Categorizing grassland vegetation in lowland hay meadows with full-waveform airborne LIDAR: a feasibility study for Natura 2000

András Zlinszky[1,2], Anke Schroiff[3], Adam Kania[4], Balázs Deák[5], Werner Mücke[1], Ágnes Vári[3], Balázs Székely[6,7], Norbert Pfeifer[1]

[1]Vienna University of Technology, Department of Geodesy and Geoinformation, Research Groups Photogrammetry and Remote Sensing, Gußhausstraße 27 – 29, 1040 Vienna, Austria, {andras.zlinszky; werner.muecke; norbert.pfeifer}geo.tuwien.ac.at
[2] Balaton Limnological Institute, Centre for Ecological Research, Hungarian Academy of Sciences, Klebelsberg Kuno út 3, 8237 Tihany, Hungary
[3] YggdrasilDiemer, Dudenstr. 38, 10965 Berlin, Germany, ankeschroiff@yahoo.de, agnesvari@yahoo.com
[4]ATMOTERM S.A., ul. Łangowskiego 4, 45-031 Opole, Poland, kania@atmoterm.pl
[5] MTA-DE Biodiversity and Ecosystem Services Research Group Egyetem tér 1., Debrecen, H-4032, Hungary, debalazs@gmail.com,
[6] Department of Geophysics and Space Science, Eötvös University, Pázmány Péter sétány 1/C, 1117 Budapest, Hungary, balazs.szekely@ttk.elte.hu
[7] Interdisziplinäres Ökologisches Zentrum, TU Bergakademie Freiberg, Akademiestraße 6, 09596 Freiberg, Germany,

1. Introduction

Grasslands are one of the most diverse and also the most threatened habitat types, especially in Europe. They are also one of the most challenging ecosystems for field habitat mapping since many grassland habitats have a complex fine-scale mosaic structure, and in several cases it is difficult to categorise the non-typical patches.

While forests and wetlands are increasingly studied with remote sensing tools, the application of earth observation methods in grasslands remains very limited. This is partly due to the complexity of grassland habitats, and in addition the low commercial interest and the uncertainty of retrieving the necessary biophysical parameters have also restricted the use of remote sensing for such habitats.

Based on an overview of many habitat mapping studies, the EEA report on terrestrial habitat mapping (NMNH & EEA 2014) concludes that while hyperspectral airborne imaging has potential to determine species composition and multi-seasonal high temporal and spatial resolution satellite images may distinguish between some kinds of grasslands, LIDAR is not useful in this setting.

This is a problem because region-wide coverage of hyperspectral images is not to be expected in the near future, and high temporal resolution mapping can have issues with cloud cover and atmospheric corrections.

Meanwhile, LIDAR is a promising tool since airborne laser scanning campaigns are less sensitive to weather than passive optical imaging, its information content is sufficient for automatic processing in most cases, sub-meter resolutions are typical and region-wide coverage is available for most of Europe.

1.1 Objectives

Our objective was to test the application of a high quality LIDAR dataset for mapping grassland vegetation classes, to evaluate the accuracy of different approaches and collect information on LIDAR-derived parameters useful for grassland mapping.

2. Data and methods

2.1 Study site

Our study site was the Soproni-hegység Natura 2000 site in Western Hungary (N 47°41', E 16°34'), which is a hill region dominated by oak-hornbeam and beech forests. The most widespread grasslands in our study site are lowland hay meadows, which build a diverse mosaic structure with other grassland types such as semi natural dry grasslands with *Bromus* and *Festuca* species, wetter *Molinia* meadows, sedge stands, disturbed and abandoned grasslands, weed patches and shrubs. Most of these meadows are mown twice a year (late spring and late summer /early autumn) regardless of vegetation type.

2.2 Sensor data and field survey

We were constrained to use sensor data collected for the purpose of forest mapping, which means the flight dates were sub-optimal for the grassland vegetation climax: July 2011 and March 2012. A Riegl LMS-Q680 system was used, flown at an altitude of ca. 500 meters above ground. The sensor operated at the wavelength of 1550 nm with full waveform recording and a nominal ground point density of 12.8 pt/m^2.

The acquisition of field data was also carried out in several field visits, timed to allow optimal determination of grassland vegetation type before the spring mowing.

Table 1, Classes used for LIDAR-based grassland mapping

Class name (5 classes)	Class name (10 classes)	Description	Natura 2000 code
Not vegetation	Not vegetation	Asphalt, buildings, water, open soil	
Shrub	Shrub	Shrubs and small trees	
Lowland hay meadow	Lowland hay meadow	Mesophilous, species-rich grassland, mown twice a year, multi-layered canopy	6510 (strictly and exclusively)
Mown meadow (except lowland hay meadow)	Dry meadow	Xeric, calcareous, *Bromus erectus* dominated	6210 (not exclusive)
	Molinia	*Molinia* dominated, tussocks	6410
	Wet high	Wet, grass, *Carex* or *Juncus* dominated	
	Lawn	Artificial lawn regularly mown	
Not mown	Fringe	Tall forbs, local or alien species, hydrophyilous or nitrophylous	includes 6430
	Abandoned	Unmanaged former meadows	
	Meadow like	Degraded grasslands, irregular mowing	

2.3 Data processing and classification

From the LIDAR data, a set of variables were calculated in rasters of 0.5 m resolution. These were based on point attributes (reflectance, echo width, normalized height) and the roughness and variability of the target surface (sigmaZ, variance, openness), both for leaf-off and leaf-on data. Bilateral filtering (Tomasi 1998) was applied in order to conserve major gradients but get rid of random noise. The difference between leaf-off and leaf-on values of each variable was also calculated, and the final set of input rasters was loaded to a multi-band pseudo-image. Pseudo-image “spectra” from the multiband dataset were calculated for each pixel of the training data, and a random forest-based machine learning algorithm was developed in Python for band selection and classification. Since random forests assign a probability to each class for each pixel, fuzzy class membership probability output products were generated together with the classical "hard boundary" vegetation map.

50% of the ground truth polygons were set aside as an independent validation dataset, and confusion matrices were generated for each classification product.

3. Results

Results for 10 classes show overall accuracies of 66%, with a Cohen's Kappa of 0.62 (representing a "good agreement", Altmann 1990) for ten different grassland categories. Not surprisingly, the best performing categories were shrubs, not vegetation, artificial lawns and wet-high vegetation, with both producer's and users accuracies above 80%. Molinia and Dry meadows have accuracies around 70%, while abandoned grasslands have 65% producer's and user's accuracy. Apparently the most difficult categories are lowland hay meadows themselves, meadow-like areas and fringe vegetation with accuracies between 40 and 50%. However, this may be due to the difficulty of identifying these categories in the field, together with the heterogeneity within these classes.

For the alternative set of 5 classes, an overall accuracy of 74% was reached, with a Cohen’s Kappa value of 0.64. All classes have accuracies above 70%, except for lowland hay meadow, which has producer’s and user’s accuracy around 45%. Using only leaf-off or only leaf-on data did not cause substantial drop in accuracy (10 percentage points), and the contribution of noise filtering was also limited (8 ppm), but their combined effect did allow considerable improvement of accuracy. Analysis of the input channels suggested that the most important variables were (calibrated) reflectance, echo width, NDSM height and the seasonality products (differences between leaf-off and leaf-on data).

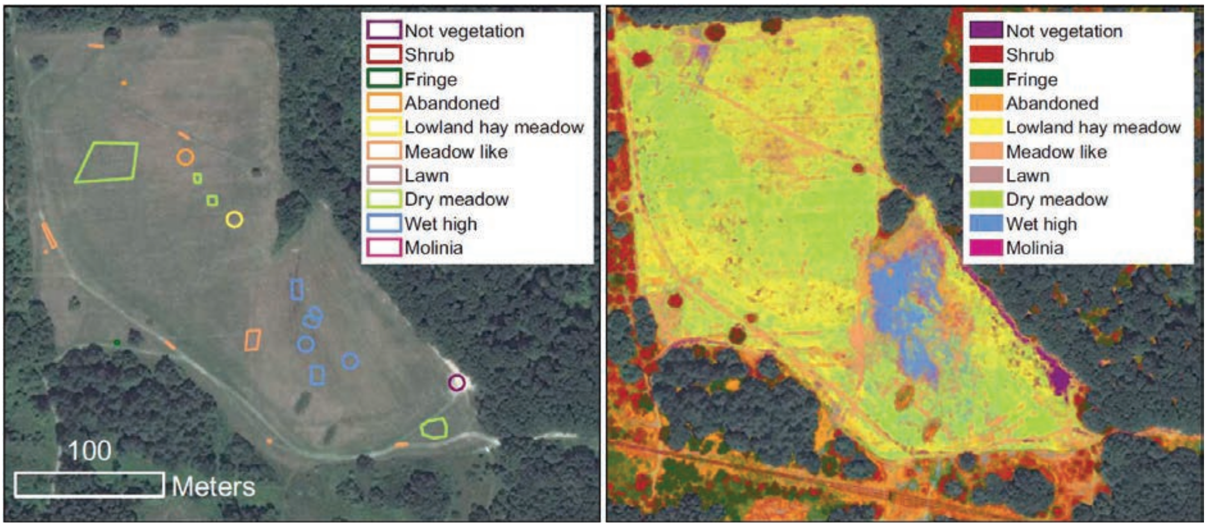

Figure 1: True colour aerial photo of a studied meadow with overlain ground truth polygons; LIDAR-based fuzzy classification of grassland categories.

4. Discussion

LIDAR-based classification of grassland vegetation was tested on a dataset with high resolution and information content, but flown with sub-optimal timing. Ground reference data collection did not include full releveés of vegetation plots, but was based on a pre-developed classification scheme with the purpose of recognizing Natura 2000 habitat types.

On one hand, the accuracy of the classification results reflects the problem of defining and identifying categories in a grassland (a problem also affecting field mapping), on the other hand they are comparable with the repeatability of field surveying itself. The information content of the point cloud was enhanced by using calibrated echo amplitude and full waveform attributes, and the large number of independent output variables was successfully handled by the machine learning algorithm. The resulting vegetation maps have a resolution of 0.5 meters, which together with the wide coverage achieved by the airborne campaign means they provide an unprecedented level of detail and pattern. The fuzzy class-membership renderings even reflect the smooth transitions between classes and the complex fine-scale mosaic structure which is so typical for a grassland, therefore we anticipate that they will be of substantial use for local conservation and monitoring. Based on these results, we believe that LIDAR has a strong potential for mapping grasslands.

Acknowledgements

This study has been carried out in the framework of ChangeHabitats2 project, an IAPP Marie Curie project of the Seventh Frame Programme of the European Commission

References

Altmann, D. G, 1990, Practical statistics for medical research. Chapman and Hall, London, UK

Muséum National d'Histoire Naturelle, European Environmental Agency, 2014, *Terrestrial Habitat Mapping in Europe: an overview* EEA Technical Report No. 1/2014, Publications Office of the European Union, Luxembourg

Tomasi, C Manduchi R., 1998, Bilateral filtering for gray and color images, In: IEEE (eds) *Sixth International Conference on Computer Vision*, Bombay, India, 839-846.