Remote Sensing of Environment 158 (2015) 140–155

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Quantifying dwarf shrub biomass in an arid environment: comparing empirical methods in a high dimensional setting

H. Zandler [a,*], A. Brenning [b,c,d], C. Samimi [a,e]

[a] University of Bayreuth, Department of Geography, 95440 Bayreuth, Germany
[b] University of Waterloo, Department of Geography and Environmental Management, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
[c] University of Heidelberg, Department of Geography, 69120 Heidelberg, Germany
[d] Friedrich Schiller University, Department of Geography, Löbdergraben 32, 07743 Jena, Germany
[e] University of Bayreuth, BayCEER, 95440 Bayreuth, Germany

ARTICLE INFO

Article history:
Received 26 May 2014
Received in revised form 3 November 2014
Accepted 7 November 2014
Available online xxxx

Keywords:
Biomass
Arid Environment
Multispectral remote sensing
Empirical modeling
Landsat OLI
RapidEye

ABSTRACT

Remote sensing based biomass estimation in arid environments is essential for monitoring degradation and carbon dynamics. However, due to the low vegetation cover in these regions, satellite-based research is challenging. Numerous potentially useful remotely-sensed predictor variables have been proposed, and several statistical and machine-learning techniques are available for empirical spatial modeling, but their predictive performance is yet unknown in this context. We therefore modeled total biomass in the Eastern Pamirs of Tajikistan, a region with extremely low vegetation cover, with a large set of satellite based predictors derived from two commonly used sensors (Landsat OLI, RapidEye), and assessed their utility in this environment using several suitable modeling approaches (stepwise, lasso, partial least squares and ridge regression, random forest). The best performing model (lasso regression) resulted in a RMSE of 992 kg ha^{-1} in spatial cross-validation, indicating that biomass quantification in this arid setting is feasible but subject to large uncertainties. Furthermore, pronounced over-fitting in some commonly used models (e.g. stepwise regression, random forest) underlined the importance of adequate variable selection and shrinkage techniques in spatial modeling of high dimensional data. The applied sensors showed very similar performance and a combination of both only slightly improved results of better performing models. A permutation-based assessment of variable importance showed that some of the most frequently used vegetation indices are not suitable for dwarf shrub biomass prediction in this environment. We suggest that predictor variables based on several bands accounting for vegetation as well as background information are required in this arid setting.

1. Introduction

Drylands cover extensive parts of the planet with about one fifth of the land surface classified as arid or drier (Lal, 2004; UNEP, 2012). Although biomass per unit area is normally low in those regions, the vast extent of the earth's arid lands gives them a significant role as a carbon pool and for the supply of essential ecosystem services (Perez-Quezada, Delpiano, Snyder, Johnson, & Franck, 2011; Trumper, Ravilious, & Dickson, 2008; Safriel & Adeel, 2005). Woody perennial vegetation has a most prominent position in drylands as it stabilizes the soil, is a year round forage source, and can be used as firewood. However, over-exploitation may lead to degradation with a loss of productivity, stored carbon, and ecosystem resilience (Eisfelder, Kuenzer, & Dech, 2012; Breckle & Wucherer, 2006; Lal, 2004). Even though desertification occurs on a local scale, it can be considered as a major global problem because of the large area affected (FAO, 2004; UNEP, 2012). Therefore, comprehensive research of desertification requires remote sensing as a tool for mapping biomass in arid environments (Eisfelder et al., 2012; Trumper et al., 2008; Yang, Weisberg, & Bristow, 2012).

In spite of these needs, remote sensing studies of vegetation in arid regions are scarce, and additional methodological research is needed to address the specific challenges faced by remote sensing techniques in these environments (Eisfelder et al., 2012). In particular, sparse and senescent vegetation may lead to a weak or ambiguous spectral response that is strongly influenced by soil background (Eisfelder et al., 2012). This also limits the utility of common vegetation indices (Asner & Heidebrecht, 2002; Montandon & Small, 2008; Yang et al., 2012). Therefore, remotely-sensed vegetation analysis in areas with plant cover under 30% has had limited success or was considered impossible (Escadafal & Chehbouni, 2008; Okin, Roberts, Murray, & Okin, 2001).

To address this methodological issue and provide biomass information in arid environments, different techniques and sensors have been applied and tested. However, existing research does not provide clear recommendations on the suitability of specific sensors for arid environments, although some studies indicate that sensors in the spectral region of the red edge such as RapidEye may be more effective than conventional sensors such as Landsat OLI (Eisfelder et al., 2012; Li, Gao, Bai, & Huang,

* Corresponding author. Tel.: +49921554636.
E-mail address: harald.zandler@uni-bayreuth.de (H. Zandler).

http://dx.doi.org/10.1016/j.rse.2014.11.007
0034-4257/

2012; Ren, Zhou, & Zhang, 2011). It is therefore our goal to assess the suitability of existing spectral indices derived from two different sensors, Landsat OLI and RapidEye in mapping biomass in a dryland environment.

An additional challenge in this context and in environmental remote sensing as a whole lies in the use of large numbers of potentially useful predictor variables to model and map a biophysical variable that is observed at a limited number of reference sites. While numerous exploratory as well as predictive tools from computational statistics and machine learning have been introduced into remote sensing in recent years, best results are achieved using varying techniques adapted to the given context (e.g., Brenning, Long, & Fieguth, 2012; Xu, Li, & Brenning, 2014).

The main intention of this study is to map dwarf shrub total biomass (TB) in an environment with extremely low vegetation cover (dwarf shrubs <20%) using a large number of predictor variables consisting of individual bands, indices, topographic attributes and texture variables derived from mutli-spectral Landsat OLI, RapidEye and ASTER GDEM satellite data. Embedded in this primary objective are two key challenges: (1) to apply and evaluate different empirical models to derive dwarf shrub biomass (stepwise, lasso, partial least squares and ridge regression as well as random forest); (2) and to compare two different commonly used relatively new sensors (Landsat OLI and RapidEye) for their suitability for vegetation detection in extremely arid environments.

2. Study area

Research was carried out in a high mountain desert landscape located in the Eastern Pamirs of Tajikistan (Fig. 1). The plateau-like region with elevations between 3500 and 5500 m above sea level is characterized by broad valleys with moderate slopes. A harsh climate with cold temperatures (Murghab annual mean 1998–2012: −1 °C, Tajik Met Service, 2013) and low precipitation (Murghab annual mean 1998–2012: 94 mm, Tajik Met Service, 2013) only allows scarce vegetation with the exception of azonal vegetation with increased water supply from orographically induced higher rainfall rates at high altitudes, surface water or groundwater (Fig. 2a & b). The majority of the non-riparian vegetation is dominated by dwarf shrubs, locally referred to as "Teresken" (Kraudzun, Vanselow, & Samimi, 2014; Vanselow, 2011). The regional dwarf shrubs are widespread woody species typical for steppe or semi-desert/desert habitats (Heklau & Röser, 2008; Heklau & von Wehrden, 2011; McArthur & Stevens, 2004).

Animal husbandry is the main economic activity and provides a livelihood for the local population. Dwarf shrubs, as the only woody plants, therefore play an important role as both forage and fuel source. For the latter purpose, the entire plant is dug up as the majority of the plant mass is located underground in the root zone (Fig. 2c). The current use has led to concerns regarding overexploitation, degradation and desertification (Breckle & Wucherer, 2006). Although much work on dwarf shrubs was done from the 1950s until the 1980s as reviewed by Walter and Breckle (1986), and generalized classifications on the occurrence of dwarf shrubs have been carried out (Kraudzun et al., 2014; Vanselow, 2011, Vanselow & Samimi, 2014), the actual occurrence and density of dwarf shrub biomass distribution remains a large research gap. The most recent remote sensing approach to quantify dwarf shrub cover by Vanselow and Samimi (2014) was abandoned due to unsatisfactory results. Besides aforementioned work, only two unpublished diploma theses exist that use satellite data and those led to very limited information regarding vegetation cover (Vanselow, 2011).

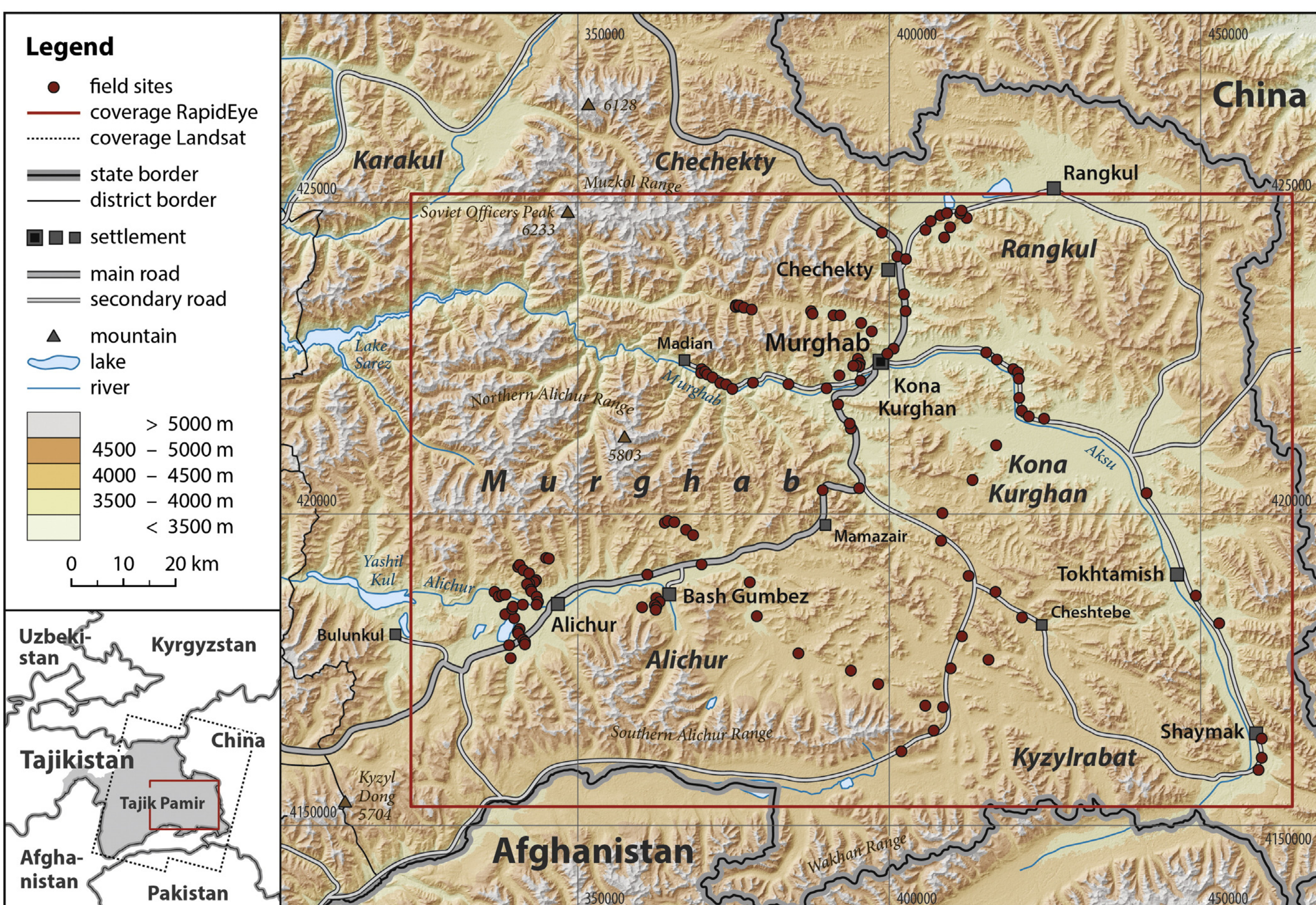

Fig. 1. Map of the study area. Research is restricted to the region covered by RapidEye images.