

## Disentangling fractional vegetation cover: Regression-based unmixing of simulated spaceborne imaging spectroscopy data



Sam Cooper<sup>a,\*</sup>, Akpona Okujeni<sup>a</sup>, Clemens Jänicke<sup>a</sup>, Matthew Clark<sup>b</sup>, Sebastian van der Linden<sup>c,d</sup>, Patrick Hostert<sup>a,d</sup>

<sup>a</sup> Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin 10099, Germany

<sup>b</sup> Center for Interdisciplinary Geospatial Analysis, Sonoma State University, Rohnert Park, CA 94928, USA

<sup>c</sup> Institute for Geography and Geology, Universität Greifswald, Domstraße 11, Greifswald 17489, Germany

<sup>d</sup> Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin 10099, Germany

### ABSTRACT

The next generation of spaceborne imaging spectrometers will enable hyperspectral analysis of vegetation cover across large spatial extents. Spectral unmixing provides a means to assess subpixel vegetation composition in such imagery. Here we implement a regression-based unmixing approach to generate fractional vegetation cover on a regional scale from a simulated Environmental Mapping and Analysis Program (EnMAP) satellite scene derived from Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) imagery acquired over the San Francisco Bay Area, California, USA, an area with a mixture of temperate and Mediterranean climate forests, woodlands and shrublands. A hierarchical classification scheme was implemented that considered fractional cover of vegetation as a whole (vegetation vs non-vegetation), vegetation life forms (woody vs non-woody vegetation; tree vs shrub vs grass), and tree leaf type (needleleaf vs broadleaf). A Gaussian Process Regression (GPR) model was trained using synthetically-mixed training data generated from an endmember library, and mapping accuracy was assessed using an independent validation dataset across four ecoregions. Our approach was able to effectively model landscape patterns at all levels of the class hierarchy. Site-wide map accuracy was highest when mapping generic vegetation fractions (MAE = 3.8%) and expectedly decreased at more complex hierarchy levels, with highest errors observed when separating tree and shrub fractions. Still, fraction estimates of needleleaf trees (MAE = 10.6%), broadleaf trees (MAE = 13.1%) and shrubs (MAE = 15.3%) were mapped with low overall error. Using Landsat imagery led to an average decrease in map accuracy of 1.9% when compared to hyperspectral image analysis and a maximum decrease of 3.5% when separating broadleaf and needleleaf trees across all sites. Further, a single regional model was shown to yield comparable results to multiple local ecoregion-based models, facilitating the analysis of large regions without creating a separate model for each region. Our results highlight the utility of regression-based approaches for quantitative vegetation mapping, which is of particular interest for future spaceborne imaging spectroscopy missions operating across large areas at moderate spatial resolution.

### 1. Introduction

The spatial characterization of vegetation composition is important for a multitude of different ecosystem assessments, in particular regarding the effects of natural and anthropogenic disturbances on ecosystems (Müller et al., 2019; Seidl et al., 2017). Vegetation cover is an important parameter for modeling carbon and nutrient fluxes (Still et al., 2003; Xu-Ri and Prentice, 2008), mapping wildlife habitat and distributions (Coops et al., 2009; Kerr et al., 2001), modeling and valuing ecosystem services (Andrew et al., 2015), or monitoring structural changes in ecosystems, e.g. related to changing land management (Leitao et al., 2015) or climate (Kelly and Goulden, 2008). From a global change and policy perspective, the mapping of vegetation composition is also important for deriving improved Essential Biodiversity Variables (Pereira et al., 2013) and Essential Climate Variables

(Bojinski et al., 2014).

Optical remote sensing provides an important means to map and monitor vegetation from local to global scales by utilizing vegetation reflectance properties to distinguish between different vegetation types (Turner et al., 2003). While multispectral systems utilize reflectance information in only a few broad spectral bands, imaging spectrometers, or hyperspectral sensors, produce images that capture spectral information in hundreds of narrow and contiguous bands from the visible to shortwave infrared regions of the electromagnetic spectrum. This increased spectral resolution is better able to capture subtle changes in spectral properties related to physical properties of materials in general, as well as spectral properties relating to vegetation biochemistry and canopy structure in particular (Asner et al., 2015; Verrelst et al., 2012). This increased sensitivity can be critical in distinguishing between spectrally similar vegetation types (Somers and Asner, 2013).

\* Corresponding author.

E-mail address: [sam.cooper@geo.hu-berlin.de](mailto:sam.cooper@geo.hu-berlin.de) (S. Cooper).

Forthcoming or recently launched spaceborne imaging spectroscopy missions such as the Environmental Monitoring and Analysis Program (EnMAP; [Guanter et al., 2015](#)), the PRRecursore IperSpettrale della Missione Applicativa (PRISMA; [Loizzo et al., 2018](#)), or the DLR Earth Sensing Imaging Spectrometer (DESiS; [Müller et al., 2019](#)), as well as proposed operationalized missions such as the Surface Biology and Geology mission (SBG; [Green, 2018](#); [National Academies of Sciences and Medicine, 2018](#)) or the Copernicus Hyperspectral Imaging Mission (CHIME; [Nieke and Rast, 2018](#)), will enable consistent and repeated collection of hyperspectral imagery across broad geographic extents. This will allow researchers to map surface characteristics relating to specific absorption bands, i.e. fine-scale spectral differences across the Earth's surface. This ability to distinguish between unique spectral characteristics allows for the description of vegetation biophysical parameters, such as plant pigments, nutrients or water content ([Blackburn, 2006](#); [Casas et al., 2014](#); [Singh et al., 2015](#)). Further, studies based on airborne hyperspectral imagery have demonstrated that the increased spectral information provided by hyperspectral imagery allows different vegetation types or species to be reliably mapped ([Boggs et al., 2019](#); [Roth et al., 2016](#); [Roth et al., 2015](#)).

While the next generation of spaceborne imaging spectrometers will offer new opportunities for large area hyperspectral image analysis, it will also present challenges to the imaging spectroscopy community. Future spaceborne missions will have moderate spatial resolution, often 30 m ground sampling distance, which is coarser than imagery used in many past hyperspectral applications which have largely relied upon airborne imagery with finer spatial resolution. Coarser spatial resolution amplifies the extent of spectral mixing, i.e. with increased pixels size the number of pixels containing multiple vegetation and non-vegetation cover types increases. This imposes new challenges regardless of a spectrally-improved sensor type. This is especially challenging in the case of ecosystems where a mosaic of different vegetation types prevails over homogenous patches of a single class, e.g. in open-canopy shrublands or woodlands ([Suess et al., 2018](#)). Even in areas of homogenous vegetation cover, patchiness and ecosystem transitions lead to mixed pixels and difficult classification cases.

Through spectral unmixing of multiple vegetation types on a per pixel basis, fractional maps of vegetation can be estimated which provide detailed sub-pixel information that is missed by conventional classification methods. This added information can be critical when trying to quantify and describe vegetation ecosystems utilizing these cover maps ([Moody and Woodcock, 1995](#)), e.g. the distribution of habitat ([Baumann et al., 2018](#); [Macchi et al., 2019](#)), vegetation change processes ([Leitao et al., 2015](#)), or quantifying tree cover and forest loss ([Hansen et al., 2003](#)). Common methods for sub-pixel fractional cover mapping are often based on linear spectral mixture analysis (SMA; [Adams et al., 1986](#)) and its variants. Multiple Endmember Spectral Mixture Analysis (MESMA; [Roberts et al., 1998](#)) and Automated Monte-Carlo Unmixing (AutoMCU; [Asner and Lobell, 2000](#)) are two commonly used SMA variants used to effectively account for the variability of cover types by means of a large number of endmembers and end-member combinations using multiple linear mixing models. We refer the reader to ([Somers et al., 2011](#)) for an overview of different linear spectral unmixing concepts and applications.

Regression-based unmixing provides an attractive alternative to linear SMA approaches as it allows for sub-pixel fractional cover mapping by means of a single globally applicable model per class which accounts for the variability of cover types. Further, this approach enables the use of powerful machine learning techniques which can be used to efficiently handle complex datasets, including noisy and highly correlated variables - a common challenge when using many adjacent spectral bands of hyperspectral imagery. Furthermore, kernel-based machine learning approaches such as Support Vector Regression (SVR) or Gaussian Process Regression (GPR) can adopt flexible kernel functions to solve non-linear problems ([Scholkopf and Smola, 2001](#); [Williams and Rasmussen, 2006](#)). However, to effectively train these

regression models, quantitative training data representing the complete range of spectral variability and mixing is required. In the context of spectral unmixing, this means that the training spectra for regression modeling must not only represent the pure endmembers, but also the spectral variation resulting from mixed pixels.

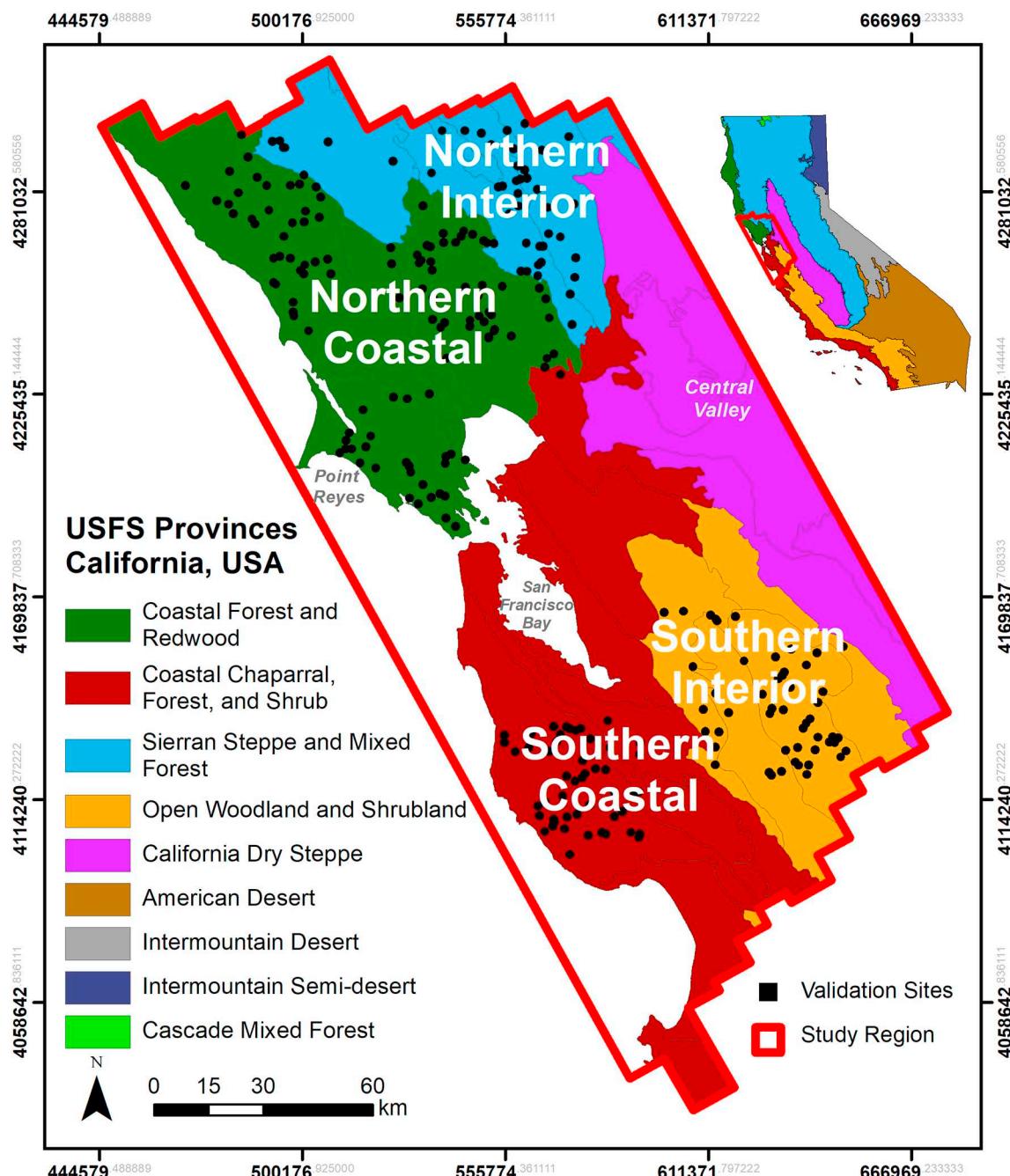
Synthetic mixing of pure endmembers has been proposed as a means to efficiently generate spectral mixtures needed for regression-based unmixing of remotely sensed imagery ([Okujeni et al., 2013](#)). Using this methodology, pure endmember spectra are linearly mixed using pre-defined fraction assignments. The resulting spectrum and associated fraction label represent a linear mixture of the input spectra. This enables the creation of large and diverse training data representing a range of possible mixed signals present in an image without having to manually relate image spectra from mixed pixels to mixing fractions calculated from high resolution land cover data, a process which can be time consuming and geographically limited due to data availability.

Synthetic mixing of pure endmember libraries in conjunction with regression-based unmixing has been demonstrated for land-cover fraction mapping in urban environments using both hyperspectral and multispectral images ([Mitraka et al., 2016](#); [Okujeni et al., 2018](#); [Schug et al., 2018](#)). Recently, this approach has been demonstrated as feasible for deriving a time-series of shrub cover fraction maps in a Mediterranean shrubland ecosystem using multi-annual Landsat images ([Suess et al., 2018](#)). However, it has not been demonstrated to what extent such a methodology is capable of separating different vegetation types over diverse ecosystems and for large areas. Specifically, it is unclear how spectral similarity or differences in structure among different vegetation types affect regression-based unmixing.

This study was therefore designed to explore the potential of future spaceborne hyperspectral data together with regression-based unmixing for quantitative mapping of vegetation in natural and semi-natural ecosystems at a regional scale. We quantified vegetation cover fractions across four ecoregions in Northern California, United States, using simulated EnMAP imagery. Mapping was done on a hierarchical basis, considering vegetation as a whole (vegetation vs non-vegetation), vegetation life forms (woody vs non-woody vegetation; tree vs shrub vs grass), and tree leaf type (needleleaf vs broadleaf). These classes allow comparison to existing studies that also unmixed one or more of these components and provide relevant sub-pixel information regarding ecosystem composition on a regional scale. Such quantitative assessments of vegetation growth forms and leaf types is a practical means for broad-scale vegetation mapping that can capture ecosystem characteristics and functioning without relying on discrete mapping of specific vegetation types ([Ustin and Gamon, 2010](#); [Defries et al., 1995](#)).

The overarching objectives of this paper are therefore to: (I) investigate the ability of regression-based unmixing to map vegetation class fractions across geographic gradients as compared to localized, ecoregion-based regression models, and (II) assess the capability of 30-m spaceborne hyperspectral imagery from the forthcoming EnMAP mission for mapping different levels of vegetation cover fractions as compared to multispectral data. To accomplish this, we implemented a regression-based unmixing workflow based on synthetically-mixed training data and compared results from four local models (based on libraries for each ecoregion) to a single regional model (based on a combined library for all ecoregions). We will accordingly address the following research questions:

- 1) To what extent can fractions of broad vegetation cover, vegetation life forms and tree leaf types in natural ecosystems be disentangled using hyperspectral imagery as provided by forthcoming spaceborne missions?
- 2) Can vegetation fractional cover models be generalized to allow for regional-scale vegetation analysis?
- 3) Does hyperspectral data improve fractional vegetation mapping efforts when compared to conventional multispectral sensors with the same spatial resolution and if so, to what extent?



**Fig. 1.** Overview of the study site showing US Forest Service Provinces, here referred to as ecoregions. The red polygon indicates the extent of the image data used in this study. Black points represent validation sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 2. Study area

This study focused on the San Francisco Bay Area in Northern California, USA (Fig. 1). The study area covers approximately 30,000 km<sup>2</sup> in a predominantly Mediterranean climate, with a wetter temperate climate along the coast. Mean annual precipitation for the region is 689 mm, with most precipitation falling between November and March followed by a prolonged dry period in summer. Water availability varies with precipitation patterns, topography, and underlying soils and is a key driver of vegetation composition, structure and function. This leads to a complex mosaic of vegetation types at very fine spatial scales.

When working at regional or broader spatial scales, these complex

patterns can be difficult to ascertain. To better assess related gradients in vegetation cover, the study region has been subdivided into four ecoregions based on the US Forest Service ecological sections (Bailey, 2004). These ecoregions are delineated by common climate, geology, and potential vegetation types.

The Coastal Forest and Redwood section along with the Coastal Chaparral Forest and Shrub section, hereafter referred to as Northern Coastal and Southern Coastal, respectively, are located along the Pacific coast. Needleleaf and evergreen broadleaf forests are common along the coast due to the presence of a cool and moist marine layer in the relatively dry summer months, giving these areas a temperate climate. These ecoregions also have interspersed shrubs and grasses, particularly on moisture limited southern slopes and exposed hilltops. The Southern

and Northern Interior ecoregions, designated as the Sierran Steppe and Mixed Forest section in the north and the Open Woodland and Shrubland section in the south, represent a transition in vegetation communities to a Mediterranean climate as one moves inland. Pure conifer stands transition to mixed and deciduous forests before giving way to open woodlands, shrublands, and finally the grass dominated interior mountains and central valley. These transitions are largely driven by moisture availability as the coastal marine effect becomes more limited towards the east. A precipitation-based moisture gradient from north (wet) to south (dry) is also reflected in the vegetation communities, though to a lesser degree than the east-west gradient.

Intensive land use in the study area includes large and densely populated urban areas, as well as perennial and annual agriculture. As the focus of this study is on delineating vegetation types in natural and semi-natural terrestrial ecosystems, these human-dominated systems were not considered in the context of this study. The Central Valley (referred to in the USFS description as ‘California Dry Steppe’) is predominantly composed of agricultural and urban areas and was therefore not investigated in this study.

### 3. Materials and methods

#### 3.1. Hyperspectral image acquisition and processing

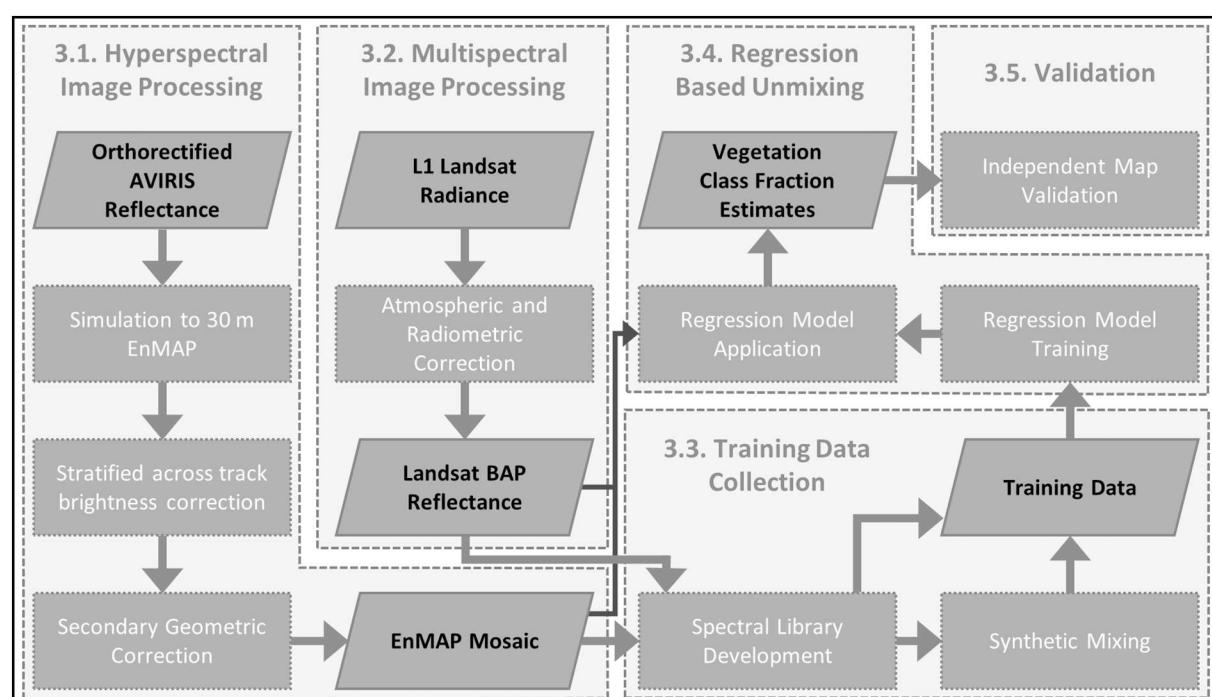
A generalized workflow for this study is provided in Fig. 2. Imagery was acquired as a part of the HypsIPI preparatory Science campaign (Lee et al., 2015). We used eleven Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) Classic images acquired on 7 June 2013. Flight lines were on average 240 km long with a swath width of approximately 12 km and a mean pixel size of 16.4 m. Overlap between adjacent flight lines was approximately 20% with a maximum view angle of  $\pm 17^\circ$ . Images were acquired between 10:37 and 15:21 local time, which together with the relatively large view angle typical of airborne platforms resulted in a large range of sun-sensor geometries across the imagery. Orthorectified AVIRIS reflectance images were obtained from the NASA Jet Propulsion Laboratory (JPL; <http://aviris.jpl.nasa.gov>). Surface reflectance was retrieved by JPL using a modified version of the

ATmospheric REMoval program (ATREM; Thompson et al., 2015).

Each AVIRIS reflectance image was converted into a 30 m EnMAP scene using the EnMAP end-to-end simulator (EeteS; Segl et al., 2012b). EnMAP will be a global sampling imaging spectroscopy mission. Spectral acquisitions will range from 420 to 2450 nm with spectral resolutions of 6.5 nm in the VNIR and 10 nm in the SWIR (Guanter et al., 2015). With a 30 m spatial resolution and 34 km swath width, a 34 degree pointing capability can potentially reduce the 24 day revisit period to 4 days. In total 242 EnMAP bands were simulated from the AVIRIS imagery using the EeteS (Segl et al., 2012a). Bands near strong atmospheric water absorption regions (1311–1465 nm and 1783–2044 nm) were removed, as were five bands in the NIR (934–952 nm) with poor reflectance retrievals. After band removal, 195 simulated EnMAP bands were used in this study.

Given the range of image acquisition times and sun-sensor geometries between flight lines, varying across track brightness gradients were observed in the imagery. Many operational reflectance retrieval processing chains for airborne hyperspectral data do not correct for such view-angle dependent artifacts, which can be especially problematic for the relatively large view angles and inconsistent sun-sensor geometries typical of airborne data and should therefore be corrected prior to analysis (Kennedy et al., 1997; Schiefer et al., 2006). To correct for this, a class-wise empirical across-track brightness correction approach following Schiefer et al. (2006) was implemented. We refer to Jänicke et al. (2020) for a detailed description of the procedure as applied to the simulated EnMAP imagery from the Bay Area. In short, images were classified into green vegetation, non-photosynthetically active vegetation and non-vegetation using empirically derived thresholds of the Normalized Difference Vegetation Index (NDVI) and Plant Senescent Reflectance Index (PSRI; Merzlyak et al., 1999) for each flight line. Reflectance for each pixel was then normalized to nadir view by fitting a quadratic model to each band across the range of view angles for each vegetation stratum and dividing the reflectance of the pixel by the appropriate correction factor (Jänicke et al., 2020).

While initial orthorectification was carried out by JPL, georeferencing errors of up to 40 m have previously been reported with these datasets (Clark and Kilham, 2016). A secondary geometric correction



**Fig. 2.** Generalized workflow for this study. Spectral library development and regression-based unmixing were carried out in parallel for both simulated EnMAP and Landsat imagery.

was therefore applied to each flight line by co-registering the simulated EnMAP flight lines to contemporaneous Landsat imagery. Tie points were automatically generated using the Automated and Robust Open-Source Image Co-Registration Software (AROSICS; Scheffler et al., 2017). Estimated local shifts ranged from less than 3 m to greater than 150 m. The EnMAP imagery was corrected using a third-degree polynomial and bilinear resampling, and remaining residual shifts were observed only at subpixel levels. The corrected images were then mosaicked into a single image, with western flight lines overlaying eastern flight lines. This dataset is freely available online via the German Research Center for the Geosciences (<http://dataservices.gfz-potsdam.de>, Cooper et al., 2020).

A water mask was built by manually selecting thresholds from the Normalized Difference Water Index (NDWI; Gao, 1996) and NDVI to visually match waterbodies present. As cloud cover was sparse and largely limited to coastal zones, clouds were manually masked from the simulated EnMAP imagery. The 2011 National Land Cover Database cover maps was then used to mask urban and agricultural areas from the imagery.

### 3.2. Multispectral image acquisition and processing

The value of spaceborne hyperspectral data for the unmixing of vegetation classes was tested against Landsat imagery. Landsat 8 OLI Collection 1 Level 1 Tier 1 images with cloud cover less than 80% were downloaded from the USGS. Cloud and shadow masking and radiometric correction (i.e. atmospheric, topographic, BRDF, and adjacency effect correction) were carried out using version 2.0 of the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE, Frantz et al., 2016). A best available pixel composite was created using the FORCE framework for June 2013 using all Landsat 8 imagery available between May 23 and June 22 (i.e.,  $\pm$  15 days from AVIRIS image acquisition). The final Landsat BAP composite contained pixels with acquisition dates ranging from May 28 to June 21, with 59% of its pixels from June 3 (close to the AVIRIS acquisition on June 7), and 15% and 17% of its pixels from June 13 and 19, mostly in the northwest corner of the box. The water and NLCD urban and agricultural masks described in 3.1 were then applied to the Landsat composites.

### 3.3. Training data collection

#### 3.3.1. Spectral library development

We developed a comprehensive image spectral library for the study region which was structured into four hierarchical levels (Table 1). Level 1 was the most generic and separated the image into vegetation and non-vegetation fractions. Level 2 further separated the Level 1 vegetation class into broad life forms (woody vs non-woody vegetation) while Level 3 separated the woody vegetation class into trees and shrubs. Finally, Level 4 considered tree leaf type, separating Level 3 trees into needleleaf and broadleaf tree categories.

Spectra were collected at Level 4 of the hierarchy. High resolution Google Earth imagery was used to identify areas of homogenous vegetation cover at the maximum density observed in the study region. Delineated homogenous areas ranged from 1 to 10 EnMAP pixels, and the mean spectrum was extracted from each area. For vegetation types associated with open canopies, a small portion of their signal may be influenced by other cover types (e.g. understory vegetation or soil), despite the assumption that these spectra represent pure cover

**Table 2**

Composition of the image spectral library. For each level, only newly differentiated classes are displayed (see Table 1). Ecoregions are labelled as (NC) Northern Coastal, (SC) Southern Coastal, (NI) Northern Interior, and (SI) Southern Interior.

	Class	Number of Spectra				
		NC	SC	NI	SI	All
Level 1	Vegetation	464	556	284	329	1633
	Non-veg	33	86	2	12	133
Level 2	Woody Veg	277	408	219	167	1071
	Non-woody Veg	187	148	65	162	562
Level 3	Tree	164	264	78	68	574
	Shrub	113	144	141	99	497
Level 4	Needleleaf	95	149	11	2	257
	Broadleaf	69	115	67	66	317

fractions. This is a common limitation in studies using image spectral libraries for fraction mapping (Quintano et al., 2017; Suess et al., 2018).

Libraries for vegetation classes higher in the hierarchy were created by combining the appropriate libraries into the new hierarchical class level, i.e., the broadleaf and needleleaf libraries were combined to create the Level 3 tree library, the tree and shrub libraries were combined to create the Level 2 woody vegetation library, and the woody and non-woody vegetation libraries were combined to create the Level 1 vegetation library. Spectral libraries were generated independently for each of the four ecoregions in the study area.

A total of 1764 pure spectra were collected throughout the study area (Table 2; Fig. 1). The libraries were collected to incorporate the full range of spectral diversity within each class, resulting in a high degree of endmember variability within each class (Fig. 3). Spectral similarity of class means between ecoregions is the result of the gradual transition of vegetation communities between these areas and similar communities found in each ecoregion. However, subtle differences in spectral means for each class are apparent when comparing different ecoregions. This is particularly evident in non-woody vegetation, which at the time of acquisition was largely senesced non-photosynthetic vegetation in the more water limited interior ecoregions but green grasses and forbs along the coast.

#### 3.3.2. Synthetic mixing

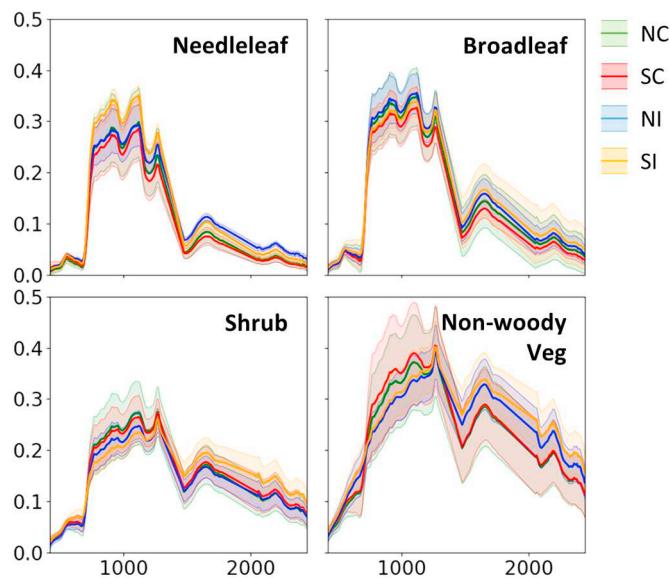
To create suitable training data for regression analysis from our library of pure endmembers, we implemented a synthetic mixing approach (Okujeni et al., 2013). This approach has been thoroughly tested for regression-based unmixing in an urban case (Okujeni et al., 2017), and we followed recommendations outlined therein. Synthetic mixing was conducted in a class-wise and randomized manner, where for each target class to be modeled we generated 2000 synthetic spectral mixtures and the associated mixing fraction. Each synthetic mixture was randomly generated using the following process:

- 1) The number of endmembers to be in the mixture was decided. This decision was weighted such that there was a 50% chance for a two endmember mixture, 40% for a three endmember mixture, and 10% for a four endmember mixture.
- 2) A random endmember was drawn from the target class.
- 3) The selected endmember was assigned a random mixing fraction

**Table 1**

Hierarchical vegetation classes.

Level 1			Vegetation			Non-Vegetation
Level 2						Non-Vegetation
Level 3		Woody Vegetation				Non-Vegetation
Level 4	Trees		Shrubs			Non-Vegetation
	Needleleaf	Broadleaf	Shrubs			Non-Vegetation



**Fig. 3.** Mean and standard deviation reflectance for Level 4 vegetation class spectra. Colors indicate ecoregion from which endmembers were drawn. Green: northern coastal, red: southern coastal, blue: northern interior, orange: southern interior. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between 0 and 1

- 4) For the remaining spectra in the mixture, endmembers were drawn randomly from the remaining endmembers.
- 5) Mixing fractions were assigned to each new endmember such that the sum of all assigned mixing fractions summed to one.
- 6) The selected endmembers were then linearly combined using the corresponding mixing fractions to produce the synthetically-mixed spectrum.

This process was repeated for each of the 2000 synthetic spectra generated. The pure library endmembers were then added to the synthetically-mixed training data and assigned mixing fractions of either 1 or 0 for spectra belonging to target and non-target classes, respectively.

#### 3.4. Regression-based Unmixing

A machine learning regression-based unmixing approach using Gaussian Process Regression (GPR) was implemented to map vegetation class fractions. GPR is a Bayesian-based learning algorithm, which provides a probabilistic framework for learning regression problems (Williams and Rasmussen, 2006). Recently, GPR has been used in the context of remote sensing in utilizing spectral information for mapping leaf chlorophyll content and other foliar traits (Verrelst et al., 2013; Wang et al., 2019), Leaf Area Index (Verrelst et al., 2012), soil salinity and moisture (Roelofsen et al., 2014), and disease detection (Ashourloo et al., 2016).

Our regression-based unmixing workflow was carried out using Python 3.6.0 and the EnMAP-Box 3.0 API (van der Linden et al., 2015). The EnMAP-box is an open-source QGIS plugin designed for advanced processing workflows of optical remote sensing data. The Scikit-learn v 0.20.0 (Pedregosa et al., 2011) implementation of GPR was used for regression mapping using the synthetically-mixed training data. To keep the sample size low, but still include a greater variability of mixed spectra and associated mixing fractions into the regression-based unmixing procedure, an ensemble based approach was implemented (Okujeni et al., 2017). Final fraction estimates were obtained using the mean of the ten regression estimates in the ensemble.

To investigate the ability to utilize this methodology across large

and diverse ecoregions, we generated four local regression models based only on spectral libraries derived from a single ecoregion (Fig. 1), as well as a regional model which combined the spectral libraries from all ecoregions in the study site. In this way, we investigate whether model generalization negatively affects the regression results due to the introduction of mixtures that would not realistically be present in the scene.

#### 3.5. Validation

A total of 260 validation sites were randomly selected within pre-defined areas of the four ecoregions dominated by natural vegetation cover (Fig. 1). Polygons consisting of  $3 \times 3$  pixel blocks were used at each validation site, and Level 4 class fractions for each pixel in the polygons were estimated independently by two experts through visual interpretation of high resolution Google Earth imagery. The two independent estimates were then averaged, and discrepancies greater than 25% between the two estimates were reviewed and appropriate fractions were agreed upon. The average mixing complexity of the validation data increases with higher class levels and regional differences are reflected in the compositional changes in vegetation cover between each ecoregion (Table 3). Mean cover fractions for each validation polygon were calculated to mitigate remaining sub-pixel spatial misregistration, as such errors will be proportionally less when simultaneously assessing multiple adjacent pixels (Stehman and Wickham, 2011).

We used the Mean Absolute Error (MAE) between the reference fractions and predicted fractions of these  $3 \times 3$  pixel validation sites as the primary measure of accuracy for the produced maps, as the MAE directly relates to the magnitude of model errors without weighting outlying residuals (Willmott and Matsuura, 2005). The Root Mean Squared Error (RMSE) is also reported to facilitate comparison to studies which favor this error metric. We additionally calculate the relative MAE as the MAE divided by the range of class fractions present in the site. In this way, the MAE is penalized when validation sites fail to encompass the full range of possible fractions (Xian et al., 2015). We further calculated the  $R^2$  and the slope and intercept of a fitted Ordinary Least Squares (OLS) regression line to assess the model fit and systematic biases, i.e. over-and underestimation of fractions.

## 4. Results

### 4.1. Regional vs local regression model performance

A general overview of the accuracy of the regional and local models are presented as the Mean Absolute Error (MAE) between predicted and reference cover fractions (Table 4), while a more detailed assessment of model performance of each class level can be found in the subsequent sections. MAEs for the regional model ranged from 1.7% to 21.8% and patterns of error varied by ecoregion, with the interior ecoregions generally having higher errors than the coastal ecoregions. Lowest errors were observed at Level 1 of the vegetation class hierarchy and in general error increased through deeper hierarchical levels. However, the largest errors were observed in Level 3 when distinguishing between tree and shrub cover for the two interior ecoregions. Highest

**Table 3**

Average and standard deviation of number of classes per pixel in the validation sites organized by ecoregion and mapping hierarchy.

	SI	NC	SC	NI	ALL
Level 1	$1.6 \pm 0.5$	$1.3 \pm 0.4$	$1.3 \pm 0.4$	$1.4 \pm 0.5$	$1.4 \pm 0.5$
Level 2	$2.4 \pm 0.6$	$1.7 \pm 0.7$	$1.6 \pm 0.7$	$2.0 \pm 0.8$	$1.9 \pm 0.8$
Level 3	$2.8 \pm 0.8$	$1.9 \pm 0.9$	$1.7 \pm 0.8$	$2.4 \pm 0.9$	$2.1 \pm 1.0$
Level 4	$2.9 \pm 0.9$	$2.2 \pm 1.0$	$2.0 \pm 1.0$	$2.6 \pm 1.0$	$2.4 \pm 1.0$

**Table 4**

Mean Absolute Error (MAE, %) of observed vs reference fractional estimates for each vegetation class by ecoregion and for all ecoregions. The regional model as applied to each ecoregion is shown in the left table and local models applied locally displayed in the right table. Ecoregions are labelled as NC: Northern Coastal, SC: Southern Coastal, NI: Northern Interior, and SI: Southern Interior.

		Regional					Local				
		NC	SC	NI	SI	All	NC	SC	NI	SI	All
Level 1	Vegetation	3.3	2.1	5.6	4.4	3.8	3.3	1.7	7.6	5.3	4.3
	Non-veg	3.4	1.7	5.8	8.4	4.6	4.4	2.2	6.9	6.5	5.0
Level 2	Woody Veg	5.7	4.9	10.9	13.6	8.2	5.6	5.7	11.0	16.0	10.2
	Non-woody Veg	7.6	4.5	6.8	9.9	7.4	7.0	5.1	7.3	10.4	7.5
Level 3	Tree	9.3	7.4	9.7	11.3	9.5	10.2	7.1	10.0	13.0	8.9
	Shrub	13.8	8.4	17.2	21.8	15.3	15.5	10.2	16.9	20.3	15.9
Level 4	Needleleaf	12.3	16.6	8.1	4.7	10.6	13.9	16.6	10.7	11.2	13.2
	Broadleaf	14.5	13.2	10.4	12.5	13.1	15.3	14.1	10.5	13.2	13.7

MAE (%)

errors in the coastal ecoregions were observed in Level 4. When assessing all validation sites in the study area together, similar patterns of error within the class hierarchy were observed, with the highest average error observed in shrubs followed by leaf type differentiation.

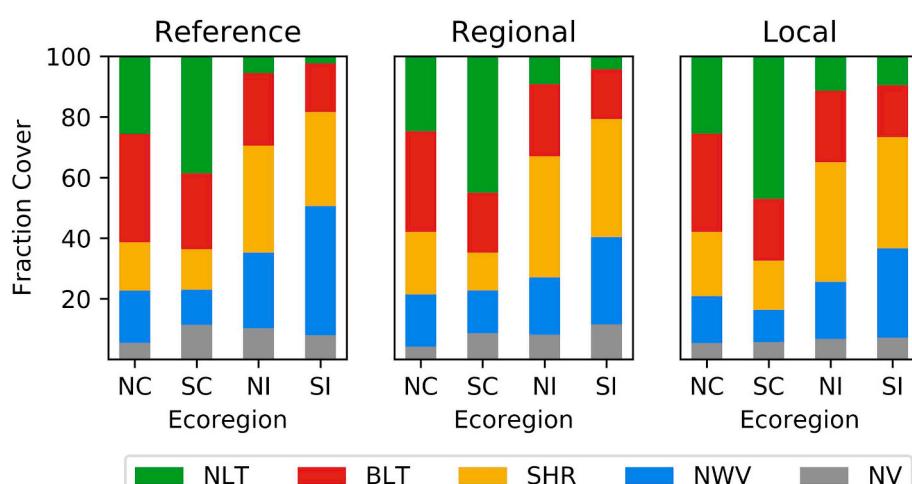
These patterns in model error reflect the distribution of cover across the study region, with the coastal ecoregion containing more needleleaf and broadleaf forests, while the interior ecoregions contained more open-canopy shrubland and woodland (Fig. 4). Overall, the distribution of the fractions match the reference data collected, and similar trends in vegetation composition can be seen among different ecoregions.

Comparisons to the local regression models indicated very similar model performance throughout classes and ecoregions (Table 4). The average absolute deviations from the regional model was 1.1%, and neither the regional nor local models consistently performed better across the study site. Further, trends in average fractional cover were similar to the regional model and reference fractions, though in general deviations from reference sites were slightly larger in the local model (Fig. 4). Because model accuracy between local and regional models was similar, the use of a generic regional library for regression training was deemed suitable, and the remainder of this section will therefore mostly focus on the regional models at each level, particularly when discussing the fraction maps.

#### 4.2. Level 1: Fractional vegetation cover

Distinguishing between vegetation and non-vegetation in Level 1 was highly accurate, with the lowest observed MAE in the classification hierarchy (Table 4). MAEs ranged from 1.7 to 8.4%, which was likely influenced by the high proportion of low residuals for high vegetation and low non-vegetation fractions. However, model slope and intercept should be carefully considered due to the sparsity of mixed and non-vegetated validation pixels (Fig. 5). This is most evident in the two northern ecoregions which showed significant deviation from the 1:1 line resulting from a slight overestimation of moderate vegetation cover fractions. This behavior was mirrored by the non-vegetation fraction estimates in these ecoregions.

Level 1 fraction maps by the regional model provided valuable insights into the spatial patterns of fractional vegetation cover across the study site (Fig. 6). Consistently high vegetation cover fractions were observed throughout the region. The primary areas dominated by non-vegetation cover were urban and agricultural land and were therefore masked out. The remaining non-vegetated areas in the study region consisted of natural soil and rock formations. These non-vegetated areas were mapped with a high degree of fidelity when compared to high resolution Google Earth imagery (Fig. 6A). Despite the low occurrence of non-vegetation dominated pixels, subpixel non-vegetation cover was observed throughout the study site. The fractional cover



**Fig. 4.** Mean fraction cover by ecoregion for Level 4 classes for reference polygons and corresponding estimates with the regional and local models. Ecoregion abbreviations: NC: Northern Coastal, SC: Southern Coastal, NI: Northern Interior, and SI: Southern Interior; Class abbreviations: NLT: Needle leaf tree, BLT: Broad leaf tree, SHR: Shrub, NWV: Non-woody vegetation, NV: Non-vegetation. Regression models are normalized to sum to one.

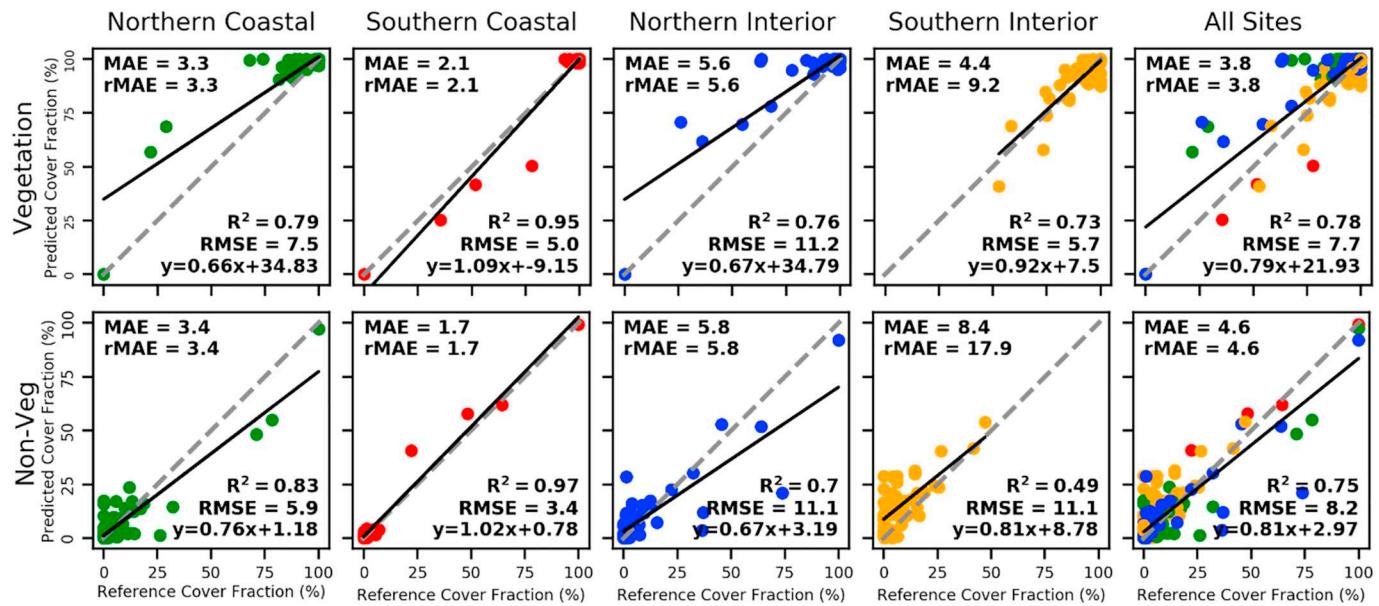


Fig. 5. Predicted (y-axis) vs reference (x-axis) fraction estimates for Level 1 classes by ecoregion and for all sites using the regional model. The black line shows the regression fit and the dotted line shows the 1:1 line.

information also allowed for the distinction of sparsely-vegetated grasslands with varying degrees of non-vegetation fractions, seen in bare and balding south facing slopes in a senesced grassland (Fig. 6B). Further, open-canopy ecosystems with non-vegetated understory could be clearly seen in the vegetation fraction maps (Fig. 6C). In general, larger subpixel non-vegetation fractions were observed in the interior ecoregions compared to the coastal ecoregions.

#### 4.3. Level 2: Fractional woody and non-woody vegetation cover

The differentiation between woody and non-woody vegetation was similarly effective (Fig. 7). All sites showed low error, with MAEs ranging from 4.6 to 13.6%. Biases were similarly low, with slopes and intercepts near 1 and 0, respectively. However, the two interior ecoregions showed a tendency to overestimate intermediate woody fractions and underestimate non-woody fractions. Additionally, these

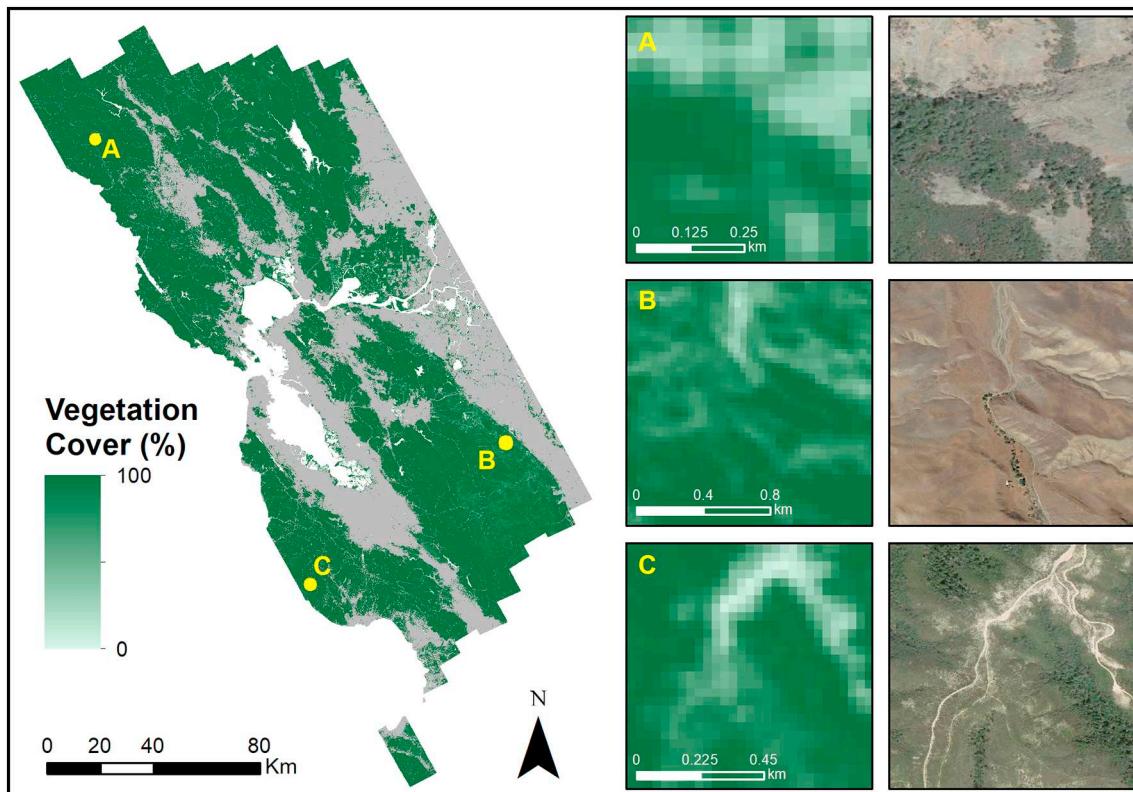
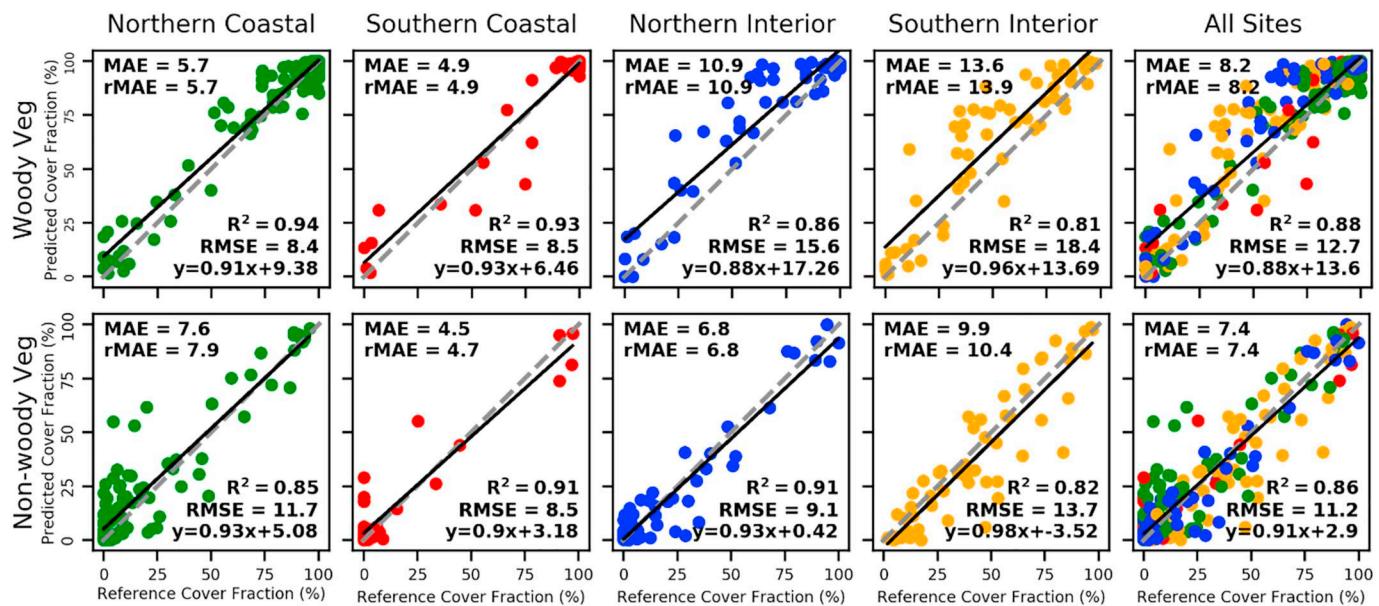


Fig. 6. Fractional vegetation cover for Level 1 showing general vegetation fractions derived from the regional model. Subset images accompanied by corresponding Google Earth imagery. Urban and agricultural lands are masked in grey.

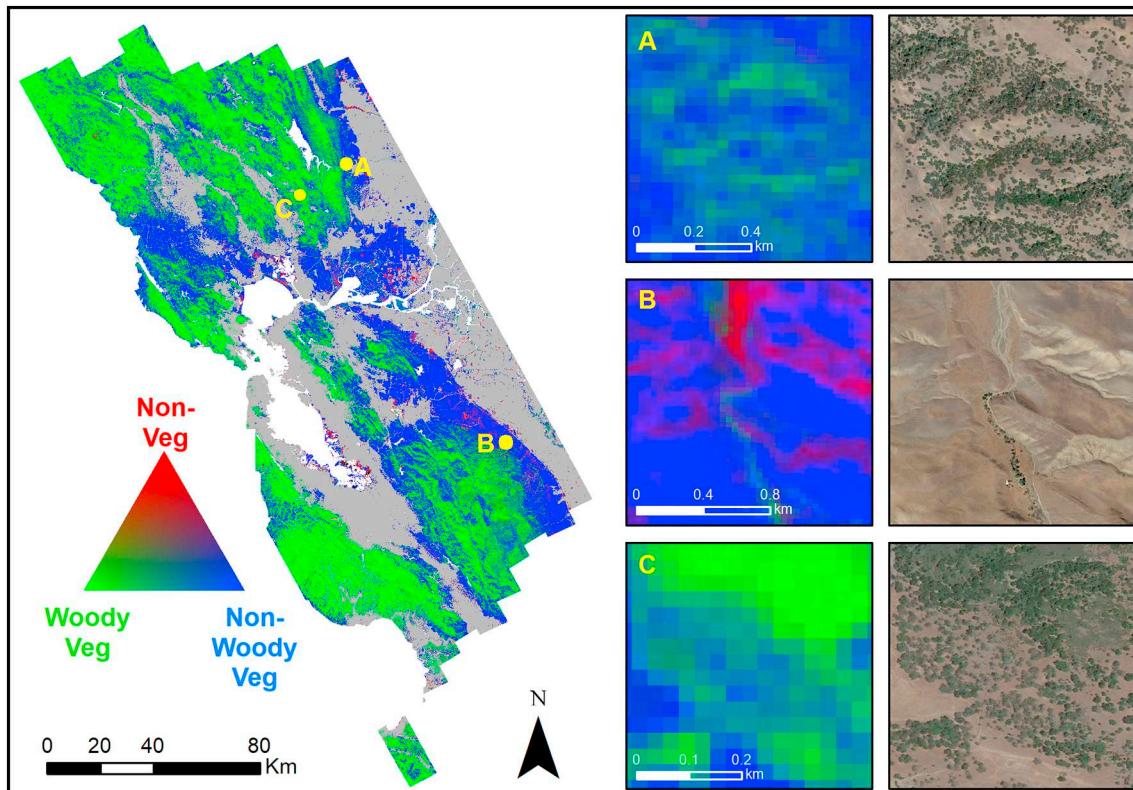


**Fig. 7.** Predicted (y-axis) vs reference (x-axis) fraction estimates for Level 2 classes by ecoregion using the regional model. The black line shows the regression fit and the dotted line shows the 1:1 line.

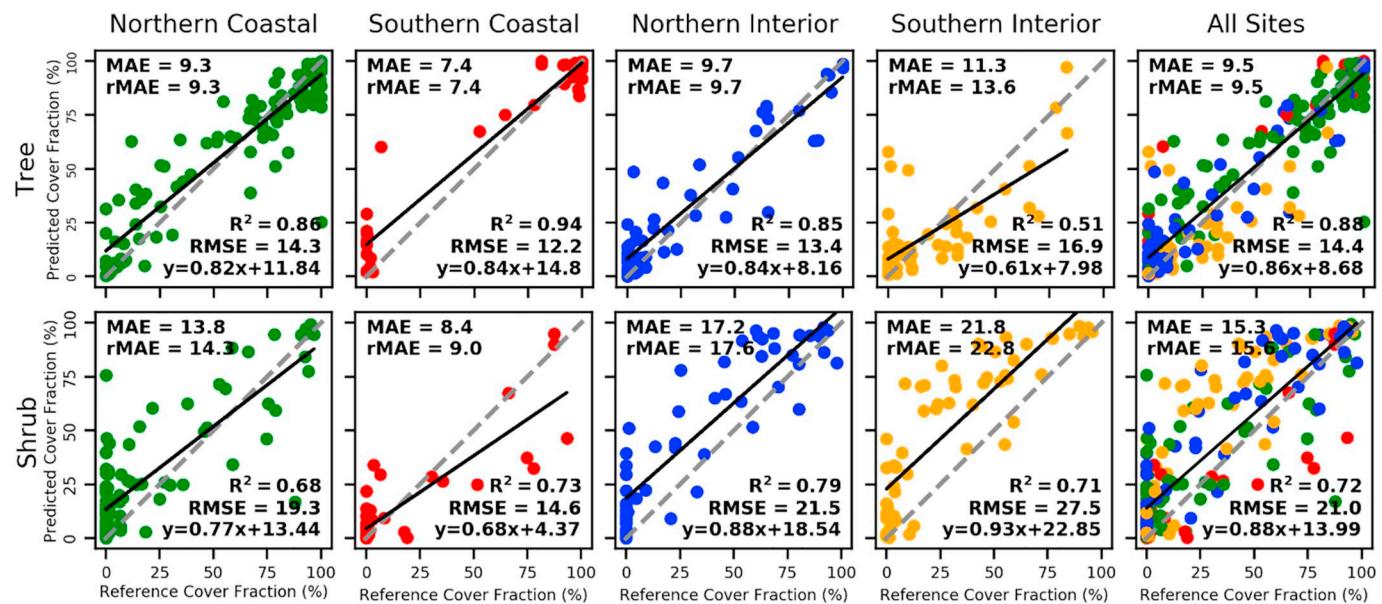
ecoregions also showed higher MAEs and lower  $R^2$  than the coastal ecoregions.

Level 2 class fraction maps at this level indicated dense woody vegetation dominating the two coastal ecoregions, transitioning into open-canopy woody vegetation cover inland (Fig. 8). Non-woody vegetation can be seen throughout, and dominates the western edges of the coastal ranges around the Central Valley, as well as many areas

around the San Francisco Bay and to the Pacific Coast north of Point Reyes National Seashore. Subpixel spatial detail at this level showed good agreement with high resolution imagery. This was observed in open-canopy woodlands containing many mixed pixels in which there was often no clearly dominant class (Fig. 8A, C). Further, subpixel features, e.g. riparian stands (Fig. 8B) were accurately depicted at this level.



**Fig. 8.** Level 2 regional model fractions showing non-vegetation (red), woody vegetation (green) and non-woody vegetation (blue). Subset images accompanied by corresponding Google Earth imagery. Urban and agricultural lands are masked in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 9.** Predicted (y-axis) vs reference (x-axis) fraction estimates for Level 3 classes by ecoregion using the regional model. The black line shows the regression fit and the dotted line shows the 1:1 line.

#### 4.4. Level 3: Fractional tree and shrub cover

The separation of trees and shrubs was effective for most ecoregions (Fig. 9). The two coastal ecoregions as well as the northern interior ecoregion showed low errors, with MAEs ranging from 7.4 to 13.8%. However, systematic biases can be seen in the overestimation of shrub and underestimation of tree cover in these areas. High shrub fractions tended to be underestimated in favor of tree cover in the Southern Coastal ecoregion, resulting in deviations from the 1:1 line. Further, the Northern Coastal ecoregion showed two opposing examples of shrub/tree confusion – one confusing nearly pure tree cover for shrub cover, and one doing the opposite. Both instances occurred in riparian zones.

The Southern ecoregions, while still effective in distinguishing between shrub and tree cover, showed more severe confusion between the two. MAEs here ranged from 9.7 to 21.8%, however, with the exception of tree cover in the Northern Interior ecoregion, significant deviations from the 1:1 line were observed. In addition to having the highest MAEs in the class hierarchy, shrub cover was consistently overestimated across the range of reference fractions for both ecoregions.

Landscape patterns of tree and shrub fractions were generally captured well, with dense tree cover in the west and increased shrub presence moving eastward (Fig. 10). The added value of subpixel cover information was again observed, for example, in open-canopy woodlands (Fig. 10A) or shrublands with sporadic tree cover (Fig. 10A, B). However, instances of overestimation of shrubs and underestimation of trees could also be seen (Fig. 10C). In this area, the model erroneously indicated a mixture of shrub and grass along stand edges while this was open-canopy oak woodlands with a grass background.

#### 4.5. Level 4: Fractional needleleaf and broadleaf cover

Needleleaf and broadleaf trees were generally modeled better in the coastal ecoregions than they were in the interior (Fig. 11). While the northern coastal ecoregion was modeled well, the southern coastal ecoregion showed an overestimation of intermediate needleleaf fractions and an underestimation of intermediate broadleaf fractions. The interior ecoregions suffered from a lack of high fractional validation data for needleleaf trees as these such species are relatively sparse in the interior ecoregions. Therefore validation model fit in these areas was quite poor as evidenced by low  $R^2$ . Further, broadleaf trees in the

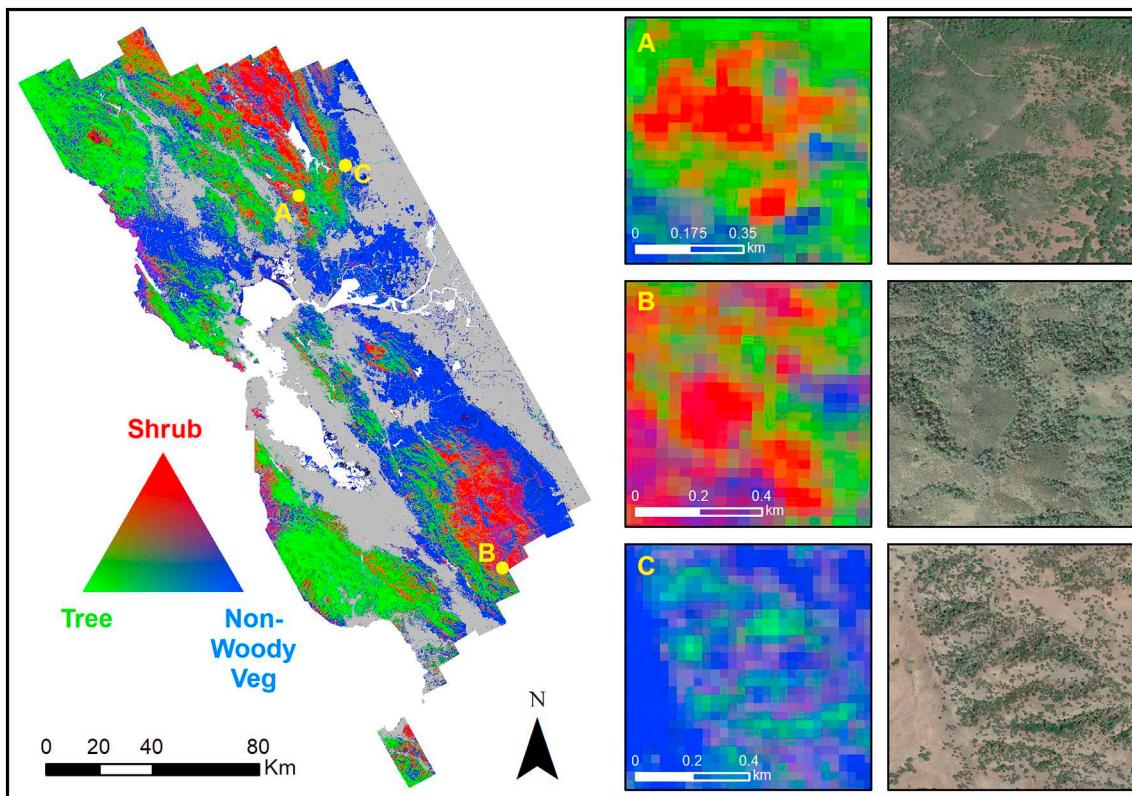
southern interior ecoregion showed overestimation of near-zero broadleaf fractions leading to higher errors.

Spatial patterns of needleleaf and broadleaf trees were captured quite well (Fig. 12). Dense needleleaf stands dominated the western coastal regions, gradually transitioning into mixed forests and eventually open-canopy broadleaf forests (Fig. 12). These transitional areas resulted in a heterogeneous landscape with high levels of mixing (Fig. 12A). At a stand level, the fractions map further demonstrated the ability to detect vegetation gradients at stand edges, showing dense needleleaf forests transitioning to broadleaf forest and finally non-forested systems (Fig. 12B, C).

#### 4.6. Multispectral unmixing

Regression-based unmixing using Landsat imagery was in general less effective than using simulated spaceborne hyperspectral data (Table 5). Across levels, Landsat results showed an increase of MAE of up to 6.3% with an average change in MAE of 2% greater than the hyperspectral analysis. Level 1 distinction of vegetation fractions showed a slight improvement over hyperspectral analysis ( $\Delta\text{MAE} = -0.5\%$ ). In general, EnMAP tended to perform better relative to Landsat in discriminating vegetation classes at increasing levels of the class hierarchy, with the exception of non-vegetation discrimination in the Southern Interior ecoregion. In contrast, Landsat-based woody vegetation discrimination performed better than the EnMAP model in the interior ecoregions, resulting in a similar performance in site-wide assessment ( $\Delta\text{MAE} = 0.2\%$ ).

We show a representative subset for Level 3 vegetation to compare Landsat results to the hyperspectral analysis (Fig. 13). Spatially, the Landsat imagery produced similar results to the EnMAP imagery. Overall vegetation patterns and transitions across the study region were depicted well, with generally subtle differences in the fraction values when compared to the EnMAP model, including a tendency to underestimate shrub cover in the Landsat based analysis. While the general patterns of lifeforms were represented well, each level tended to model intermediate fractions more poorly. This results in more distinct boundaries between vegetation classes, e.g., transitions between trees and shrubs, as well as less variation in canopy cover. Further, some mixed pixels showed an erroneous shift in cover fractions. For example, some areas containing isolated trees surrounded by shrub cover shifted



**Fig. 10.** Fractional vegetation cover for Level 3 regional model showing shrub (red) tree (green) and non-woody vegetation (blue). Subset images accompanied by corresponding Google Earth imagery. Urban and Agricultural lands are masked in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

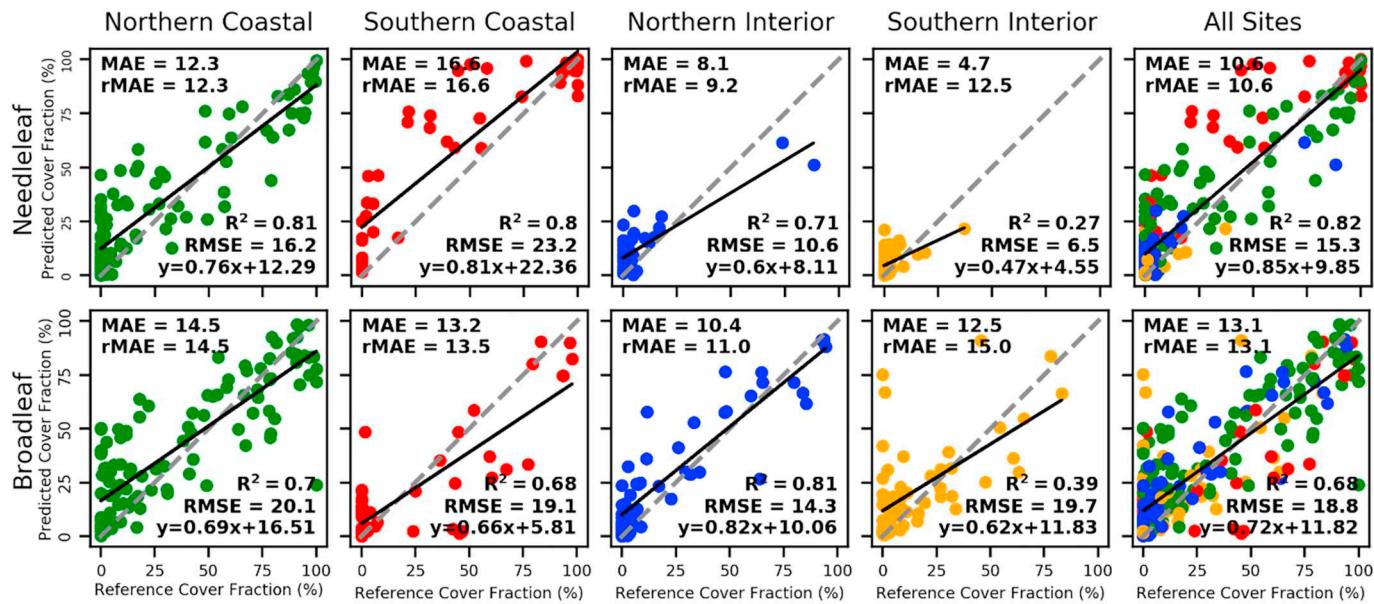
to being mapped as tree dominated in the Landsat-based analysis.

## 5. Discussion

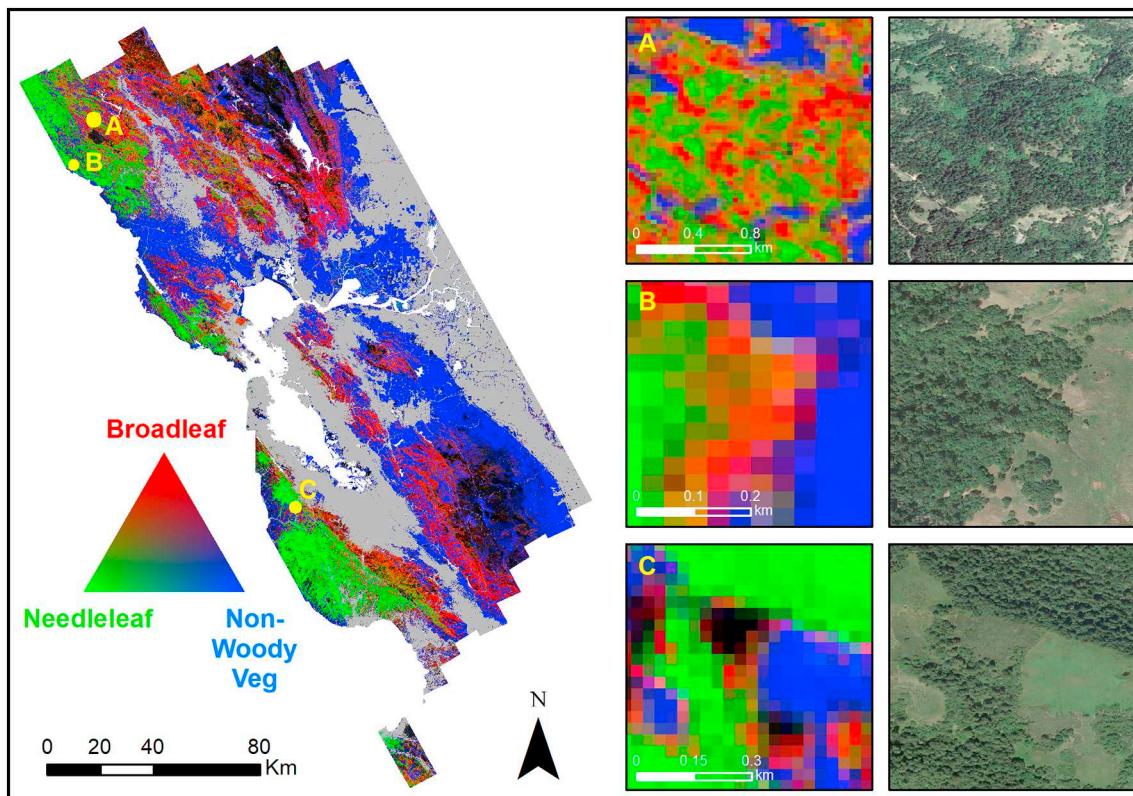
### 5.1. Unmixing of complex ecosystems

This study was designed to explore the potential of future

spaceborne hyperspectral data for mapping of vegetation types in natural and semi-natural ecosystems at a regional scale. To assess the limits of our methodology, we implemented a hierarchical classification scheme that was comprised of increasingly specific vegetation types, and therefore also increasing spectral similarities between the classes to separate. We found regression-based unmixing to be highly effective for differentiating vegetation types across the entire class hierarchy, both



**Fig. 11.** Predicted (y-axis) vs reference (x-axis) fraction estimates for Level 4 classes by ecoregion using the regional model. The black line shows the regression fit and the dotted line shows the 1:1 line.



**Fig. 12.** Fractional cover for Level 4 vegetation classes showing broadleaf trees (red) needleleaf trees (green) and non-woody vegetation (blue) derived from the regional model. Subset images accompanied by corresponding Google Earth imagery. Urban and agricultural lands are masked in grey. Shrub dominated pixels, as well as remaining non-vegetation dominated pixels, are black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in terms of independently evaluated validation sites as well as in our visual assessments of the resulting fraction maps.

Accuracies varied through the levels of vegetation hierarchy, with highest accuracies obtained with the Level 1 separation of vegetation (MAE = 3.8%) and non-vegetation (MAE = 4.6%) fractions. Our method performed similarly to other methods that unmixed vegetation fractions with hyperspectral data (e.g., Dennison et al., 2019), and indicates the high degree of separability between vegetation and non-vegetation. Visual interpretation of the fraction maps further showed a high degree of agreement with high resolution imagery and also indicated that our methodology enabled mapping of subtle vegetation

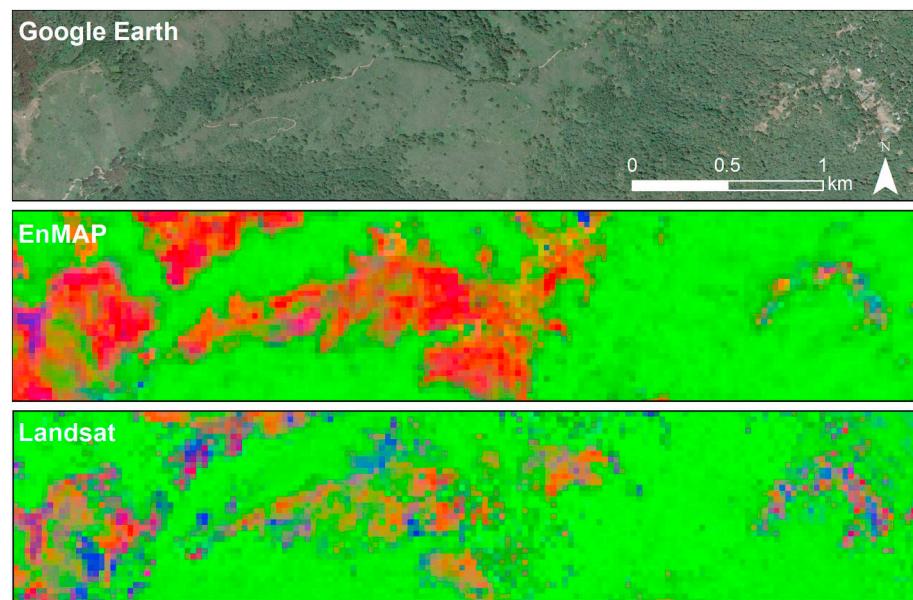
cover types, such as vegetation with a substantial sub-canopy soil cover or sparse vegetation cover. Unmixing at this thematically coarse level still revealed important details regarding sub-canopy vegetation cover, which can be useful in the assessment of ecosystem health (Carlson and Ripley, 1997) or when monitoring erosion in sparsely vegetated areas (Alejandro and Omasa, 2007).

With more detailed vegetation classes, the mixing complexity observed across the study site increased, and the utility of unmixing was readily apparent. Fractional mapping of Level 2 woody (MAE = 8.2%) and non-woody vegetation (MAE = 7.4%) was most effective, with the lowest MAEs of the resulting vegetation classes. Fractional mapping of

**Table 5**

Mean Absolute Error (MAE, %) of observed vs reference fractional estimates for each vegetation type by ecoregion using Landsat imagery for the regional regression model. Differences to the regional EnMAP results (Table 4) are presented on the right. Ecoregions are labelled as NC: Northern Coastal, SC: Southern Coastal, NI: Northern Interior, and SI: Southern Interior.

	Level 1	Real Landsat					Difference from EnMAP					Difference	
		NC	SC	NI	SI	All	NC	SC	NI	SI	All		
		Vegetation	3.0	1.4	4.4	4.4	3.3	-0.3	-0.7	-1.2	0.0	-0.5	
Level 2	Level 2	Non-veg	4.8	3.6	7.1	14.7	7.1	1.5	1.9	1.2	6.3	2.5	7
		Woody Veg	7.4	8.0	9.6	9.5	8.4	1.7	3.0	-1.3	-4.1	0.2	
Level 3	Level 3	Non-woody Veg	10.8	8.9	10.4	7.9	9.8	3.2	4.3	3.6	-2.0	2.4	-7
		Tree	12.2	11.9	12.6	12.0	12.2	2.9	4.6	2.9	0.7	2.7	
		Shrub	17.0	11.8	18.7	22.1	17.5	3.2	3.4	1.5	0.3	2.3	
Level 4	Level 4	Needleleaf	16.0	18.1	8.7	6.4	12.9	3.7	1.5	0.5	1.7	2.3	MAE (%)
		Broadleaf	20.6	19.5	11.2	11.1	16.6	6.1	6.3	0.8	-1.4	3.5	



**Fig. 13.** Fractional cover for Level 3 vegetation classes for an example subset in the Southern Coastal ecoregion for regional model using simulated EnMAP (middle) and Landsat (bottom). Shrubs are shown in red, trees in green and non-woody vegetation in blue. Google Earth imagery is on top. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

non-woody vegetation was effective across the entire study region, both at the landscape scale as well as pixel level. Similarly, woody vegetation was captured well, both as a non-dominant understory class as well as in grassland dominated systems. It is important to note though that the seasonality of these systems plays a role in the separability of different classes, as the June imagery provided a stark spectral contrast between senesced grasses comprising the non-woody vegetation class and green woody vegetation. In the more arid, interior Californian ecoregions, larger soil fractions were accurately depicted and allowed differentiating the nuances between ecoregions. This suggests that our regression-based unmixing approach is able to distinguish between non-photosynthetic vegetation (NPV) and non-vegetation fractions, i.e., bare soil. While differentiation between these two endmembers was evident in our results, quantitative validation was not possible at this detailed level as the fractions of green vegetation and NPV are temporally dynamic and requires reference data collection on site during overflight which was not available for our retrospective analyses. However, this distinction has previously been shown to be possible with hyperspectral image analysis in airborne data (Roberts et al., 1993), and our findings suggest that it will also be possible with regression-based unmixing on a regional scale from spaceborne data.

Level 3 trees (MAE = 9.5%) and shrubs (MAE = 15.3%) were also accurately modeled at a regional scale. Regional patterns of both woody vegetation types were depicted well, as were complex landscape level mosaics, such as elevation or terrain dependent differences in vegetation cover. Given the spectral similarities of these classes, challenges still remain. Complex canopy structures were difficult to model, as was apparent in open-canopy woodlands such as those typical of the interior ecoregions. In these systems, overstory trees and understory herbaceous vegetation and soil were highly mixed at fine spatial scales. When synthetically mixing these spectra for model training, however, confusion with other vegetation types may arise from the ambiguity of the mixed spectra. In our case, for example, oak and grass mixtures resembled shrub spectra, leading to poor model performance in areas dominated by such vegetation classes (Figs. 8 & 9B). In the Southern Interior ecoregion, the imperfect model fit is likely related to this confusion between mixed tree and shrub cover. This is underpinned by the overestimation of shrubs and underestimation of trees. However, the overestimation of shrub cover in the Northern Interior ecoregion was driven more by a confusion with background non-vegetation cover fractions (Figs. 5 & 9). Such difficulty in distinguishing between tree and shrub has been previously noted, and the incorporation of

structural information from radar or LIDAR data could facilitate the unmixing of these classes (Baumann et al., 2018; Sankey et al., 2017). However, while model accuracies for tree cover were comparable to previous studies (e.g. Senf et al., 2020; Baumann et al., 2018), our shrub model consistently underperformed other unmixing efforts with both hyperspectral (Schwieder et al., 2014; Suess et al., 2015) and multispectral data (Homer et al., 2012; Xian et al., 2015; Lippet et al., 2018), though these studies were largely focused on shrub dominated landscapes.

In the interior ecoregions, Level 4 discrimination of tree leaf types as assessed by MAE (broadleaf = 10.4 & 12.5%, needleleaf = 8.1 & 4.7%) was more effective than the Level 3 mapping of tree (9.7 & 11.3%) and shrub (17.2 & 21.8%) fractions (Table 4). This should be interpreted with caution, however, as the relative sparsity of needleleaf and broadleaf dominated pixels in the interior ecoregions reduces the reliability of this metric for validation. Here, the relative MAE (rMAE) provides a more appropriate error metric. For example, needleleaf predictions in the Southern Interior ecoregion had an MAE of 4.7%, which would suggest high model performance. However, fractions in the validation data have maximum coverage of about 25%, meaning that while the absolute error is small, the predictive capability of the model is limited. The rMAE effectively weights this metric, and the resulting rMAE of 12.5% is more in line with conclusions drawn from visual interpretation of the scatterplots, i.e. modeling performed similarly in these areas as compared to the coastal ecoregions. Overall we were able to effectively model these classes at Level 4, and our model accuracies performed similarly to other studies that unmixed needleleaf and broadleaf trees (Ju et al., 2003; Senf et al., 2020). On a landscape scale, transitions of needleleaf-dominated systems near the coast into mixed forest and broadleaf dominated systems in the interior are obvious from the analysis. This was also reflected at the sub-pixel level, with stand-level mixtures of needleleaf and broadleaf being captured well by the fraction maps.

In assessing the scatterplots for Levels 3 and 4, one major source of error is found in a general overestimation of low fractional cover (Figs. 9 & 11). This tendency to overestimate plots with low vegetation cover has been observed in other unmixing exercises (Ju et al., 2003; Schwieder et al., 2014; Senf et al., 2020), though much of the literature fails to report comprehensive assessments (e.g. actual vs observed scatterplots) with which we can make this comparison. Still, this can be seen as a limitation to our approach, and an indication of model confusion between the classes.

Our regression-based unmixing approach provided valuable information that would be otherwise missed with conventional, ‘hard’ classification approaches. In open-canopy systems in particular, such classifications would underestimate areas with target classes falling below 50% vegetation cover and overestimate areas with vegetation cover over 50%. This is of particular importance for quantitative studies such as biomass mapping or carbon sequestration assessment when gradual vegetation differences prevail, as in semi-arid to arid systems (Guerschman et al., 2009). Here, we show that under such conditions, using an image-based library is possible. However, our approach has limitations in open-canopy systems or in case of rare vegetation types being an important target class, as the collection of pure image spectra at 30 m spatial resolution can be difficult. In such cases, the use of a field-based library as an alternative or to complement the image spectral library will be beneficial (Okin et al., 2001).

Our first research question explored to what extent fractions of broad vegetation cover, vegetation life forms and tree leaf types in natural ecosystems can be disentangled using hyperspectral imagery as provided by forthcoming spaceborne missions. For this class scheme, we found that a regression-based unmixing approach in conjunction with synthetically-mixed training data was able to effectively map vegetation class fractions as compared to an independent reference dataset. Overall accuracies were encouraging and suggest that unmixing based on synthetically-mixed spectra holds potential for disentangling vegetation classes in Mediterranean woodland and temperate forest ecosystems.

## 5.2. Local vs regional modeling

The utility of hyperspectral data for regional studies across ecosystems has been explored with regard to the classification of vegetation and other land cover classes (Clark and Kilham, 2016; Clark, 2017), forest alliances (Clark et al., 2018), crop type (Shivers et al., 2018), and species (Roth et al., 2015). Large area fraction mapping of natural areas using hyperspectral unmixing is relatively understudied, with few studies either focusing on limited areas (e.g., Roberts et al., 2015), or other methodologies, such as using MESMA as a classifier (e.g., Roth et al., 2015; Dudley et al., 2015; Clark, 2017).

To explore how our approach generalizes across large areas for vegetation analyses, a regional regression approach incorporating training data from the entire study site was compared to a local, ecoregion-based regression approach. The regional regression method was shown to be just as effective as local, ecoregion-based modeling. Vegetation patterns for each class were similar throughout the study site, and model accuracies from independent validation data had an average absolute difference of only 1.1%. This indicated that a generic, regional approach was suitable for unmixing complex landscapes, and signified that the added complexity of the synthetic mixtures did not introduce additional spectral ambiguity into the model. The regional model was thus deemed suitable for analysis of a broad-scale study area, indicating that a generic approach across multiple ecoregions is comparable to more locally tuned and ecoregion-specific modeling.

The two interior ecoregions did show some systematic deviations between local and regional estimated fractions, with non-woody vegetation having lower fractional estimates for the regional models while non-vegetation estimates were higher. This could potentially relate to the inclusion of green grass spectra from the temperate, coastal ecoregions in the regional library. This would have increased spectral ambiguity between non-woody vegetation and other green vegetation classes, leading to NPV being misclassified as non-vegetation in the regional models. These discrepancies were minor, however, and were not captured in the validation data.

While the results presented here suggest the ability of a generic regression-based unmixing model to be applied over a large area, we do note that in terms of global biodiversity, the ecoregions in this study are all relatively similar, with the study region ranging from temperate to

Mediterranean climates. Further research is therefore needed into the transferability of this methodology to the full range of terrestrial biomes.

Our second research question asked if vegetation fractional cover models can be generalized to allow for large-area vegetation analysis. We found that using a more generic spectral library representative for multiple ecoregions was suitable for unmixing across the study site. This is a promising finding in the context of spaceborne hyperspectral sensors, as it supports the assumption that broad-scale studies assessing vegetation across different Mediterranean and temperate ecosystems around the globe will be feasible. This study therefore represents an important step in moving beyond local case studies hyperspectral analyses have typically been limited to due to the reliance on aerial acquisitions.

## 5.3. Simulated spaceborne hyperspectral vs multispectral imagery

The added value of hyperspectral imagery when compared to conventional multispectral imagery for vegetation mapping is an ongoing topic of research, and to the best of our knowledge fractional mapping of vegetation classes on a regional scale using hyperspectral imagery is largely untested. An exception is a recent study by Clark (2017), who used MESMA fractions from simulated spaceborne hyperspectral imagery (HypIRI) and Landsat 8 imagery, respectively, to map discrete land-cover classes in our Bay Area study region. With the same summer 2013 AVIRIS base data as used in our study, the Clark (2017) analysis found significant increases in classification accuracy with hyperspectral over multispectral bands. In light of these findings, we sought to better understand the added value of hyperspectral imagery for vegetation fraction mapping, by comparing simulated spaceborne hyperspectral imagery with multispectral imagery acquired by Landsat 8.

The general patterns of errors across ecoregions for each class was similar to those of the hyperspectral analysis, and MAEs were on average 1.9% higher with multispectral imagery when compared to hyperspectral imagery (Table 5). Furthermore, multispectral image analysis consistently underperformed on a site-wide basis apart from one exception. The Level 1 discrimination of generic vegetation fractions with Landsat performed similarly or slightly better than the hyperspectral analysis across all ecoregions. This suggests that the level of spectral detail needed to unmix vegetation and non-vegetation fractions is already satisfied by multispectral imagery, at least at a 30 m spatial resolution. Furthermore, potentially confounding factors resulting from the use of airborne data, in particular the spatial resampling from 16 m to 30 m, likely impacts the accuracy of the simulated EnMAP results. This could help explain why Landsat showed slight improvement in some of these classes where hyperspectral imagery has no clear benefit. Despite this, we see that the discrimination of non-vegetation is still less accurate with Landsat than with hyperspectral imagery, particularly in the Southern Interior ecoregion. This could result from confusion with NPV which dominated the non-woody vegetation class, particularly in the interior ecoregions. This would also explain why woody vegetation was readily separable, while non-woody vegetation showed mixed results when compared to hyperspectral imagery. These discrepancies further highlight the benefits of hyperspectral imagery for unmixing even generic classes, but more specifically in discriminating between NPV and non-vegetation.

Higher levels of vegetation complexity (i.e. Levels 3 and 4) clearly show that Landsat analysis underperformed the hyperspectral assessment. This can be seen with visual interpretation of the results (Fig. 13), where, for example, mixtures of shrub and tree cover were less accurately represented and confusion between the two was greater with multispectral analysis. The exception here was with the broadleaf class in the Southern Interior ecoregion, in which multispectral assessment showed slight improvements. As previously stated (5.1), this should be interpreted with caution, as needleleaf and broadleaf trees in the interior ecoregion were relatively sparse, and a bias towards low

fractional cover could explain the relatively low MAEs.

Our third research question asked if hyperspectral data improves vegetation mapping efforts when compared to conventional multispectral imagery. We found that hyperspectral analyses outperformed multispectral analyses in almost all cases. Hyperspectral imagery generally improves spectral separability of pure endmembers, as also demonstrated in other cases where vegetation type classification was the focus (Clark, 2017). Here, we additionally show that spectral mixtures of multiple cover types within a single pixel are more easily separable with hyperspectral imagery. Further, with high quality field data, hyperspectral analysis provides promising avenues to distinguishing more specific vegetation types, e.g., vegetation alliances (Clark et al., 2018) or species (Roth et al., 2015), and the inclusion of multi-temporal hyperspectral data could allow for even greater separability of vegetation classes, species or related traits (Lausch et al., 2016; Clark et al., 2018).

#### 5.4. Further considerations

In this study, we focused on assessing the potential for vegetation fraction mapping of various growth forms and leaf types with hyperspectral imagery. Previous studies have shown promising advances in the use of hyperspectral imagery for the mapping of more detailed or dominant vegetation classes at the pixel level, e.g. mixtures of diagnostic species in alliances (Clark et al., 2018) or high-cover species (Roth et al., 2015). To achieve these thematic levels of mapping at scales that spaceborne hyperspectral missions will enable requires reference information that is difficult to obtain for broad-scale studies. The ability to unmix vegetation classes at regional scales with our approach at the species or alliance level therefore remains unknown. Given this, the tradeoff between the hard classification of detailed vegetation types and the unmapping of thematically broader classes should be considered in future studies, and will vary with the intended application (Ustin and Gamon, 2010).

Time series analysis of multispectral imagery has proven its added value for distinguishing land cover types based on their phenological behavior (Clark, 2017; Hansen and Loveland, 2012; Clark, 2020). This can be critical for classes with a strong phenological signal (e.g., deciduous broadleaf forests or NPV). The recent and planned launches of numerous spaceborne hyperspectral sensors will greatly facilitate studies exploring the possible synergies between the spectral and temporal aspects of vegetation remote sensing, even if these scientific missions will not produce the temporal density of ongoing operational missions such as Landsat or Sentinel-2. However, as our single date hyperspectral imagery outperformed a single date multispectral image for fraction estimation, we can expect multi-temporal hyperspectral analyses having the potential to outperform multi-temporal multispectral approaches for vegetation mapping once spaceborne hyperspectral sensors begin to collect consistent multitemporal data. Initial studies looking at the classification of vegetation species with multi-temporal hyperspectral data have shown limited improvements or even slight decreases in accuracy when compared to single date assessments (Dudley et al., 2015; Meerdink et al., 2019). However, for forest alliance classification in our Bay Area study area, Clark et al. (2018) found a significant increase in overall accuracy with multi-temporal (spring, summer, fall) over summer-only hyperspectral data, while a parallel analysis found multi-temporal hyperspectral data were significantly more accurate compared to Landsat 8 and Sentinel-2 data (Clark, 2020). Imaging spectroscopy data acquired at ecologically relevant repeat times, i.e., at time steps that are able to capture the seasonal variability of vegetation types (Clark et al., 2018; Clark, 2020), will certainly offer novel analysis pathways for the unmapping of vegetation cover.

Spatial resolution of imagery also plays a key role in the unmapping of fractional cover, as it directly impacts the degree of mixing complexity of selected classes. For instance, Sentinel-2 now provides regular multispectral acquisitions at a 10 m spatial resolution. This will decrease

the number of mixed pixels in the scene compared to what is observed at 30 m (Table 3); however, the mixing of multiple land cover classes will still be prevalent to some degree. For example, in forest alliance classification, results indicate that higher spectral resolution of hyperspectral data is more important than Sentinel-2's high spatial resolution (Clark, 2020). In addition to the temporal component of data, more research is needed into the tradeoff between the higher spatial resolution of multispectral spaceborne sensors such as Sentinel-2 and the increased spectral resolution of spaceborne hyperspectral missions which will have a coarser spatial resolution.

LIDAR data has also been shown to be useful in distinguishing between vegetation classes, particularly those with distinct structural differences such as shrubs and trees. However, as with hyperspectral image acquisition, LIDAR acquisitions have typically been acquired from airborne platforms which limits their applicability in both space and time. Spaceborne missions such as the Global Ecosystem Dynamics Investigation (GEDI, Dubayah et al., 2014) will further open synergies with optical data sources, while their spatial resolution and footprint poses new challenges to analyzing such data.

## 6. Conclusions

In this study we demonstrated the ability of spaceborne imaging spectroscopy data in combination with regression-based unmixing for fractional mapping across of broad vegetation classes for a large area and across different ecoregions. As such, this study represents a step towards using and validating methods for hyperspectral image analysis on spaceborne imagery at broad spatial scales. This was done using simulated EnMAP data and a synthetically-mixed training dataset generated from an image endmember library.

Our regression-based spectral unmixing approach was able to capture fractional vegetation cover well for various classes of vegetation lifeforms and leaf types. Despite the complexity of the ecosystems present in the study area, a regional regression model was shown to be suitable for fractional vegetation cover, indicating the efficacy of a generic approach for large scale assessments. Unsurprisingly, overall accuracies in unmapping vegetation fractions depended on the spectral similarity and dissimilarity of the unmixed classes. Consequently, analysis of simulated EnMAP imagery outperformed Landsat based analyses, highlighting the utility of hyperspectral imagery for unmapping vegetation classes facilitated by the increased spectral separability of these classes as a result of increased spectral resolution.

The global scope of spaceborne imaging spectroscopy missions will enable hyperspectral analysis of diverse ecoregions. When moving beyond the Mediterranean and temperate ecosystems of California explored here, further research into the transferability of this methodology to the full range of terrestrial biomes will be needed. The addition of multi-temporal hyperspectral acquisitions is expected to enhance such hyperspectral unmapping efforts through the incorporation of phenological information (Dudley et al., 2015). Such enhanced information will allow for ascertaining both abrupt (e.g., fire) and subtle (e.g. climate-induced) changes in vegetation cover over time. Further, with high quality reference data, hyperspectral analysis provides promising avenues to distinguishing more specific vegetation types, e.g., vegetation alliances (Clark et al., 2018) or individual species (Roth et al., 2015). Such studies are critical for integrating remote sensing based analyses into international frameworks such as the definition of Essential Biodiversity Variables (Pereira et al., 2013). These frameworks seek to standardize ongoing efforts for monitoring environmental change of ecosystems around the world – a realm for which spaceborne remote sensing will play a key role (Pettorelli et al., 2018). In particular, through the use of spaceborne imaging spectroscopy, the methods presented here can be used to obtain high quality information about the vegetation composition of complex ecosystems across the globe. This will be especially important in regions which lack information regarding ecosystem composition and where spaceborne data

will be the only source of information.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was conducted as part of the EnMAP Core Science Team and was funded by the German Aerospace Centre Project Management Agency and granted by the German Ministry of Economic Affairs and Energy (BMWi; grant number 50EE1622). The authors would like to thank NASA Jet Propulsion Laboratory (JPL) for providing AVIRIS imagery, K. Segl from the German Research Centre (GFZ) for EnMAP simulation, and A. Rabe for assistance with algorithm development.

## References

- National Academies of Sciences, E., & Medicine, 2018. Thriving on our Changing Planet: A Decadal Strategy for Earth Observation from Space. National Academies Press.
- Adams, J.B., Smith, M.O., Johnson, P.E., 1986. Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site. *J. Geophys. Res.* 91, 8098–8112.
- Alejandro, M., Omasa, K., 2007. Estimation of vegetation parameter for modeling soil erosion using linear spectral mixture analysis of Landsat ETM data. *ISPRS J. Photogramm. Remote Sens.* 62, 309–324.
- Andrew, M.E., Wulder, M.A., Nelson, T.A., Coops, N.C., 2015. Spatial data, analysis approaches, and information needs for spatial ecosystem service assessments: a review. *GI Sci. Remote Sens.* 52, 344–373.
- Ashourloo, D., Aghighi, H., Matkan, A.A., Mobasher, M.R., Rad, A.M., 2016. An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 9, 4344–4351.
- Asner, G.P., Lobell, D.B., 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sens. Environ.* 74, 99–112.
- Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., 2015. Quantifying forest canopy traits: imaging spectroscopy versus field survey. *Remote Sens. Environ.* 158, 15–27.
- Bailey, R.G., 2004. Identifying ecoregion boundaries. *Environ. Manag.* 34, S14–S26.
- Baumann, M., Levers, C., Macchi, L., Bluhm, H., Waske, B., Gasparri, N.I., Kuemmerle, T., 2018. Mapping continuous fields of tree and shrub cover across the Gran Chaco using Landsat 8 and Sentinel-1 data. *Remote Sens. Environ.* 216, 201–211.
- Blackburn, G.A., 2006. Hyperspectral remote sensing of plant pigments. *J. Exp. Bot.* 58, 855–867.
- Bogan, S.A., Antonarakis, A.S., Moorcroft, P.R., 2019. Imaging spectrometry-derived estimates of regional ecosystem composition for the Sierra Nevada, California. *Remote Sens. Environ.* 228, 14–30.
- Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* 95, 1431–1443.
- Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* 62, 241–252.
- Casas, A., Riaño, D., Ustin, S.L., Dennison, P., Salas, J., 2014. Estimation of water-related biochemical and biophysical vegetation properties using multitemporal airborne hyperspectral data and its comparison to MODIS spectral response. *Remote Sens. Environ.* 148, 28–41.
- Clark, M.L., 2017. Comparison of simulated hyperspectral HypsIPI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping. *Remote Sens. Environ.* 200, 311–325.
- Clark, M.L., 2020. Comparison of multi-seasonal Landsat 8, Sentinel-2 and hyperspectral images for mapping forest alliances in Northern California. *ISPRS J. Photogramm. Remote Sens.* 159, 26–40.
- Clark, M.L., Kilham, N.E., 2016. Mapping of land cover in northern California with simulated hyperspectral satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 119, 228–245.
- Clark, M.L., Buck-Diaz, J., Evans, J., 2018. Mapping of forest alliances with simulated multi-seasonal hyperspectral satellite imagery. *Remote Sens. Environ.* 210, 490–507.
- Cooper, S., Okujeni, A., Jänicke, C., Segl, K., van der Linden, S., Hostert, P., 2020. 2013 Simulated EnMAP Mosaics for the San Francisco Bay Area USA. GFZ Data Services. Coops, N.C., Wulder, M.A., Iwanicka, D., 2009. Exploring the relative importance of satellite-derived descriptors of production, topography and land cover for predicting breeding bird species richness over Ontario, Canada. *Remote Sens. Environ.* 113, 668–679.
- DeFries, R.S., Field, C.B., Fung, I., Justice, C.O., Los, S., Matson, P.A., ... Sellers, P.J., 1995. Mapping the land surface for global atmosphere-biosphere models: Toward continuous distributions of vegetation's functional properties. *J. Geophys. Res.* 100 (D10), 20867–20882.
- Dennison, P.E., Qi, Y., Meerdink, S.K., Kokaly, R.F., Thompson, D.R., Daughtry, C.S., ... Numata, I., 2019. Comparison of methods for modeling fractional cover using simulated satellite hyperspectral imager spectra. *Remote Sensing* 11 (18), 2072.
- Dudley, K.L., Dennison, P.E., Roth, K.L., Roberts, D.A., Coates, A.R., 2015. A multi-temporal spectral library approach for mapping vegetation species across spatial and temporal phenological gradients. *Remote Sens. Environ.* 167, 121–134.
- Frantz, D., Röder, A., Stellmes, M., Hill, J., 2016. An operational radiometric landsat preprocessing framework for large-area time series applications. *IEEE Trans. Geosci. Remote Sens.* 54 (7), 3928–3943.
- Gao, B.C., 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266.
- Green, R.O., 2018. Global VSWIR imaging spectroscopy and the 2017 Decadal Survey. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 183–185.
- Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Muhle, H., Muller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitao, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H.Y., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP Spaceborne imaging spectroscopy Mission for earth observation. *Remote Sens.* 7, 8830–8857.
- Guerszman, J.P., Hill, M.J., Renzullo, L.J., Barrett, D.J., Marks, A.S., Botha, E.J., 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sens. Environ.* 113, 928–945.
- Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sens. Environ.* 122, 66–74.
- Hansen, M.C., DeFries, R.S., Townshend, J.R., Carroll, M., DiMiceli, C., Sohlberg, R.A., 2003. Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *Earth Interact.* 7, 1–15.
- Homer, C.G., Aldridge, C.L., Meyer, D.K., Schell, S.J., 2012. Multi-scale remote sensing sagebrush characterization with regression trees over Wyoming, USA: laying a foundation for monitoring. *Int. J. Appl. Earth Obs. Geoinf.* 14 (1), 233–244.
- Jänicke, C., Okujeni, A., Cooper, S., Clark, M., Hostert, P., van der Linden, S., 2020. Brightness gradient corrected hyperspectral image mosaics for mapping vegetation cover fractions in the San Francisco Bay Area. *Remote Sens. Lett.* 11 (1), 1–10.
- Ju, J., Kolaczyk, E., Gopal, S., 2003. Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing. *Remote Sens. of Environ.* 84 (4), 550–560.
- Kelly, A.E., Goulden, M.L., 2008. Rapid shifts in plant distribution with recent climate change. *Proc. Natl. Acad. Sci.* 105, 11823–11826.
- Kennedy, R.E., Cohen, W.B., Takao, G., 1997. Empirical methods to compensate for a view-angle-dependent brightness gradient in AVIRIS imagery. *Remote Sens. Environ.* 62 (3), 277–291.
- Kerr, J.T., Southwood, T., Cihlar, J., 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *Proc. Natl. Acad. Sci.* 98, 11365–11370.
- Lausch, A., Erasmi, S., King, D., Magdon, P., Heurich, M., 2016. Understanding forest health with remote sensing-part I—a review of spectral traits, processes and remote-sensing characteristics. *Remote Sens.* 8, 1029.
- Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J., Middleton, E.M., 2015. An introduction to the NASA Hyperspectral InfraRed imager (HypIPI) mission and preparatory activities. *Remote Sens. Environ.* 167, 6–19.
- Leitao, P.J., Schwieder, M., Suess, S., Okujeni, A., Galvao, L.S., van der Linden, S., Hostert, P., 2015. Monitoring natural ecosystem and ecological gradients: perspectives with EnMAP. *Remote Sens.* 7, 13098–13119.
- van der Linden, S., Rabe, A., Held, M., Jakimow, B., Leitao, P.J., Okujeni, A., Schwieder, M., Suess, S., Hostert, P., 2015. The EnMAP-box-A toolbox and application programming Interface for EnMAP data processing. *Remote Sens.* 7, 11249–11266.
- Loizzo, R., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetto, C., Varacalli, G., 2018. PRISMA: the Italian hyperspectral mission. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 75–178.
- Macchi, L., Baumann, M., Bluhm, H., Baker, M., Levers, C., Grau, H.R., Kuemmerle, T., 2019. Thresholds in forest bird communities along woody vegetation gradients in the South American Dry Chaco. *J. Appl. Ecol.* 56 (3), 629–639.
- Meerdink, S.K., Roberts, D.A., Roth, K.L., King, J.Y., Gader, P.D., Koltunov, A., 2019. Classifying California plant species temporally using airborne hyperspectral imagery. *Remote Sens. Environ.* 232, 111308.
- Merzlyak, M.N., Gitelson, A.A., Chikunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* 106, 135–141.
- Mitraka, Z., Del Frate, F., Carbone, F., 2016. Nonlinear spectral unmixing of landsat imagery for urban surface cover mapping. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 9, 3340–3350.
- Moody, A., Woodcock, C.E., 1995. The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landsc. Ecol.* 10, 363–379.
- Müller, J., Noss, R.F., Thorn, S., Bässler, C., Leverkus, A.B., Lindenmayer, D., 2019. Increasing disturbance demands new policies to conserve intact forest. *Conserv. Lett.* 12, e12449.
- Nieke, J., Rast, M., 2018. Towards the copernicus hyperspectral imaging mission for the environment (CHIME). In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 157–159.
- Okin, G.S., Roberts, D.A., Murray, B., Okin, W.J., 2001. Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sens. Environ.* 77, 212–225.
- Okujeni, A., van der Linden, S., Tits, L., Somers, B., Hostert, P., 2013. Support vector regression and synthetically mixed training data for quantifying urban land cover.

- Remote Sens. Environ. 137, 184–197.
- Okujeni, A., van der Linden, S., Suess, S., Hostert, P., 2017. Ensemble learning from synthetically mixed training data for quantifying urban land cover with support vector regression. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 10, 1640–1650.
- Okujeni, A., Canters, F., Cooper, S.D., Degerickx, J., Heiden, U., Hostert, P., Priem, F., Roberts, D.A., Somers, B., van der Linden, S., 2018. Generalizing machine learning regression models using multi-site spectral libraries for mapping vegetation-im-pervious-soil fractions across multiple cities. *Remote Sens. Environ.* 216, 482–496.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. *Science* 339, 277–278.
- Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queiroz, A.M., Keith, D., Wegmann, M., Schrott, F., Stellmes, M., Sonnenschein, R., Geller, G., Roy, S., Somers, B., Murray, N., Bland, L., Geijzendorffer, I., Kerr, J., Broszeit, S., Leitão, P., Duncan, C., Serafy, G., He, K., Blanchard, J., Lucas, R., Mairoti, P., Webb, T., Nicholson, E., 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation* 4 (2), 71–93.
- Quintano, C., Fernandez-Manso, A., Roberts, D.A., 2017. Burn severity mapping from Landsat MESMA fraction images and land surface temperature. *Remote Sens. Environ.* 190, 83–95.
- Roberts, D., Smith, M., Adams, J., 1993. Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sens. Environ.* 44, 255–269.
- Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., Green, R., 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sens. Environ.* 65, 267–279.
- Roberts, D.A., Dennison, P.E., Roth, K.L., Dudley, K., Hulley, G., 2015. Relationships between dominant plant species, fractional cover and land surface temperature in a Mediterranean ecosystem. *Remote Sens. Environ.* 167, 152–167.
- Roelofs, H.D., Kooistra, L., van Bodegom, P.M., Verrelst, J., Krol, J., Witte, J.-P.M., 2014. Mapping a priori defined plant associations using remotely sensed vegetation characteristics. *Remote Sens. Environ.* 140, 639–651.
- Roth, K.L., Casas, A., Huesca, M., Ustin, S.L., Alsina, M.M., Mathews, S.A., Whiting, M.L., 2016. Leaf spectral clusters as potential optical leaf functional types within California ecosystems. *Remote Sens. Environ.* 184, 229–246.
- Roth, K.L., Roberts, D.A., Dennison, P.E., Alonzo, M., Peterson, S.H., Beland, M., 2015. Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. *Remote Sens. Environ.* 167, 135–151.
- Sankey, T., Donager, J., McVay, J., Sankey, J.B., 2017. UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sens. Environ.* 195, 30–43.
- Scheffler, D., Hollstein, A., Diedrich, H., Segl, K., Hostert, P., 2017. AROSICS: an automated and robust open-source image co-registration software for multi-sensor satellite data. *Remote Sens.* 9, 676.
- Schiefer, S., Hostert, P., Damm, A., 2006. Correcting brightness gradients in hyperspectral data from urban areas. *Remote Sens. Environ.* 101 (1), 25–37.
- Scholkopf, B., Smola, A.J., 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond. MIT press.
- Schug, F., Okujeni, A., Hauer, J., Hostert, P., Nielsen, J.Ø., van der Linden, S., 2018. Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series. *Remote Sens. Environ.* 210, 217–228.
- Schwieder, M., Leitão, P.J., Suess, S., Senf, C., Hostert, P., 2014. Estimating fractional shrub cover using simulated EnMAP data: a comparison of three machine learning regression techniques. *Remote Sens.* 6 (4), 3427–3445.
- Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., Sang, B., Mogulsky, V., Hofer, S., 2012a. EeteS—the EnMAP end-to-end simulation tool. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 5, 522–530.
- Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., Sang, B., Mogulsky, V., Hofer, S., 2012b. EeteS—the EnMAP end-to-end simulation tool. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 5, 522–530.
- Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., 2017. Forest disturbances under climate change. *Nat. Clim. Chang.* 7, 395.
- Senf, C., Lastovicka, J., Okujeni, A., Heurich, M., van der Linden, S., 2020. A generalized regression-based unmixing model for mapping forest cover fractions throughout three decades of Landsat data. *Remote Sens. Environ.* 240.
- Shivers, S., Roberts, D., McFadden, J., Tague, C., 2018. Using imaging spectrometry to study changes in crop area in California's Central Valley during drought. *Remote Sens.* 10, 1556.
- Singh, A., Serbin, S.P., McNeil, B.E., Kingdon, C.C., Townsend, P.A., 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecol. Appl.* 25, 2180–2197.
- Somers, B., Asner, G.P., 2013. Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests. *Remote Sens. Environ.* 136, 14–27.
- Somers, B., Asner, G.P., Tits, L., Coppin, P., 2011. Endmember variability in spectral mixture analysis: a review. *Remote Sens. Environ.* 115, 1603–1616.
- Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment. *Remote Sens. Environ.* 115 (12), 3044–3055.
- Still, C.J., Berry, J.A., Collatz, G.J., DeFries, R.S., 2003. Global distribution of C3 and C4 vegetation: carbon cycle implications. *Global Biogeochem. Cycles* 17 6-1-6-14.
- Suess, S., Van der Linden, S., Okujeni, A., Leitão, P.J., Schwieder, M., Hostert, P., 2015. Using class probabilities to map gradual transitions in shrub vegetation from simulated EnMAP data. *Remote Sens.* 7 (8), 10668–10688.
- Suess, S., van der Linden, S., Okujeni, A., Griffiths, P., Leitão, P.J., Schwieder, M., Hostert, P., 2018. Characterizing 32 years of shrub cover dynamics in southern Portugal using annual Landsat composites and machine learning regression modeling. *Remote Sens. Environ.* 219, 353–364.
- Thompson, D.R., Gao, B.C., Green, R.O., Roberts, D.A., Dennison, P.E., Lundein, S.R., 2015. Atmospheric correction for global mapping spectroscopy: ATREM advances for the HypsIPI preparatory campaign. *Remote Sens. Environ.* 167, 64–77.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* 18, 306–314.
- Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. *New Phytol.* 186 (4), 795–816.
- Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., Moreno, J., 2012. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Trans. Geosci. Remote Sens.* 50, 1832–1843.
- Verrelst, J., Alonso, L., Caicedo, J.P.R., Moreno, J., Camps-Valls, G., 2013. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 6, 867–874.
- Wang, Z., Townsend, P.A., Schweiger, A.K., Couture, J.J., Singh, A., Hobbie, S.E., Cavender-Bares, J., 2019. Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. *Remote Sens. Environ.* 221, 405–416.
- Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning. MIT Press Cambridge, MA.
- Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* 30 (1), 79–82.
- Xian, G., Homer, C., Rigge, M., Shi, H., Meyer, D., 2015. Characterization of shrubland ecosystem components as continuous fields in the Northwest United States. *Remote Sens. Environ.* 168, 286–300.
- Xu-Ri, & Prentice, I.C., 2008. Terrestrial nitrogen cycle simulation with a dynamic global vegetation model. *Glob. Chang. Biol.* 14, 1745–1764.