreement in delineation between the ground reference and the classification results, due to the randomness of the pixel raster's location (Lechner et al. 2009). Consequently, a percentage composition of the various land-cover types was estimated for each plot, rather than just a single class label. These percentage compositions provide more information about transition regions between classes than single isolated pixels; hence, they aid in the validation of contextual classification results, because preserving class transitions is one of the key challenges for contextual classifiers such as MRF (Tso and Olsen 2005). The plots were randomly split in a training and a validation set, each containing approximately half of the plots.

To assess the accuracy of the classification result using the percentage composition of the plots in the validation set, a fuzzy error matrix was constructed, based on the minimal percentage membership of classification data and reference data, summed over each plot (Binaghi et al. 1999). If we adopt the notation from Binaghi et al. (1999), the element $\tilde{\mathrm{M}}(m,n)$ on row m and column n of the fuzzy error matrix $\tilde{\mathrm{M}}$ is given by

$$\tilde{\mathrm{M}}(m,n) = \sum_{x \in \mathrm{X}} \min\left(\mu_{\tilde{C}_m}(x), \mu_{\tilde{R}_n}(x)\right), \tag{8}$$

where X is the set of all validation sample plots and $\mu_{\tilde{C}_m}(x)$ and $\mu_{\tilde{R}_n}(x)$ are the percentage memberships of sample plot x for the classification data $\tilde{C}$ in class m and for the reference data in class n, respectively. The overall accuracy (OA) measure is extracted from the fuzzy error matrix in a similar way as with the traditional error matrix, i.e. by taking the sum of all the elements on the diagonal and dividing by the total number of plots:

$$\mathrm{OA} = \frac{\sum_{n=1}^{N} \tilde{\mathrm{M}}(n,n)}{|\mathrm{X}|}, \tag{9}$$

where N is the number of classes and $|\mathrm{X}|$ is the cardinality of the set of validation plots.

All classification results have been produced using the same spectral classifier, namely a weighted linear discriminant analysis (LDA) classifier. Because this article looks at the effects of hierarchical and spatial methods, relative to spectral classification results, other classifiers may be used to produce similar results. However, LDA was selected here for simplicity. The training plots, just like the validation plots, are small areas for which only the percentage occurrence of each class is given. Therefore, training is performed by weighting the plots according to the occurrence of the classes. For instance, in case of the LDA classifier, if a training plot contains 20% of class A and 80% of class B, this plot is taken into account for 20% when calculating the mean vector of class A and for 80% when calculating the mean vector of class B. Because the covariance matrix of an LDA classifier is considered to be the same for all the classes, the percentages do not need to be taken into account while computing it.

Looking at the classification results, two distinct effects are considered: first, the effect of classifying on each level separately versus classifying hierarchically, i.e. the influence of the classification scheme, and second, the effect of using a flat MRF versus TS-MRF approach. For both the flat MRF and the TS-MRF techniques, a second-order neighbourhood and two-site cliques were used. In order to make a fair comparison, parameter estimation for the flat MRF has been performed, similarly to the TS-MRF approach, for each sub-image that results from laying a raster over the image as described in Section 3.3. The four resulting techniques that are compared are:

- a one-pixel-at-a-time classification performed on each level separately (direct);
- a classification performed on each level separately, using a flat MRF model (flat MRF);
- a one-pixel-at-a-time classification, classified hierarchically using the classification scheme (hierarchical); and
- a classification using the TS-MRF technique with the extended tree that resembles the classification scheme, (TS-MRF).

4.2. Results and discussion

Figure 7 illustrates the classification result of Kalmthoutse Heide, for all four levels, using the hierarchical approach. The six general classes that are present are heathland (purple), forest (dark green), grasslands (light green), sand dunes (yellow), waterbodies (blue), and arable fields (red). Figure 8 shows a more detailed view of two small areas (50×50 pixels), classified at level 4 using the four strategies. The hierarchical approach already shows more structure and less noise than the result classified directly on level 4. Furthermore, we can see that both MRF results provide smoother thematic maps, yet they also manage to preserve the transitions between classes with respect to the spectral results. Consider, for instance, the sandy road structures in the top row results of Figure 8. Additionally, the TS-MRF result benefits from the tree structure that reflects the dependencies in the field.

Aside from substantially facilitating visual interpretation, vegetation characterization is also improved, as can be seen from the accuracy assessment. Table 2 shows the OA on levels 2, 3, and 4, respectively, for the classification result using all four strategies. Apart from the OA for all the validation plots, the OA for the homogeneous and heterogeneous plots are shown separately. Homogeneous plots are those plots that contain 100% of the same class, so no transitions occur. Heterogeneous plots are those plots that contain multiple classes, so one or more transitions occur. Consequently, the homogeneous plots provide information on how much smoothing has occurred, whereas the heterogeneous plots, in addition, show how well transitions have been preserved. The table shows that hierarchical classification using the classification scheme improves the OA with respect to direct classification, not only considering the homogeneous plots but also considering the heterogeneous plots, especially as we move further down the hierarchy. Level 4 result shows the largest effect from classifying hierarchically, which is to be expected, because the hierarchy itself is the most extensive for this level.

The MRF result mainly provides smoothing at the higher levels of the hierarchy (levels 2 and 3), increasing the OA for the homogeneous plots, but less so for the heterogeneous plots. For these levels, performances of flat MRF and TS-MRF provide more similar results. Although the differences between the results of the extended TS-MRF and flat MRF strategies are very small at level 2 of the hierarchy, the increase in complexity is equally small. Only when descending down the hierarchy of the tree, the complexity increases substantially and the effects of the extra parameters come into play. Additionally, at levels 2 and 3, patches of ground cover are relatively large and homogeneous. In other words, transitions between patches are more similar, irrespective of the class they belong to. As a consequence, estimating multiple penalty parameters can be considered unnecessary at these levels, because the difference between the parameters will not be very large. At level 4, on the other hand, the benefit of using the hierarchical approach together with MRF becomes much more clear. Although the accuracies for the homogeneous plots of flat MRF and

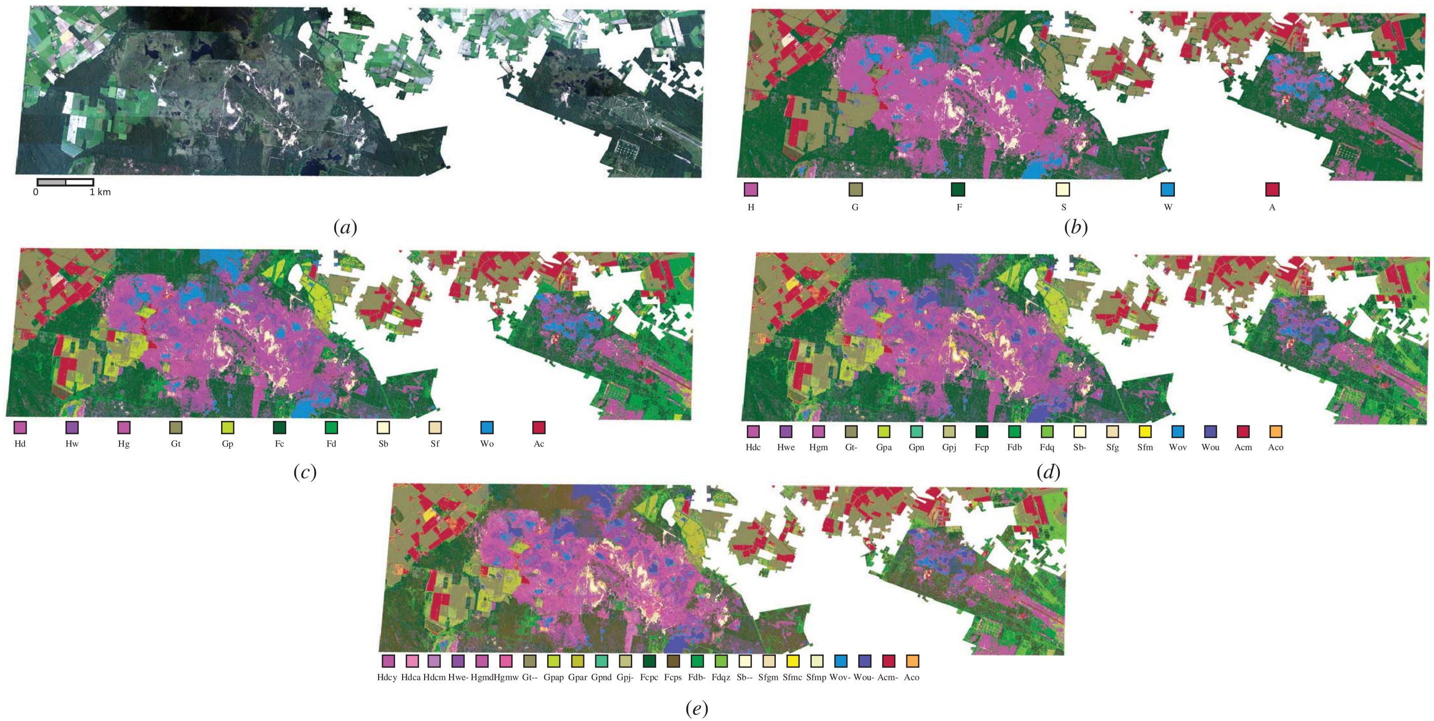
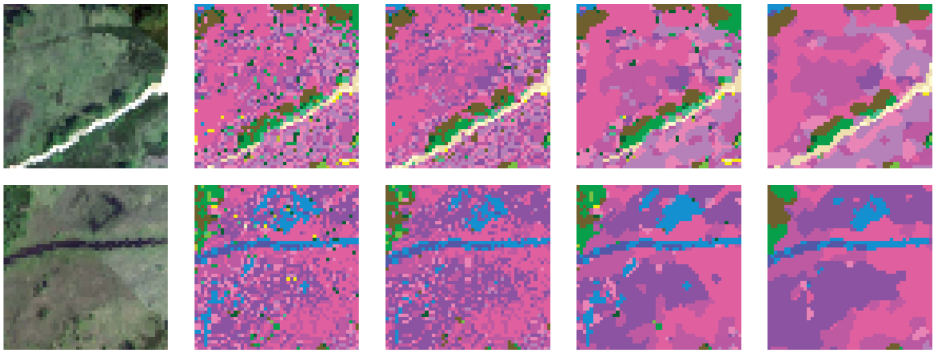


Figure 7. Hierarchical classification result of Kalmthoutse Heide. (*a*) True colour, (*b*) level 1, (*c*) level 2, (*d*) level 3, (*e*) level 4.



(*a*) (*b*) (*c*) (*d*) (*e*)

Figure 8. Details of classification results at level 4 using the various strategies. (*a*) True colour, (*b*) direct, (*c*) hierarchical, (*d*) flat MRF, (*e*) TS-MRF.

Table 2. Overall accuracy for classification results on levels 2, 3, and 4 using all four strategies.

	Direct (%)	Hierarchical (%)	Flat MRF (%)	TS-MRF (%)
Level 2				
All plots	67.08	67.75	68.56	68.62
Homogeneous plots	70.19	71.10	74.92	75.48
Heterogeneous plots	64.31	64.76	62.88	62.49
Level 3				
All plots	61.19	63.53	63.90	64.55
Homogeneous plots	62.60	65.87	69.80	70.56
Heterogeneous plots	60.01	61.56	58.95	59.50
Level 4				
All plots	45.39	48.54	50.78	52.40
Homogeneous plots	46.37	48.77	55.45	55.04
Heterogeneous plots	44.59	48.35	46.90	50.21

TS-MRF results are comparable at this level, TS-MRF performs better on heterogeneous plots. Considering all the plots together, TS-MRF provides the highest overall accuracy on all levels.

5. Conclusion

In this article, we presented a TS-MRF technique based on a tree that reflects the hierarchy of a classification tree built from ecological knowledge. This way, the dependencies laid out in a meaningful hierarchical scheme were incorporated in the classification process, while being further incorporated as contextual information for the classification. When applied to the classification of a large heathland reserve, this method outperformed alternative methods such as a flat MRF approach and a simple hierarchical classification approach. The hierarchical classification approach shows improvements with respect to the per-level classification, as does the extended TS-MRF method, especially on the most detailed levels of the classification scheme. At the same time, the MRF approach increases the ease of visual interpretation of the classification results.

Due to the flexibility of the classification scheme, which can easily be fine-tuned to the local features of any heathland site, we expect that our approach will be well suitable to map and assess the quality of a variety of heathland areas. Moreover, the methodology may even prove valuable for other complex ecosystems such as coastal dunes, scrublands, or tidal areas, but this will require further investigations. Irrespective of its performance in other ecosystems, we have shown that the hierarchical and contextual classification method performs well and can easily be combined with the recently developed habitat classification framework. This will definitely facilitate the mapping, the monitoring, and the assessment of the conservation status of natural heathlands, and may ultimately lead to a better management and conservation of these natural areas.

Acknowledgements

This work was supported by the Belgian Science Policy Office under the framework of the STEREO II programme – project HABISTAT (contract SR/00/103). The partners involved in the project are the Flemish Institute for Technological Research (VITO), Vrije Universiteit Brussel (VUB), Research Institute for Nature and Forest (INBO), and Alterra and University of Antwerp (Vision Lab). The authors thank VITO for the acquisition and pre-processing of the data set and Martin Knotters (Alterra) for advising on the sample design.

References

Alexandridis, T. K., E. Lazaridou, A. Tsirika, and G. C. Zalidis. 2009. "Using Earth Observation to Update a Natura 2000 Habitat Map for a Wetland in Greece." *Journal of Environmental Management* 90: 2243–51.

Berthod, M., Z. Kato, S. Yu, and J. Zerubia. 1996. "Bayesian Image Classification Using Markov Random Fields." *Image and Vision Computing* 14: 285–95.

Besag, J. 1986. "On the Statistical Analysis of Dirty Pictures." *Journal of the Royal Statistical Society* 48: 259–302.

Besag, J. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society* 36: 192–236.

Binaghi, E., P. Brivio, P. Ghezzi, and A. Rampini. 1999. "A Fuzzy Set-Based Accuracy Assessment of Soft Classification." *Pattern Recognition Letters* 20: 935–48.

Bock, M., P. Xofis, J. Mitchley, G. Rossner, and M. Wissen. 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.

De Blust, G. 2007. "Heathland, an Ever Changing Landscape." In *Europes Living Land-Scapes: Essays On Exploring Our Identity in the Countryside*, edited by B. Pedroli, A. Van Doorn, G. De Blust, M. Paracchini, D. Wascher, and F. Bunce, 178–92. Wageningen: Landscape Europe & KNNV Publishing.

Driesen, J., G. Thoonen, and P. Scheunders. 2009. "Spatial Hyperspectral Image Classification by Prior Segmentation." In *Proceedings of the Geoscience and Remote Sensing Symposium*, 2009 IEEE International, IGARSS 2009, Cape Town, South Africa, July 12–17, 709–12. Cape Town: IEEE Conference Publications.

European Commission. 2007. *Interpretation Manual of European Union Habitats – EUR27*. Brussels, Belgium: DG Environment, 142pp.

European Commission. 2005. *Natura 2000 in the Atlantic region*. Luxembourg: Office for Official Publications of the European Communities, 12pp.

Fauvel, M., J. Benediktsson, J. Chanussot, and J. Sveinsson. 2008. "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles." *IEEE Transactions on Geoscience and Remote Sensing* 46: 3804–14.

Gaetano, R., G. Poggi, and G. Scarpa. 2005. "Multitemporal Image Classification with Automatic Building of Tree-Structured MRF Models." In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, September 4–7. Antalya (Turkey): EURASIP.

Geman, S., and D. Geman. 1984. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–41.
Gruijter, J., D. Brus, M. Bierkens, and M. Knotters. 2006. *Sampling for Natural Resource Monitoring*. Berlin: Springer-Verlag.
Haest, B., J. Vanden Borre, G. Thoonen, T. Spanhove, S. Delalieux, L. Kooistra, C. A. Mücher, D. Paelinckx, P. Scheunders, and P. Kempeneers. 2011. "Hyperspectral Remote Sensing for Detailed Heathland Habitat Assessment in the Framework of NATURA 2000." (Submitted).
Hill, M. 1979. *TWINSPAN: A Fortran Program for Arranging Multivariate Data in an Ordered Two-way Table by Classification of the Individuals and Attributes*. Ithaca, NY: Ecology and Systematics, Cornell University.
Langanke, T., G. Rossner, B. Vrščaj, S. Lang, and J. Mitchley. 2005. "Selection and Application of Spatial Indicators for Nature Conservation at Different Institutional Levels." *Journal for Nature Conservation* 13: 101–14.
Lechner, A. M., A. Stein, S. D. Jones, and J. G. Ferwerda. 2009. "Remote Sensing of Small and Linear Features: Quantifying the Effects of Patch Size and Length, Grid Position and Detectability on Land Cover Mapping." *Remote Sensing of Environment* 113: 2194–204.
Li, S. 2001. *Markov Random Field Modeling in Image Analysis*, 2nd ed. New York: Springer-Verlag.
McCune, B., and M. Mefford. 2006. *PC-ORD. Multivariate Analysis of Ecological Data. Version 5.12*. Gleneden Beach, OR: MjM Software.
Poggi, G., G. Scarpa, and J. Zerubia. 2005. "Supervised Segmentation of Remote Sensing Images Based on a Tree-structured MRF Model." *IEEE Transactions on Geoscience and Remote Sensing* 43: 1901–11.
Smits, P., and S. Dellepiane. 1999. "Discontinuity-Adaptive Markov Random Field Model for the Segmentation of Intensity SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 37: 627–31.
Tso, B., and R. C. Olsen. 2005. "A Contextual Classification Scheme Based on MRF Model with Improved Parameter Estimation and Multiscale Fuzzy Line Process." *Remote Sensing of Environment* 97: 127–36.
Vanden Borre, J., D. Paelinckx, C. Mücher, L. Kooistra, B. Haest, G. De Blust, and A. Schmidt. 2011. "Integrating Remote Sensing in Natura 2000 Habitat Monitoring: Prospects on the Way Forward." *Journal for Nature Conservation* 19, no. 2: 116–125.