ISPRS Journal of Photogrammetry and Remote Sensing 130 (2017) 150–161

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Reducing classification error of grassland overgrowth by combing low-density lidar acquisitions and optical remote sensing data

T.P. Pitkänen *, N. Käyhkö

Geography Section, Department of Geography and Geology, University of Turku, 20014 Turku, Finland

ARTICLE INFO

Article history:
Received 28 August 2016
Received in revised form 17 May 2017
Accepted 24 May 2017

ABSTRACT

Mapping structural changes in vegetation dynamics has, for a long time, been carried out using satellite images, orthophotos and, more recently, airborne lidar acquisitions. Lidar has established its position as providing accurate material for structure-based analyses but its limited availability, relatively short history, and lack of spectral information, however, are generally impeding the use of lidar data for change detection purposes. A potential solution in respect of detecting both contemporary vegetation structures and their previous trajectories is to combine lidar acquisitions with optical remote sensing data, which can substantially extend the coverage, span and spectral range needed for vegetation mapping. In this study, we tested the simultaneous use of a single low-density lidar data set, a series of Landsat satellite frames and two high-resolution orthophotos to detect vegetation succession related to grassland overgrowth, i.e. encroachment of woody plants into semi-natural grasslands. We built several alternative Random Forest models with different sets of variables and tested the applicability of respective data sources for change detection purposes, aiming at distinguishing unchanged grassland and woodland areas from overgrown grasslands. Our results show that while lidar alone provides a solid basis for indicating structural differences between grassland and woodland vegetation, and orthophoto-generated variables alone are better in detecting successional changes, their combination works considerably better than its respective parts. More specifically, a model combining all the used data sets reduces the total error from 17.0% to 11.0% and omission error of detecting overgrown grasslands from 56.9% to 31.2%, when compared to model constructed solely based on lidar data. This pinpoints the efficiency of the approach where lidar-generated structural metrics are combined with optical and multitemporal observations, providing a workable framework to identify structurally oriented and dynamically organized landscape phenomena, such as grassland overgrowth.

1. Introduction

Extensively managed semi-natural grasslands, characterised by a high level of species diversity and consequently high ecological values, have significantly declined in Northern Europe due to recent shifts in land use regimes (Kull and Zobel, 1991; Reidsma et al., 2006; Reitalu et al., 2010; Strijker, 2005). Many of these previous extensively managed grasslands have either been converted into cultivated fields or purposely afforested, and, particularly in remote areas, abandoned leading to succession and encroachment of woody plants i.e. overgrowth (Cousins et al., 2007; Luoto et al., 2003). This has reduced the connectivity of the previous semi-natural grassland patches, which has had negative effects on the reproduction and survival of species specialized in these habitats (Luoto et al., 2003; Gajour et al., 2012). Overgrowth of semi-natural habitats is an opposite trajectory to agricultural intensification but, however, both of them will gradually lead to declined biodiversity and a less diverse landscape, thus lowering the ecosystem resilience and posing potential threats to ecosystem and landscape services (Bengtsson et al., 2003; Lindborg et al., 2008; O'Connell et al., 2015).

Since abandoned and partially overgrown grasslands play a key role in the maintenance of the grassland ecosystem network and their ecological vitality in Northern Europe, it is essential to identify those habitat fragments, which are in a state of transition to forested environments, but have not yet lost their grassland characteristics. Through designated management, these characteristics can potentially be restored and the decline in biodiversity halted (Pitkänen et al., 2014; Pykälä, 2003). To distinguish such grassland restoration targets over large geographical areas, new mapping based approaches which are capable of identifying grassland over-

* Corresponding author.
E-mail address: timo.pitkanen@utu.fi (T.P. Pitkänen).

http://dx.doi.org/10.1016/j.isprsjprs.2017.05.016
0924-2716/

growth are needed. However, overgrown grasslands are challenging to map since they represent stages of dynamic and gradual transitions from open grasslands to closed wooded environments. Successful overgrowth detection methods should be able to simultaneously capture present vegetation structures in relation to their previous changes, and therefore be able to depict essential traces of the past in the current structural habitat setting.

Optical remote sensing data such as satellite images and orthophotos have a long history as regards vegetation mapping, and their use has lately been characterised by the rapidly increased availability of images and processing methods at various scales (Xie et al., 2008). Recently, lidar systems have also been developing, and have gradually become an essential source of data for detection of structural vegetation patterns (Bradbury et al., 2005; Gaveau and Hill, 2003). Lidar has been found particularly useful in forestry-related applications to produce accurate estimates of various structural features, such as, the amount of biomass or the location of single trees (Bouvier et al., 2015; Drake et al., 2002; Dubayah and Drake, 2000; Mongus and Žalik, 2015; Sumnall et al., 2016; Yao et al., 2012). These approaches rely on multiple lidar returns, and backscattering from partially overlapping vegetation layers, which makes it possible to calculate various structural metrics (Suárez et al., 2005). Many studies have also successfully tested lidar capabilities to detect other vegetation-related patterns, such as the identification of specific plant species, their compositions, or detailed land cover classes (Donoghue et al., 2007; Heinzel and Koch, 2011; Rosso et al., 2006; Vierling et al., 2008). In addition, numerous studies have complemented lidar acquisitions with data from optical sensors or fieldwork, to increase the accuracy or provide the ground truth data needed for the analysis (Leckie et al., 2003; Margolis et al., 2015; Naidoo et al., 2012; McRoberts and Tomppo, 2007; Tonolli et al., 2011a). This has also been done due to the limited availability and applicability of lidar acquisitions and the capabilities of optical data to fill these gaps, thus, making the fusing of lidar with other data sources an essential study topic by itself (Ahmed et al., 2015; Lim et al., 2003; Ota et al., 2014). In the context of grasslands, lidar data has been used primarily for mapping various vegetation types both as a single data source (Zlinszky et al., 2014), as well as combined with multispectral acquisitions (Bork and Su, 2007; Rapine et al., 2015). These examples have indicated that lidar is an efficient tool by itself but when lidar-generated structural metrics are connected with optical and multispectral data, the combination can be even more effective in distinguishing various land cover components than its respective parts (Neuenschwander et al., 2009).

Detecting grassland overgrowth, or other dynamic landscape characteristics, must be recognised either based on multitemporal observations or explicit change-related indicators in the contemporary landscape. Repetitive measurements using optical sensors have been used, for example, to detect deforestation, regeneration and seasonal changes (Singh, 1989; Lu et al., 2004, 2016) whereas lidar has provided opportunities for e.g. calculating variations in biomass, measuring forest growth, detecting gap dynamics, and mapping damage caused by fire or snow (Englhart et al., 2013; Vastaranta et al., 2012; Vepakomma et al., 2008; Wulder et al., 2009; Yu et al., 2006). While these studies demonstrate the high applicability of lidar data for structurally oriented change detection purposes, however, the costs of lidar acquisitions over extensive areas can transpire to be unreasonably high, despite the markedly lower costs compared to charges in the past (Chen, 2007; Huang et al., 2014). In a paper by Hummel et al. (2011), the production of airborne small-footprint lidar data in Oregon with a mean pulse density of 6.31 points/m^2 over an area of 12,794 ha was concluded to result in a cost of US$ 3.34/ha. In another study, conducted by Baccini and Asner in Peru and Columbia (2013), the costs were reported to be considerably smaller (US$ 0.05–0.20/ha) which is partially due to the extensive study area (770,380 ha) and lower point density (approximately 2 points/m^2). These examples of the expenses incurred using lidar are far lower when compared to field-based sampling (Baccini and Asner, 2013; García-Gutiérrez et al., 2015), but it is evident that without a substantial source of funding, lidar is best suited to covering relatively limited areas.

It would be a tempting opportunity to benefit from the nationwide lidar programmes which have recently been established in various countries. The advantage of these lidar acquisitions is both their generally large coverage and free-of-charge dissemination, which also makes them a good option for extensive vegetation mapping approaches. The major disadvantages, however, often include their relatively low point density, contemporary lack of multitemporal data, and acquisition during the leaf-off period due to the primary focus of the topographic mapping (Valbuena et al., 2016). Multitemporal data may not always be precondition to use lidar data for detecting vegetation succession (e.g. Falkowski et al., 2009; Martinuzzi et al., 2013; van Ewijk et al., 2011), but this usually confines the detection to predefined and structurally differing phases which are delineated solely by their contemporary structure. For grassland overgrowth focused on in this study, however, this strategy is deficient as the target is on a range of succession phases which have a certain trajectory as a common denominator, but which cannot be defined based on a specific existing structure. To emphasise both contemporary vegetation structure and previous habitat dynamics, a feasible solution could be to combine a single lidar data set with other multitemporal materials such as satellite images, which provide data over a considerably longer span. This strategy has been applied in a few studies in order to, for example, detect biomass dynamics or forest floor inundation (Badreldin and Sanchez-Azofeifa, 2015; Huang et al., 2014). Furthermore, small samples of lidar data have been used to train models which are based on satellite images, thus aiming at to find a feasible compromise between the acquisition costs and frequency (Ahmed et al., 2015; Maselli et al., 2011). These studies have relied on using lidar data to build an initial model, which has then been further applied to produce predictions using other remote sensing materials. Such approaches are workable in a context where the modelling target has relatively stable spectral properties, but not over a range of various conditions.

In this study, we have developed, tested, and evaluated a procedure that aims at detecting overgrowth of a semi-natural grassland mosaic with freely available data sources and their combined usage. In the context of this paper, the definition of overgrowth is relatively wide and refers to a notable succession development which has occurred in areas dominated by grassland vegetation and deciduous trees; the definition, however, does not specifically take additional factors such as the type of overgrowth or soil nutrition into account. Thus, the results should be regarded as defining a limited pool of potentially overgrown semi-natural grasslands, although further fieldwork and analyses would be needed to confirm this status. We based our model on a single low-density lidar data set, indicating the present vegetation structure, together with multitemporal orthophoto and satellite materials, with the aim of assessing the contemporary state as well as previously incurred changes in the optically detected vegetation properties. More accurately, we used variables extracted from a publicly available small-footprint lidar data set acquired in 2009, a series of Landsat TM/ETM + satellite images between years 2001 and 2010, and two orthophoto layers from years 2002 and 2012/13. These variables, along with training data, were modelled using a Random Forest algorithm to predict whether grassland overgrowth had occurred. Models were created both based on single data sets as well as their combinations, to identify the performance of the used data sources and their applicability for overgrowth detection. Finally, we dis-