

# Comparison modeling for alpine vegetation distribution in an arid area

Jihua Zhou · Liming Lai · Tianyu Guan · Wetao Cai · Nannan Gao · Xiaolong Zhang · Dawen Yang · Zhentao Cong · Yuanrun Zheng

Received: 9 November 2015 / Accepted: 12 June 2016 / Published online: 15 June 2016  
© Springer International Publishing Switzerland 2016

**Abstract** Mapping and modeling vegetation distribution are fundamental topics in vegetation ecology. With the rise of powerful new statistical techniques and GIS tools, the development of predictive vegetation distribution models has increased rapidly. However, modeling alpine vegetation with high accuracy in arid areas is still a challenge because of the complexity and heterogeneity of the environment. Here, we used a set of 70 variables from ASTER GDEM, WorldClim, and Landsat-8 OLI (land surface albedo and spectral vegetation indices) data with decision tree (DT), maximum likelihood classification (MLC), and random forest (RF) models to discriminate the eight vegetation groups and 19 vegetation formations in the upper reaches of the Heihe River Basin in the Qilian Mountains, northwest China. The combination of variables clearly discriminated vegetation groups but failed to discriminate vegetation formations. Different variable combinations performed

differently in each type of model, but the most consistently important parameter in alpine vegetation modeling was elevation. The best RF model was more accurate for vegetation modeling compared with the DT and MLC models for this alpine region, with an overall accuracy of 75 % and a kappa coefficient of 0.64 verified against field point data and an overall accuracy of 65 % and a kappa of 0.52 verified against vegetation map data. The accuracy of regional vegetation modeling differed depending on the variable combinations and models, resulting in different classifications for specific vegetation groups.

**Keywords** Classification tree · Random forest · Landsat8 OLI · Spectral vegetation indices · Vegetation mapping · Qilian Mountains

J. Zhou · L. Lai · T. Guan · W. Cai · N. Gao · X. Zhang · Y. Zheng (✉)

Key Laboratory of Resource Plants, Beijing Botanical Garden, West China Subalpine Botanical Garden, Institute of Botany, Chinese Academy of Sciences, No. 20 Nanxincun, Xiangshan, Beijing 100093, China  
e-mail: zhengyr@ibcas.ac.cn

J. Zhou · T. Guan · W. Cai · N. Gao · X. Zhang  
University of Chinese Academy of Sciences, Beijing 100049, China

D. Yang · Z. Cong  
State Key Laboratory of Hydro-science and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China

## Introduction

Vegetation maps provide important baseline information on the spatial distribution of species and vegetation communities, and they provide an important reference for resource management, biodiversity conservation, and ecological services assessment (Newell and Leathwick 2005; Ohmann et al. 2014). Accurate vegetation mapping and classification has become a fundamental topic for vegetation ecology research (van der Maarel and Franklin 2013) and a key requirement for ecosystem management and global change research (Xie et al. 2008; Corbane et al. 2015).

Traditional field surveys are the first step and main data resource for vegetation mapping. Field surveys can collect a wide range of reliable data, but they are costly and labor intensive, especially in remote areas such as alpine regions and arid ecosystems where ground-based data mapping becomes more logistically challenging (Schmidt and Skidmore 2003). For large-scale mapping, the boundaries of different vegetation units are manually defined using information based on the external aspects of the landscape, such as elevation or soil type (Zhang et al. 2008), and this can cause inaccuracy in transition areas.

Remote sensing techniques offer new ways to map vegetation, especially for complex and inaccessible areas (Burai et al. 2015). Today, a broad variety of data are available from different sensors, e.g., multi-resolution optical (multispectral and hyper spectral) imagery, radar, and LiDAR products. These produce useful information for mapping natural habitats and their status (Corbane et al. 2015). The main sensors used for vegetation mapping were reviewed by Xie et al. (2008). Although technological advances have produced innovative remote sensors, including hyper spectral sensors (Schmidt and Skidmore 2003) and synthetic aperture radar (SAR) (van Beijma et al. 2014), the multispectral images from Landsat ETM+ are still among the most useful platforms for landscape-scale vegetation mapping (Cohen and Goward 2004; Hansen and Loveland 2012; Loveland and Dwyer 2012). Digital environmental data layers (terrain, geology, and soils) were incorporated as either ancillary or collateral data in the early efforts to develop regional vegetation maps from remote sensing data, and image classification methods have been well described (Richards and Richards 1999; Schowengerdt 2006). However, mapping plant communities through image classification is still limited by spectral similarities among different communities (Zhang et al. 2014). Predictive vegetation mapping determines the geographic distribution of vegetation composition across a landscape from mapped environmental variables using different models (Franklin 1995). As described by Franklin (2010), machine learning modeling methods, which are free from distributional assumptions (Guisan and Zimmermann 2000), can produce more realistic results than traditional vegetation mapping methods (Sluiter 2005). These new methods include support vector machines, artificial neural networks, and classification trees (Hastie et al. 2009). The hope is that predictive vegetation mapping will solve

some of the shortcomings of traditional vegetation mapping and image classification (Franklin 1995, 2010; Cawsey et al. 2002; Newell and Leathwick 2005).

Alpine regions are more sensitive to climate change than warmer climates and there are few refuges for alpine plants; therefore, vegetation mapping in alpine areas is essential (Zimmermann and Kienast 1999; Mark et al. 2000). However, vegetation mapping in alpine regions is challenging because of the rapidly changing environmental gradients and variations in microtopography that accompany increasing elevation. In alpine regions, spatial vegetation patterns are mainly controlled by water and temperature (Zhang et al. 2008), although solar radiation and wind also influence vegetation distribution (Frank 1988). Though climate variables control the vegetation distribution in alpine regions, climate is influenced by local terrain and can be hard to measure. Therefore, terrain variables are widely used as surrogates (Zhang et al. 2008). Compared to alpine environments in wetter regions, mapping alpine vegetation in arid areas is even more complicated, because distributions cover a range of vegetation types from desert in the lower reaches to glaciers in the upper reaches. Developing reliable, accurate mapping methods for such vegetation types is essential, especially in remote and undeveloped areas (Wang et al. 2003; Zhao et al. 2006; Cheng et al. 2014).

The Qilian Mountains in northwestern China, which include the upper reaches of the second longest inland river in China, the Heihe River, form the ecotone of the Qinghai–Tibet Plateau, the Loess Plateau, the Central Asian desert, and the Qaidam desert (Chen et al. 1994). The region's vegetation is typical alpine located in an arid area, and it is an ideal study site for developing an integrated vegetation mapping method for highly heterogeneous vegetation (Zhao et al. 2006; Cheng et al. 2014). Furthermore, vegetation changes in the Qilian Mountains will significantly affect the water yield of the Heihe River. This in turn will affect desert ecosystem stability in Inner Mongolia in the lower reaches and oases that supply water for large human populations and maintain environmental stability along the Silk Road in the middle reaches. There are some existing reports on land cover and species distribution in the region (Han 2002; Zhao et al. 2006; Gong et al. 2014). However, there has been no attempt to develop a method for mapping vegetation using multiple data sources for this complicated landscape and crucial region (Cheng et al. 2014).

In this paper, we developed a framework to integrate multi-temporal satellite data with geographic information and ground-based vegetation mapping data to enhance the accuracy of vegetation distribution modeling. The objectives were to determine which spectral and geospatial predictor variables had a significant relationship with different types of vegetation and enhanced classification accuracy for vegetation mapping and to identify a suitable modeling method for arid alpine vegetation distribution in the upper reaches of the Heihe River Basin in the Qilian Mountains, northwestern China.

## Methods

### Study area and vegetation

The upper reaches of the Heihe River Basin are located in the middle section of the Qilian Mountains, which range from 98° 34' to 101° 11' E and 37° 41' to 39° 05' N (Fig. 1) and cover an area of approximately 10,009 km<sup>2</sup>. The Qilian Mountains lie to the north of the Qinghai–Tibet Plateau and to the south of the Hexi Corridor (Silk Road). Elevation ranges from 1668 to 5062 m (calculated from ASTER GDEM, <http://westdc.westgis.ac.cn/>). The area is a typical arid and semi-arid alpine region. The mean annual precipitation ranges from 149 to 486 mm. The mean annual temperature ranges from 6.9 to −9.8 °C, with cooler averages at higher elevations (climate data sourced from WorldClim, <http://www.worldclim.org/>). Precipitation decreases from east to west and increases from north to south in the study area, but temperature shows the reverse pattern (Zhao et al. 2005; Qin et al. 2013; Gao et al. 2014).

The most recent vegetation map is the Vegetation Map of the People's Republic of China (VMC, 1:1,000,000) (Editorial Committee of Vegetation Map of China, the Chinese Academy of Sciences 2007). The map contains 11 vegetation groups, 55 vegetation types, and 859 vegetation formations and sub-formations. It was completed in 2007 based on field survey data. There are vegetation, cultivated vegetation, eight vegetation types, and 18 formations in the study area (Table 1). The lowlands (1600–2400 m) are mainly desert and the upper regions (2400–2800 m) are steppe, consisting of *Stipa* spp., with needle-leaf forest in the north ranging from 2400 to 3200 m, scrub-meadows from 3200 to 4000 m, and alpine vegetation, mainly

*Saussurea* spp., in areas higher than 4000 m. Glaciers form at the peaks of some mountains. The main land use is grazing; the forest is protected by the government, and logging has been forbidden in recent years. Some cultivated vegetation is grown near county towns, but cultivation covers an area of less than 1 % in this region.

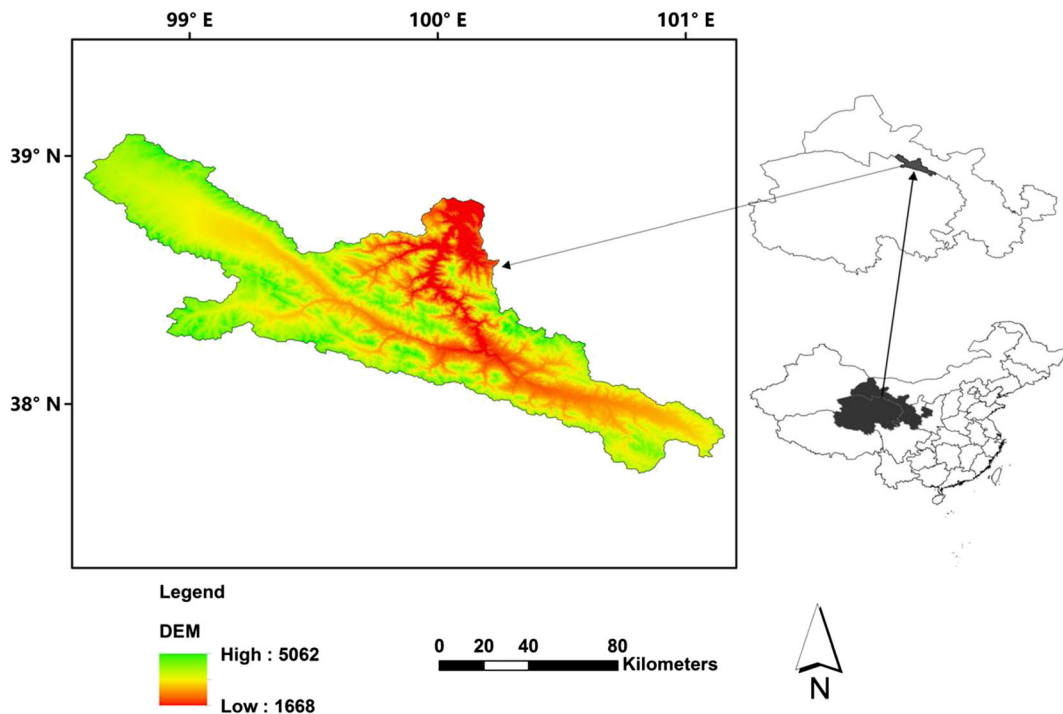
### Vegetation and training data

The training and model assessment data were obtained from multiple sources, including field data collected in April 2013, July 2013, and September 2014, Heihe eco-hydrological transect survey data (Feng et al. 2014), and data published in the literature. These data contained information on total vegetation cover, plant species cover and abundance, and vegetation height. A total of 1220 vegetation plots were surveyed, of which 1007 were used for model training and 213 for model assessment. For every vegetation formation, 51–60 datasets were used, whereas for every vegetation group, all data in the formations belonging to the group were used, totaling between 51–260 datasets. In this area, there was only one additional vegetation type classified from the higher level vegetation groups, and this vegetation type (*Stipa* spp.) is difficult to distinguish from others using remote sensing data, or even by field survey. Therefore, for vegetation modeling, we only used two units, i.e., vegetation group and vegetation formation. The VMC vegetation map was used for model assessment.

### Spectral and geospatial data

ASTER GDEM (30-m grid) was acquired from the Cold and Arid Regions Science Data Center at <http://westdc.westgis.ac.cn/>. Elevation, aspect, and slope were derived from the ASTER GDEM. These are widely used terrain variables in vegetation distribution modeling (Franklin 2010).

Climate data with a 1-km resolution was downloaded from WorldClim (Hijmans et al. 2005) at <http://www.worldclim.org/> and was resampled to a 30 × 30 m grid cell size using a nearest-neighbor method in ArcGIS v.10.0 (ESRI, Redlands, CA, USA 2010). The downloaded data included minimum, maximum, and average annual temperature and average precipitation. These data have a proven relationship with vegetation distribution (Franklin 2010) and are commonly used as bioclimatic constraints in vegetation models (Sitch et al. 2003).



**Fig. 1** The location and DEM of the upper reach of Heihe

Landsat 8 OLI images (part of WRS-2 path 133, row 34; path 133, row 33; path 134, row 31) in summer (July 2013), autumn (early October 2013), and winter (January 2014) were acquired from the United States Geological Survey at <http://www.usgs.gov/> and Geospatial Data Cloud at <http://www.gscloud.cn/>. Image pre-processing included geographic correction, radiometric correction, FLASSH atmosphere correction, and image subset mosaicking in ENVI v.5.1 (Exelis VIS, Boulder, CO, USA 2013).

The spectral response of vegetated areas represents a complex mixture of vegetation, soil brightness, environmental effects, shadow, soil color, and moisture (Bannari et al. 1995). Preliminary studies have shown that spectral vegetation indices (SVIs) are more sensitive to vegetation type and phenology than individual spectral bands (Bannari et al. 1995). SVIs are widely used to discriminate vegetation type and land cover and are at the heart of ecological applications of remote sensing (Cohen and Goward 2004). SVIs are beneficial in landscape classification and deriving continuous estimates of vegetation biophysical characteristics. The most widely used SVI is the normalized difference vegetation index (NDVI) (Price et al. 2002; Cohen and Goward 2004). Various SVIs have been tested to find

the best index for vegetation discrimination (Table 2). The soil-adjusted vegetation index (SAVI) combines soil and vegetation reflectance. Different vegetation types have different reflectance ratios for SAVI calculation. We used fixed adjustment factors of 1.16 and 1.5 in SAVI (Huete 1988) and in the optimized soil-adjusted vegetation index (OSAVI) (Rondeaux et al. 1996).

For simplicity, we refer to Landsat8 OLI land surface albedos (LSA) and spectral vegetation indices (SVIs) as spectral variables and others as geospatial variables. There were seven geospatial variables and 63 spectral variables (seven SLAs and 14 SVIs for summer, autumn, and winter). All of the grid data (70 variables) were resampled to the OLI pixel size because this was the principal data source for discerning major vegetation units. Different variable combinations were used when decision tree (DT) and random forest (RF) classifications were derived (Table 3). Variable combinations 1–8 contained either one type of geospatial or seasonal spectral variable, combination 9 contained geospatial variables, combinations 10–15 contained geospatial variables with one additional set of either SLAs or SVIs in the same season, combinations 16–18 contained geospatial variables with one additional seasonal SLAs and SVIs, and combination 19 used all of the variables.

**Table 1** Classification units of the vegetation of China

Vegetation groups	Vegetation types	Formations (sub-formations)
1. Needle-leaf forest	(1) Cold-temperate and temperate mountains needle-leaf forest	(1) <i>Picea crassifolia</i> forest
2. Scrub	(2) Subalpine broadleaf deciduous scrub	(2) <i>Salix gilashania</i> scrub
2. Scrub	(2) Subalpine broadleaf deciduous scrub	(3) <i>Salix oritrepha</i> scrub
2. Scrub	(2) Subalpine broadleaf deciduous scrub	(4) <i>Salix oritrepha</i> , <i>Dasiphora fruticosa</i> scrub
2. Scrub	(2) Subalpine broadleaf deciduous scrub	(5) <i>Dasiphora fruticosa</i> scrub
3. Desert	(3) Temperate semi-shrubby and dwarf semi-shrubby desert	(6) <i>Sympegma regelii</i> desert
4. Steppe	(4) Temperate needle-grass arid steppe	(7) <i>Stipa krylovii</i> steppe
4. Steppe	(4) Temperate needle-grass arid steppe	(8) <i>Stipa penicillata</i> steppe
4. Steppe	(4) Temperate needle-grass arid steppe	(9) <i>Stipa breviflora</i> , <i>S. bungeana</i> steppe
4. Steppe	(5) Alpine grass, <i>Carex</i> steppe	(10) <i>Stipa purpurea</i> alpine steppe
5. Meadow	(6) Alpine <i>Kobresia</i> spp., forb meadow	(11) <i>Kobresia pygmaea</i> alpine meadow
5. Meadow	(6) Alpine <i>Kobresia</i> spp., forb meadow	(12) <i>Kobresia humilis</i> alpine meadow
5. Meadow	(6) Alpine <i>Kobresia</i> spp., forb meadow	(13) <i>Kobresia filifolia</i> alpine meadow
5. Meadow	(6) Alpine <i>Kobresia</i> spp., forb meadow	(14) <i>Kobresia schoenoides</i> , <i>Carex</i> spp. swamp alpine meadow
5. Meadow	(6) Alpine <i>Kobresia</i> spp., forb meadow	(15) <i>Elymus nutans</i> , <i>Roegneria nutans</i> alpine meadow
6. Alpine vegetation	(7) Alpine sparse vegetation	(16) <i>Saussurea medusa</i> , <i>Saussurea</i> spp. sparse vegetation
6. Alpine vegetation	(7) Alpine sparse vegetation	(17) <i>Saussurea</i> spp., <i>Rhodiola rosea</i> , <i>Cremanthodium</i> spp. sparse vegetation
7. Culture vegetation	(8) One crop annually short growing period cold-resistant crops	(18) Spring barley, spring wheat, potatoes, turnip, pea, rapessed
8. Land without vegetation	(9) Land without vegetation	(19) Glaciers and snow limit

Vegetation distribution models

Decision tree (DT), random forest (RF), and maximum likelihood classification (MLC) models were used to model vegetation distribution. The DT model has a visual structure and uses classification and regression tree algorithms (Hastie et al. 2009) to generate classification rules. Different DTs have different classification rules. We used a DT with five layers, in which the smallest parent node had 40 samples and the smallest child node had 10 samples.

RF is an ensemble method and a black box model and has been applied successfully in a number of ecological land cover studies (Cutler et al. 2007; Corcoran et al. 2013). The RF algorithm improves classification accuracy considerably, as well as being largely insensitive to noisy data sets (Gislason et al. 2006). RF can also produce estimates of variable importance, providing quantitative analysis of variable contributions (Gislason et al. 2006; Corcoran et al. 2013). The RF

models were generated using the default settings in EnMAPBox, with 100 trees, and a Gini coefficient was used for the node impurity function (van der Linden et al. 2015). The values of each variable and classification precision for each model were calculated.

MLC is a widely used supervised classification method and one of the best methods for many classification problems (Franklin 2010; Burai et al. 2015). The algorithm is based on a Gaussian probability density function model and each pixel is allocated to the class with the highest probability. The number of training samples must be larger than the number of variables in MLC; therefore, only variable combinations 1–5, 9–12, and 19 were used.

Predicted vegetation maps were generated for DT and RF using 19 variable combinations, and for MLC using ten variable combinations, resulting in 48 maps. The DT and RF model results contained variable importance, indicating which variables were important for vegetation discrimination.



**Table 2** The vegetation index

Index	Abbreviation	Formula	Author and year
Ratio vegetation index	RVI	$NIR/Red$	Pearson and Miller 1972
Brightness index	BI	$0.2909Blue + 0.2493Green + 0.4806Red$ $+ 0.5568NIR + 0.4438SWIR1 + 0.1706SWIR2$	Crist et al. 1986
Green vegetation index	GI	$-0.2728Blue - 0.2174Green$ $- 0.5508Red + 0.7221NIR$ $+ 0.0733SWIR1 - 0.1648SWIR2$	Crist et al. 1986
Wetness index	WI	$0.1446Blue + 0.1761Green + 0.3322Red$ $+ 0.3396NIR - 0.6210SWIR1 - 0.4186SWIR2$	Crist et al. 1986
Differenced vegetation index	DVI	$NIR - Red$	Clever 1986
Green ratio	GR	$NIR/Green$	Price et al. 2002
MIR ratio	MR	$NIR/SWIR1$	Price et al. 2002
Soil-adjusted vegetation index	SAVI	$(1.5(NIR - Red))/((NIR + Red + 0.5))$	Huete 1988
Optimization of soil-adjusted vegetation index	OSAVI	$(1.16(NIR - Red))/((NIR + Red + 0.16))$	Rondeaux et al. 1996
Atmospherically resistant vegetation index	ARVI	$(NIR - (2 * Red - Blue))/((NIR + (2 * Red - Blue)))$	Kaufman and Tanre 1992
Normalized difference vegetation index	NDVI	$(NIR - Red)/(NIR + Red)$	Rouse et al. 1974
Enhanced vegetation index	EVI	$2.5[(NIR - Red)/(NIR + 6 * Red - 7.5Blue + 1)]$	Huete et al. 2002
Normalized difference tillage index	NDTI	$(SWIR1 - SWIR2)/(SWIR1 + SWIR2)$	van Deventer et al. 1997
Normalized difference senescent vegetation index	NDSVI	$(SWIR1 - Red)/(SWIR1 + Red)$	Marsset et al. 2006

## Model assessment

The remaining 213 field data points that were not used in the training process and the VMC were used to assess the overall accuracy and kappa coefficient of the models. For the kappa coefficient, Landis and Koch (1977) defined values from 0.7 to 1.0 as indicating very good-to-perfect agreement, 0.55–0.70 for good-to-very good agreement, 0.40–0.55 for fair-to-good agreement, 0.2–0.4 for poor-to-fair agreement, and 0.0–0.20 for no-to-poor agreement.

## Results

### Vegetation modeling and accuracy assessment

None of the models accurately discriminated vegetation distribution at the vegetation formation level, with overall accuracy and a kappa coefficient of <40 % and 0.2, respectively.

Vegetation distribution was clearly discriminated at the vegetation group level. The accuracy and kappa coefficients of the maps generated by complex variable combinations (9–19) were better than with simple

variable combinations (1–8) in all of the models (Tables 4 and 5). In the simple variable combinations (1–8), topography had the best accuracy and kappa coefficient results. When assessed using field point data, the RF model with variable combination 13 gave the best result: the overall accuracy was 75 % and the kappa coefficient was 0.64. The best MLC model used variable combination 12 with an overall accuracy of 67.86 % and a kappa coefficient of 0.52. The best DT model used variable combination 8 with an accuracy of 68.75 % and a kappa coefficient of 0.55. When assessed using the vegetation map, variable combinations 9–19 were better than variable combinations 1–8 in all of the models. The RF model that used either variable combination 15 or 19 had the highest overall accuracies at >65 % with a kappa value of 0.52. The best MLC model used combination 19 and had similar results with an overall accuracy of 63.84 % and a kappa coefficient of 0.51. The best DT model used variable combination 18 and had an accuracy of 57.31 % and a kappa coefficient of 0.47 (Fig. 2).

Compared with the assessment using field point data, the assessment of the models against the vegetation map showed low overall accuracy and kappa coefficients. Using the same variable combinations, the different models had similar performances assessed against the

**Table 3** Variable combinations

Variables	Variable combinations																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
<i>Geospatial variables</i>																			
Topography variables	✓								✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Climate variables		✓							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>Spectral variables</i>																			
Summer LSAs			✓							✓						✓			✓
Autumn LSAs				✓							✓						✓		✓
Winter LSAs					✓							✓						✓	✓
Summer SVIs						✓							✓			✓			✓
Autumn SVIs							✓							✓			✓		✓
Winter SVIs								✓							✓			✓	✓

LSA stands for land surface albedo, SVI stands for spectral vegetation index

field point data; however, DT had a significantly lower predictive power than RF and MLC when assessed against the vegetation map.

**Significant variables**

Elevation was the most important of the 70 variables in both the DT and RF models (Table 6). Maximum temperature was the second most important, followed by a number of other WorldClim variables, especially in the RF models. Slope was important in RF but not in DT (Table 6). The spectral variables of summer and autumn were highly important in DT, but in RF some winter SVIs were important. Summer and autumn SAVI, OSAVI, NDVI, and GI were important SVIs in DT, while winter BI was important in RF. SLAs showed a medium level of importance.

**Discussion**

We aimed to develop a framework to integrate widely available multi-temporal satellite data with geographic information and ground-based vegetation data to enhance the accuracy of vegetation distribution modeling in arid alpine landscapes. ASTER GDEM, WorldClim data, and Landsat images are commonly used in vegetation and species distribution models (Sesnie et al. 2008; Franklin 2010). Although other data, such as geology, soil, radiation, and high spectral and spatial resolution remote sensing images, are also effective for

vegetation mapping, these data are not easily available for some regions. The data used in this research are easily obtainable for most parts of the world, including alpine regions for which field data are lacking.

**Factors affecting vegetation classification levels**

Vegetation classification is a fundamental issue in ecology, and most classification schemes have two or more classification levels. For global or large areas, a hierarchical classification scheme is usually applied (Faber-Langendoen et al. 2014). In the VMC, the highest classification levels (vegetation groups and vegetation types) mainly stem from the appearance of communities and climate, the middle classification level (vegetation formations) stems from dominant species, while the lowest level (associations) stems from community structure and the dominant species in each synusia. For climate and hydrology models, vegetation groups are suitable for use as land surface types, but for biological conservation and resource utilization studies, especially for rare species, the lower levels of classification are needed (Newell and Leathwick 2005). In vegetation distribution modeling, the level of discrimination depends on the input variables. Terrain and climate variables indicate the vegetation environment, while spectral variables reflect land surface objects. The input variables can discriminate vegetation when there are significant differences (Price et al. 2002; Franklin 2010). When the input variables can distinguish most types in a classification level, this is called the

**Table 4** Model assessment by filed data

Variable combination	Decision tree		Random forest		Maximum likelihood classification	
	Overall accuracy (%)	Kappa coefficient	Overall accuracy (%)	Kappa coefficient	Overall accuracy (%)	Kappa coefficient
1	54.91	0.39	63.39	0.50	63.39	0.50
2	37.95	0.25	52.68	0.37	47.77	0.33
3	54.91	0.34	54.46	0.37	37.05	0.25
4	50.89	0.34	52.23	0.37	41.52	0.28
5	50.00	0.35	58.93	0.46	59.82	0.47
6	49.11	0.29	57.14	0.41		
7	60.27	0.43	54.46	0.40		
8	68.75*	0.55*	60.27	0.47		
9	54.46	0.38	68.30	0.56	62.50	0.48
10	57.14	0.42	72.77	0.62	61.61	0.49
11	64.73	0.50	63.84	0.50	62.95	0.49
12	61.61	0.48	70.09	0.59	67.86*	0.56*
13	59.82	0.45	75.00*	0.64*		
14	57.59	0.42	66.52	0.53		
15	59.38	0.45	73.21	0.62		
16	59.82	0.45	71.88	0.61		
17	57.59	0.42	60.27	0.45		
18	59.38	0.45	71.43	0.60		
19	59.82	0.46	70.98	0.59	66.07	0.52

Data with \* were the best results. Variable combinations were shown in Table 3

distinguishable level. In this study, the distinguishable level was vegetation group.

In alpine regions, although climate varies considerably, some vegetation formations have a similar habitat and appearance (Wang et al. 2003). They lack significant differences in spectral features, which make them difficult to distinguish. For example, the dominant species found in steppe and meadow formations are mainly *Kobresia* spp. and *Stipa* spp. These are hard to distinguish even in the field and often grow together, especially in this ecotone (Editorial Committee of Vegetation Map of China and the Chinese Academy of Sciences 2007). At elevations from 2700 to 3400 m, there is a large transition zone for steppe and meadow, with parts of the steppe distributed up to 4000–5000 m (mainly *Stipa purpurea* in the alpine steppe). Generally, steppe is a relatively xeromorphic habitat, while meadow is mesomorphic; however, both are common in this ecotone (Wang et al. 2003; Editorial Committee of Vegetation Map of China and, Chinese Academy of Sciences 2007). Similarly, scrub is mainly distributed from 3200

to 3800 m, and meadows are also found at these elevations. In this study, *Salix gilashania* was 1 to 2 m in height, but *Dasiphora fruticosa* was <0.5 m, making it hard to distinguish from meadow vegetation by its spectral features. Distinguishing forest is easy in alpine regions because there are only a few forest types (Zhao et al. 2006; Han 2002).

Phenology is important for vegetation discrimination, especially forest vegetation. There is a higher accuracy when distinguishing between broad-leaved and needle-leaved forest, although identifying species within each forest type is difficult because they often have similar spectral reflectance curves (Gao et al. 2015). In Kansas, different Landsat band combinations and vegetation indices could distinguish six grassland types because the grasses in this region have different phenologies (Price et al. 2002). When using satellite images, naturally mixed pixels always create problems with identifying vegetation and to some extent reduce the accuracy of classifications because of spectral confusion (Domaç and Süzen 2006). Different resolutions can also lead to problems



**Table 5** Model assessment by vegetation map

Variable combination	Decision tree		Random forest		Maximum likelihood classification	
	Overall accuracy (%)	Kappa coefficient	Overall accuracy (%)	Kappa coefficient	Overall accuracy (%)	Kappa coefficient
1	39.85	0.20	41.80	0.23	43.29	0.24
2	29.30	0.14	41.87	0.23	11.89	0.00
3	15.27	-0.02	4.98	0.00	0.08	0.00
4	3.15	0.00	3.22	0.00	15.42	0.00
5	3.28	0.00	2.70	0.00	0.17	0.00
6	40.92	0.24	46.93	0.31		
7	47.11	0.29	45.91	0.31		
8	49.76	0.34	51.76	0.38		
9	42.53	0.20	62.66	0.48	62.37	0.50
10	37.24	-0.02	63.80	0.51	59.89	0.47
11	38.28	-0.02	63.09	0.49	60.38	0.47
12	38.04	-0.02	63.73	0.50	63.18	0.51
13	57.61	0.42	64.28	0.51		
14	56.33	0.40	64.50	0.51		
15	57.31	0.42	65.60*	0.52*		
16	57.61*	0.42	64.72	0.52		
17	38.12	-0.02	63.98	0.50		
18	57.31	0.47*	64.83	0.52		
19	38.11	-0.02	65.42	0.52	63.84*	0.51*

Data with \* were the best results. Variable combinations were shown in Table 3

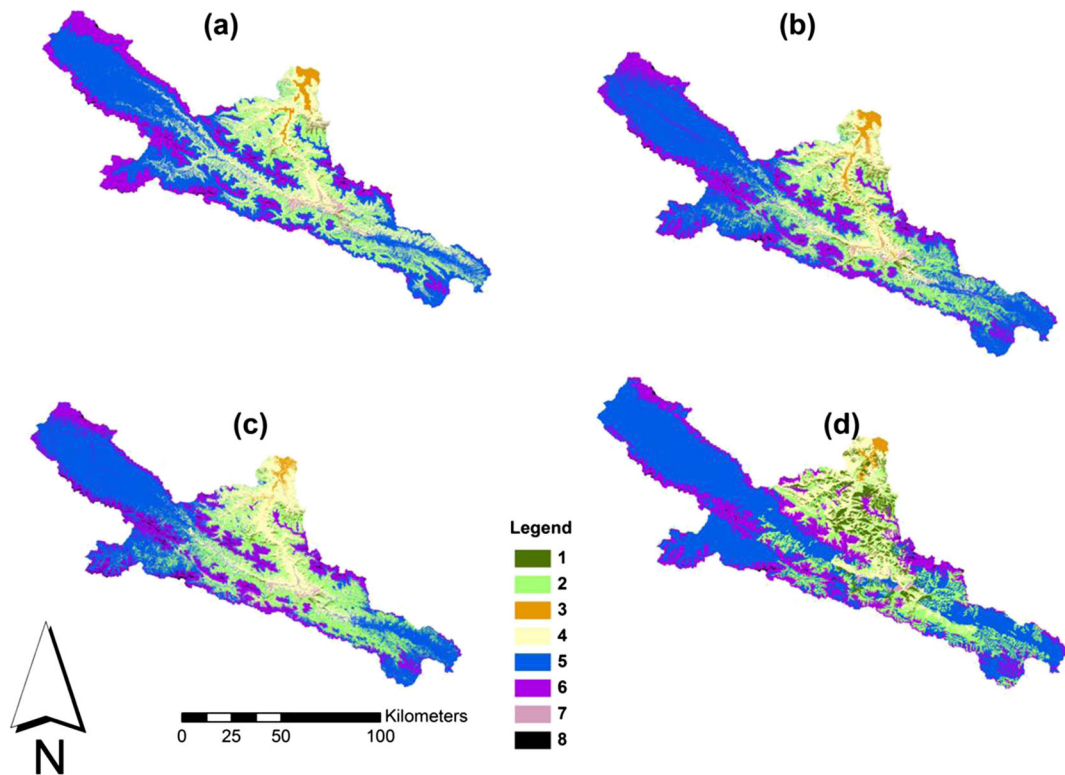
because some pixel sizes are too coarse to detect narrow gullies (van Beijma et al. 2014). In our study, all alpine vegetation had a short growing season over a similar time period and phenology, and the WorldClim data was at a 1-km resolution, both of which resulted in classification errors for several vegetation types.

Different remote sensing sensors may overcome these issues. For example, hyper spectral data contains many narrow, contiguous spectral bands, while SAR is relatively unaffected by atmospheric effects. Both are promising for future vegetation classification (Corbane et al. 2015; Landmann et al. 2015), and integrating hyper spectral and SAR data may provide opportunities for a more comprehensive understanding of land cover and change. However, those new methods are under development and are mostly used on airborne platforms. Their images tend to cover smaller areas than coarser resolution sensors. For example, the hyper spectral satellite, Hyperion, has an image width of 7.5 km, which is not suitable for large area vegetation mapping. Therefore, although they have the advantage of being able to

discriminate between vegetation species, their airborne platform sensors and lack of repeat coverage currently limit their use in landscape-scale vegetation mapping and monitoring (van Beijma et al. 2014).

The importance of variables in modeling

Elevation is the most important variable for vegetation mapping in alpine regions, especially in mountains where there are large elevation variances (Dirnbock et al. 2003; Dobrowski et al. 2008; Sesnie et al. 2008). Elevation is also an important factor in species distribution models (Oke and Thompson 2015). In the Qilian Mountains, the elevation range is >3000 m. The WorldClim climate variables are highly correlated with elevation and in some regions may not contain much more additional information than elevation (Sesnie et al. 2008). However, we found that the WorldClim variables were important in both the DT and RF models. In DT, many mid-level nodes were WorldClim variables, especially maximum temperature and annual precipitation.



**Fig. 2** The modeling vegetation map with highest accuracy by decision tree model (a), random forest model (b), maximum likelihood classification (c), and the Vegetation Map of the

People’s Republic of China in the upper reach of Heihe (d). The legend represents vegetation groups shown in Table 1

The separation of desert from other zones depended on precipitation being <221 mm, while the separation of alpine vegetation from other vegetation occurred at temperatures below 2 °C. In this research, slope also had a

high importance value and may influence the distribution of some vegetation.

SVIs were combinations of LSAs, and contained more information than single LSAs (Bannari et al.

**Table 6** Top ten most important variables

Decision tree model variables	Standardized importance (%)	Random forest model variables	Normalized importance
Elevation	100.0	Elevation	1.37
Maximum - temperature	81.0	Maximum - temperature	1.19
Autumn - SAVI	75.0	Slope	1.08
Autumn - OSAVI	75.0	Precipitation	1.06
Autumn - NDVI	75.0	Minimum - temperature	0.95
Autumn - NDTI	74.2	Mean - temperature	0.93
Summer - GI	73.5	Summer - WI	0.82
Summer - DVI	72.8	Summer - MR	0.76
Mean - temperature	69.7	Summer - LSA7	0.74
Summer - SAVI	69.0	Winter - BI	0.73

The abbreviations were shown in Tables 2 and 3

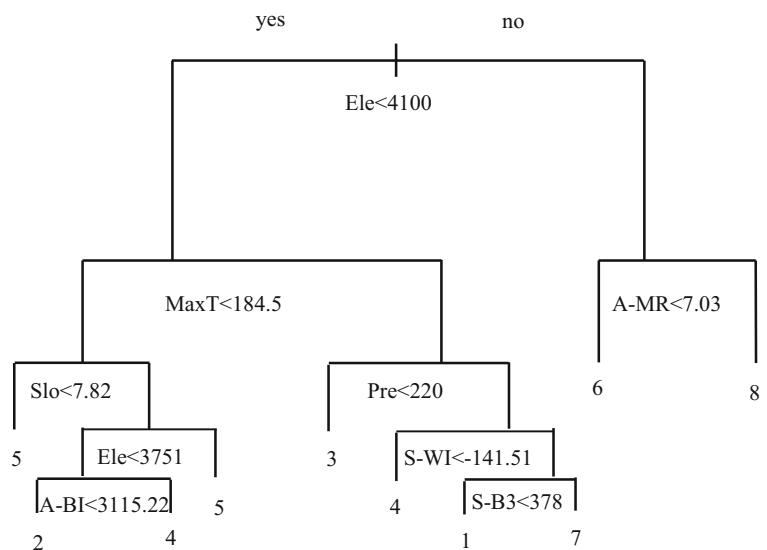
1995). SVIs had higher important values than LSAs or raw bands in our study and other studies (Price et al. 2002; Sennie et al. 2008). For the DT model, summer and autumn spectral variables were more important than winter ones, because summer and autumn images contain more information in alpine regions. In the winter false color image, forest differed significantly from other vegetation and some shrub vegetation also could be distinguished, but other vegetation showed no visual difference. In the summer and autumn false color images, forest, scrub, desert, alpine vegetation, and cultural vegetation showed visual differences, but the steppe and meadow were not distinguishable from each other. In the RF model, seasonal differences were not significant, because RF is an ensemble method (Cutler et al. 2007) that reduces the difference between spectral variables. BI, GI, and WI were combinations of all TM bands and they contained more information than the SVIs, which contained only two bands. NDVI, SAVI, and DVI were important in the DT. These are combinations of near infrared and red bands, which reflect important information on vegetation, because visible radiation in the red wavelengths (630–690 nm) is absorbed by chlorophyll while radiation in the near infrared (760–900 nm) is strongly reflected by leaf cellular structures (Bannari et al. 1995). To compare the importance of individual variables for each season and group seasonal variables in RF assessed by field data, the model had >50 % accuracy and kappa coefficient >0.37 when group seasonal variables were used; the model had <35 %

accuracy and kappa coefficient <0.2 when an individual variable was used.

The performance of different models

The modeled vegetation maps show that multi-source geospatial and spectral variables successfully mapped vegetation groups. Previous research has found that the RF model performs better when modeling vegetation distribution compared with other methods (Sluiter 2005). In this research, RF performed better than MLC and DT. RF was more stable with an increase in variables, while MLC was limited by the input variable numbers. DT produced low accuracy prediction maps when the training points were limited, but because DT had a visual structure, it was useful in extracting classification rules. In this research, elevation was the first node, while summer and autumn SVIs were used at the lowest levels of classification, so that it was clear how the classification was carried out (Fig. 3). The main vegetation groups that were confused were scrub, steppe, and meadow. Steppe and meadow cover a large transition region, while scrub is also scattered among the meadow vegetation communities. Meadow and alpine vegetation, steppe, and desert ecotones might change in different years, which may contribute to low accuracy in the model results. Furthermore, the Chinese vegetation map is an ecological classification scheme, and not a remote sensing classification one, so some vegetation had similar spectral characteristics (Ouyang et al. 2015),

**Fig. 3** The structure of the decision tree model using variable combine 19 shown in Table 3. The numbers represent vegetation groups in Table 1. *Ele* represents elevation, *Slo* represents slope, *MaxT* represents maximum annual temperature, and *Pre* represents average precipitation, *A-BI* represents autumn brightness index, *A-MR* represents autumn MIR ratio, *S-WI* represents summer wetness index, and *S-3* represents summer land surface albedo for green band



which increased the discrepancy between the modeled and existing vegetation maps.

Compared with other studies using similar variables and similar size study areas, our results had similar accuracy using both point and map assessment, with the point accuracy being higher than the map accuracy. Vegetation mappers in the Sierra Nevada obtained their best result using DT models with an overall accuracy of 75 % and kappa coefficient of 0.69 using point assessment (Dobrowski et al. 2008). In Costa Rica and Nicaragua, the overall accuracy was 81 % using point assessment (Sesnie et al. 2008), while in Svalbard, Norway, an overall accuracy of 55.36 % and a kappa coefficient of 0.48 were obtained using a conventional vegetation map assessment (Johansen et al. 2012). In this study, assessment points were taken from either field surveys or visual interpretation, and we are confident of their accuracy. Compared with conventional vegetation maps, the transition zone and natural mixed pixels always result in lower accuracy (Domaç and Süzen 2006); therefore, mapping accuracy is lower than it is with point assessment. This suggests that point assessment is the best choice for areas without conventional vegetation mapping but may not be useful over wide areas. However, while conventional vegetation map assessment may be less accurate than point assessment, these maps are useful for understanding the modeled vegetation distribution trend. A combination of conventional vegetation map assessment and point assessment is recommended for accurate assessment when both types of data are available. In this study, the vegetation group map predicted by the best RF model and the VMC map had similar distributions. However, there were some differences, e.g., boundary for meadow vegetation, and false classification for vegetation in specific areas, e.g., vegetation close to riversides were classified as desert at low elevations and alpine sparse vegetation at high elevations; however, in the VMC, they were classified into nearby vegetation groups because they had similar species compositions. Although there were some differences, the results were acceptable based on the accuracy and kappa coefficient values.

## Conclusion

The combination of ASTER GDEM, WorldClim, and multi-season Landsat-8 OLI data can be used to discriminate alpine vegetation at relatively coarse

vegetation classification levels, i.e., vegetation groups in the Qilian Mountains in an arid area of northwestern China. However, the data failed to classify vegetation into more detailed classes, i.e., vegetation formation. Among the 70 variables used for modeling, elevation was the most important. The RF model classified vegetation more accurately than the MLC and DT models.

**Acknowledgments** This work was funded by National Natural Science Foundation of China [91225302].

## References

- Bannari, A., Morin, D., Bonn, F., & Huete, A. (1995). A review of vegetation indices. *Remote Sensing Reviews*, 13(1–2), 95–120.
- Burai, P., Deak, B., Valko, O., & Tomor, T. (2015). Classification of herbaceous vegetation using airborne hyperspectral imagery. *Remote Sensing*, 7(2), 2046–2066.
- Cawsey, E., Austin, M., & Baker, B. L. (2002). Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodiversity and Conservation*, 11(12), 2239–2274.
- Chen, G., Peng, M., Huang, R., & Lu, X. (1994). Vegetation characteristics and its distribution of Qilian mountain region. *Acta Botanica Sinica*, 36, 63–72 (in Chinese).
- Cheng, G., Li, X., Zhao, W., Xu, Z., Feng, Q., Xiao, S., et al. (2014). Integrated study of the water–ecosystem–economy in the Heihe River Basin. *National Science Review*, 1(3), 413–428.
- Clevers, J. (1986) The application of a vegetation index in correcting the infrared reflectance for soil background. In *Remote sensing for resources development and environmental management. International symposium*. 7, (pp. 221–226)
- Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, 54(6), 535–545. doi:10.1641/0006-3568(2004)054[0535:lrieao]2.0.co;2.
- Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., Milian, V. E. G., et al. (2015). Remote sensing for mapping natural habitats and their conservation status—new opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 37, 7–16. doi:10.1016/j.jag.2014.11.005.
- Corcoran, J. M., Knight, J. F., & Gallant, A. L. (2013). Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. *Remote Sensing*, 5(7), 3212–3238. doi:10.3390/rs5073212.
- Crist, E. P., Laurin, R., & Cicone, R. C. (1986) Vegetation and soils information contained in transformed Thematic Mapper data. In *Proceedings of IGARSS'86 Symposium*, (pp. 1465–1470): European Space Agency Publications Division Paris
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., & Hess, K. T. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. doi:10.1890/07-0539.1.

- Dirnbock, T., Dullinger, S., Gottfried, M., Ginzler, C., & Grabherr, G. (2003). Mapping alpine vegetation based on image analysis, topographic variables and canonical correspondence analysis. *Applied Vegetation Science*, 6(1), 85–96. doi:10.1658/1402-2001(2003)006[0085:mavboj]2.0.co;2.
- Dobrowski, S. Z., Safford, H. D., Cheng, Y. B., & Ustin, S. L. (2008). Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification. *Applied Vegetation Science*, 11(4), 499–508.
- Domaç, A., & Süzen, M. (2006). Integration of environmental variables with satellite images in regional scale vegetation classification. *International Journal of Remote Sensing*, 27(7), 1329–1350.
- Editorial Committee of Vegetation Map of China, the Chinese Academy of Sciences. (2007). *the Vegetation Map of the People's Republic of China (1:1 000 000)*, Geological Publishing House.
- ESRI, Redlands, CA, USA. (2010). ArcGIS v.10.0.
- Exelis VIS, Boulder, CO, USA. (2013). ENVI v.5.1.
- Faber-Langendoen, D., Keeler-Wolf, T., Meidinger, D., Tart, D., Hoagland, B., Josse, C., et al. (2014). EcoVeg: a new approach to vegetation description and classification. *Ecological Monographs*, 84(4), 533–561. doi:10.1890/13-2334.1.
- Feng, Q., Su, Y., Hou, H., Zhang, Y., & Gao, H. (2014). Dataset of investigation of eco-hydrology transect in Heihe river basin. *Heihe Plan Science Data Center*. doi:10.3972/heihe.041.2014.db.
- Frank, T. D. (1988). Mapping dominant vegetation communities in the Colorado rocky-mountain front range with Landsat thematic mapper and digital terrain data. *Photogrammetric Engineering and Remote Sensing*, 54(12), 1727–1734.
- Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19(4), 474–499. doi:10.1177/030913339501900403.
- Franklin, J. (2010). *Mapping species distributions: spatial inference and prediction*. Cambridge: Cambridge University Press.
- Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., & Savenije, H. H. G. (2014). Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China. *Hydrology and Earth System Sciences*, 18(5), 1895–1915. doi:10.5194/hess-18-1895-2014.
- Gao, T., Zhu, J. J., Zheng, X., Shang, G. D., Huang, L. Y., & Wu, S. R. (2015). Mapping spatial distribution of larch plantations from multi-seasonal Landsat-8 OLI Imagery and multi-scale textures using random forests. *Remote Sensing*, 7(2), 1702–1720.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. doi:10.1016/j.patrec.2005.08.011.
- Gong, J., Xie, Y., Jia, Z., & Qian, D. (2014). Recent progress in land use and cover change in Heihe River Basin. *Journal of Lanzhou University (Natural Sciences)*, 50(3), 390–397. in Chinese.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. doi:10.1016/S0304-3800(00)00354-9.
- Han, T. (2002). Study on classification between coniferous forest and shrubbery based on TM data in the part of Qilian mountainous area. *Remote Sensing Technology and Application*, 17, 317–321 (in Chinese).
- Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122, 66–74. doi:10.1016/j.rse.2011.08.024.
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning* (Vol. 2, Vol. 1): Springer.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. doi:10.1002/joc.1276.
- Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195–213.
- Johansen, B., Karlsen, S., & Tømmervik, H. (2012). Vegetation mapping of Svalbard utilising Landsat TM/ETM+ data. *Polar Record*, 48, 47–63.
- Kaufman, Y., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 261–270.
- Landis, J., & Koch, G. (1977). The measurement of observed agreement for categorical data. *Biometrics*, 33, 159–174.
- Landmann, T., Piironen, R., Makori, D. M., Abdel-Rahman, E. M., Makau, S., Pellikka, P., et al. (2015). Application of hyperspectral remote sensing for flower mapping in African savannas. *Remote Sensing of Environment*, 166, 50–60. doi:10.1016/j.rse.2015.06.006.
- Loveland, T. R., & Dwyer, J. L. (2012). Landsat: building a strong future. [Article]. *Remote Sensing of Environment*, 122, 22–29. doi:10.1016/j.rse.2011.09.022.
- Mark, A. F., Dickinson, K. J. M., & Hofstede, R. G. M. (2000). Alpine vegetation, plant distribution, life forms, and environments in a perhumid New Zealand region: oceanic and tropical high mountain affinities. *Arctic Antarctic and Alpine Research*, 32(3), 240–254. doi:10.2307/1552522.
- Marsett, R. C., Qi, J., Heilman, P., Biedenbender, S. H., Watson, M. C., Amer, S., et al. (2006). Remote sensing for grassland management in the arid southwest. *Rangeland Ecology & Management*, 59(5), 530–540. doi:10.2111/05-201r.1.
- Newell, C. L., & Leathwick, J. R. (2005). *Mapping Hurunui forest community distribution, using computer models*. Wellington: Department of Conservation.
- Ohmann, J. L., Gregory, M. J., & Roberts, H. M. (2014). Scale considerations for integrating forest inventory plot data and satellite image data for regional forest mapping. *Remote Sensing of Environment*, 151, 3–15.
- Oke, O. A., & Thompson, K. A. (2015). Distribution models for mountain plant species: the value of elevation. *Ecological Modelling*, 301, 72–77.
- Ouyang, Z., Zhang, L., Wu, B., Li, X., Xu, W., Xiao, Y., et al. (2015). An ecosystem classification system based on remote sensor information in China. *Acta Ecologica Sinica*, 35(2), 219–226 (in Chinese).



- Pearson, R. & Miller, L. (1972). Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie, Pawnee National Grasslands, Colorado. *Proceedings of the 8th International Symposium on Remote Sensing of the Environment II*: 1355–1379.
- Price, K. P., Guo, X., & Stiles, J. M. (2002). Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing*, 23(23), 5031–5042.
- Qin, J., Ding, Y. J., Wu, J. K., Gao, M. J., Yi, S. H., Zhao, C. C., et al. (2013). Understanding the impact of mountain landscapes on water balance in the upper Heihe River watershed in northwestern China. *Journal of Arid Land*, 5(3), 366–383. doi:10.1007/s40333-013-0162-2.
- Richards, J. A., & Richards, J. (1999). *Remote sensing digital image analysis* (Vol. 3): Springer.
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55, 95–107.
- Rouse, J. W., Haas, R. W., Schell, J. A., Deering, D. W., & Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (Greenwave effect) of natural vegetation*. Greenbelt: NASA/GSFCT Type III Final Report.
- Schmidt, K., & Skidmore, A. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85(1), 92–108.
- Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing*. USA: Academic press.
- Sesnie, S. E., Gessler, P. E., Finegan, B., & Thessler, S. (2008). Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment*, 112(5), 2145–2159. doi:10.1016/j.rse.2007.08.025.
- Sitch, S., Smith, B., Prentice, I. C., Armeth, A., Bondeau, A., Cramer, W., et al. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9(2), 161–185.
- Sluiter, R. (2005). *Mediterranean land cover change: modelling and monitoring natural vegetation using GIS and remote sensing*. Utrecht: Utrecht University.
- van Beijma, S., Comber, A., & Lamb, A. (2014). Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sensing of Environment*, 149, 118–129. doi:10.1016/j.rse.2014.04.010.
- van der Linden, S., Rabe, A., Held, M., Jakimow, B., Leitão, P., Okujeni, A., et al. (2015). The EnMAP-Box—a toolbox and application programming interface for EnMAP data processing. *Remote Sensing*, 7(9), 11249.
- van der Maarel, E., & Franklin, J. (2013). *Vegetation ecology*. Oxford: Wiley-Blackwell.
- van Deventer, A. P., Ward, A. D., Gowda, P. H., & Lyon, J. G. (1997). Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*, 63, 87–93.
- Wang, G. H., Zhou, G. S., Yang, L. M., & Li, Z. Q. (2003). Distribution, species diversity and life-form spectra of plant communities along an altitudinal gradient in the northern slopes of Qilianshan Mountains, Gansu, China. *Plant Ecology*, 165(2), 169–181. doi:10.1023/a:1022236115186.
- Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1(1), 9–23.
- Zhang, Z., De Clercq, E., Ou, X., De Wulf, R., & Verbeke, L. (2008). Mapping dominant vegetation communities at Meili Snow Mountain, Yunnan Province, China using satellite imagery and plant community data. *Geocarto International*, 23(2), 135–153.
- Zhang, Z., van Coillie, F., Ou, X., & de Wulf, R. (2014). Integration of satellite imagery, topography and human disturbance factors based on canonical correspondence analysis ordination for mountain vegetation mapping: a case study in Yunnan, China. *Remote Sensing*, 6(2), 1026–1056.
- Zhao, C. Y., Nan, Z. R., & Cheng, G. D. (2005). Methods for modelling of temporal and spatial distribution of air temperature at landscape scale in the southern Qilian mountains, China. *Ecological Modelling*, 189(1–2), 209–220. doi:10.1016/j.ecolmodel.2005.03.016.
- Zhao, C., Nan, Z., Cheng, G., Zhang, J., & Feng, Z. (2006). GIS-assisted modelling of the spatial distribution of Qinghai spruce (*Picea crassifolia*) in the Qilian Mountains, northwestern China based on biophysical parameters. *Ecological Modelling*, 191(3–4), 487–500. doi:10.1016/j.ecolmodel.2005.05.018.
- Zimmermann, N. E., & Kienast, F. (1999). Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, 10(4), 469–482. doi:10.2307/3237182.