

# Unmixing urban hyperspectral imagery using probability distributions to represent endmember variability

Yuan Zhou<sup>a,1,\*</sup>, Erin B. Wetherley<sup>b</sup>, Paul D. Gader<sup>a</sup>

<sup>a</sup>*Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA*

<sup>b</sup>*Department of Geography, University of California Santa Barbara, Santa Barbara, CA, USA*

---

## Abstract

Urban composition can be analyzed through spectral unmixing of images from airborne imaging spectrometers. Unmixing given a spectral library can be accomplished by set-based methods or distribution-based methods. For computational efficiency and optimal accuracy, set-based methods employ a library reduction procedure when applied to large spectral libraries. On the other hand, distribution-based methods model the library by only a few parameters, hence innately accept large libraries. A natural question arises that can distribution-based methods with the original large spectral library achieve comparable performance to set-based methods in urban imagery.

In this study, we aim to investigate the unmixing capability of several distribution-based methods, Gaussian mixture model (GMM), normal compositional model (NCM), and Beta compositional model (BCM) by comparing them to set-based methods MESMA and alternate angle minimization (AAM). The data for validation were collected by the AVIRIS sensor over the Santa Barbara region: two 16 m spatial resolution and two 4 m spatial resolution images. 64 validated regions of interest (ROI) (180 m by 180 m) were used to assess estimate accuracy. Ground truth was obtained using 1 m images leading to the following 6 classes: turfgrass, non-photosynthetic vegetation (NPV), paved, roof, soil, and tree. Spectral libraries were built by manually identifying and extracting pure spectra from both resolution images, resulting in 3,287 spectra at 16 m and 15,426 spectra at 4 m. The libraries were further reduced to 61 spectra at 16 m and 95 spectra at 4 m for set-based methods. The results show that in terms of mean absolute error (MAE), GMM performed best among the distribution-based methods while MESMA performed best among the set-based methods. For 16 m data, there is no significant difference between GMM and MESMA (MAE = 0.069 vs. MAE = 0.074,  $p = 0.25$ ). For 4 m data, though GMM is not as accurate as MESMA (MAE = 0.056 vs. MAE = 0.046,  $p = 7e-5$ ), it is better than AAM (MAE = 0.056 vs. MAE = 0.065,  $p = 0.02$ ) which is a re-implementation of MESMA. Further evidence on a reconstructed synthetic dataset implies possible overfitting of the reduced library to the images for MESMA. These findings suggest that the distribution-based method GMM could achieve comparable unmixing accuracy to set-based methods without the need of library reduction, it may also be more stable across datasets, and the current 2-step workflow could be replaced by a single model in applying a universal spectral library<sup>1</sup>.

*Keywords:* Spectral mixture analysis, MESMA, AVIRIS, GMM, NCM

---

## 1. Introduction

Global climate change and urbanization will bring many challenges to our environment. Urban areas are a major source of greenhouse gas emissions. To facilitate monitoring, management, development of urban environment, land cover information is a prerequisite and used in many physical models describing the urban ecosystem, such as climate (De Ridder et al., 2015), energy and water flux (Wang et al., 1996). It is also

---

\*Corresponding author

Email addresses: zhouyuanzxcv@gmail.com (Yuan Zhou), wetherley@umail.ucsb.edu (Erin B. Wetherley), pgader@cise.ufl.edu (Paul D. Gader)

<sup>1</sup>Present address: Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA

<sup>1</sup>The data and code of this research are available online (<https://github.com/zhouyuanzxcv/Hyperspectral>).

associated with urban heat island intensity (Zhou et al., 2017a), which in turn impacts the urban resident health. Urban land cover fractions can be estimated from airborne or spaceborne hyperspectral imagery (Van der Linden et al., 2019). Compared to color/panchromatic imagery that shows similar reflectance between different urban materials, imaging spectroscopy can measure the reflectance at narrow bands covering visible, near infrared, and short wave infrared (VSWIR) range, hence can differentiate materials with more subtle details (Herold et al., 2004). Applying high spatial and spectral resolution imagery for mapping urban vegetation and surfaces has been performed before (Alonzo et al., 2013). However, accurate mapping through classifying the acquired pixels requires a spatially fine-scale image, which is expensive to collect since it requires the airborne imaging spectrometer to fly at a low altitude. For example, the airborne visible infrared imaging spectrometer (AVIRIS) collects two types of images, 16 m spatial resolution at a 20 km altitude and 4 m spatial resolution at a 4 km altitude, between which only the 4 m data may be fine enough for this task. Considering the difficulty of acquiring these airborne data for multiple urban areas or time periods, a more applicable way is to use an orbital imaging spectrometer to measure urban composition globally and track changes over time.

Currently, several upcoming orbital imaging spectrometers are able to achieve this goal, such as NASA's Hyperspectral Infrared Imager (HyspIRI) (Lee et al., 2015), Germany's Environmental Mapping and Analysis Program (EnMAP) (Guanter et al., 2015). They will significantly increase the coverage range and revisit frequency of the same sites, scaling up the current local case studies using infrequent hyperspectral measurements to a regional and global level. However, due to their high altitude collection process and the tradeoff between spatial and spectral details, data from these orbital spectrometers typically have very low spatial resolutions (e.g. 30 m/pixel for HyspIRI and EnMAP (Lee et al., 2015)). At such a coarse resolution, multiple materials can exist in a pixel and contribute to the measured spectrum, also known as a mixed pixel (Small, 2001). From the spectra of these mixed pixels, sub-pixel composition can be estimated by using spectral mixture analysis (SMA), which tries to find the underlying constituting material spectra (*endmember*) and their fractions (*abundance*). By unmixing these pixels to obtain their sub-pixel fractions, we can measure and track urban composition globally, which can serve as input for advanced ecosystem models.

### 1.1. Spectral mixture analysis

The most common SMA that relates endmembers and abundances to a pixel relies on the *linear mixing model* (LMM), which assumes that the reflectance measured within each pixel is a unique linear combination of a fixed set of endmember spectra, weighted by their abundances, plus some noise (Settle and Drake, 1993). The intuition behind this model is that given a flat surface the fractional area of a material determines its representation in the measured signal. There are two problems to this model. First, spectral reflectance for identical materials can be highly variable. For example, asphalt spectra can vary significantly based on age, shadowing, and composite materials (Herold and Roberts, 2005). This is called *endmember variability*, and it is caused by several extrinsic factors and intrinsic factors, such as illumination, atmospheric condition, and measurement scales (objects or materials considered "pure" may in reality be composed of materials at smaller scales) (Somers et al., 2011; Zare and Ho, 2014). Second, multiple scattering and shadowing exist inevitably due to topographic complexity in urban scenes, which breaks the single scattering assumption in the LMM. To overcome the second problem, we may introduce more complicated nonlinear models, such as bilinear models (Heylen et al., 2014). However, the complexity in urban scenes makes accurate modeling of multiple scattering a difficult task. On the other hand, if we allow the endmember spectra to vary per pixel according to endmember variability, we may mitigate these two problems simultaneously.

The first advantage of this scheme comes from the fact that endmember is a hierarchical notion. For example, a tree is composed of trunk, bark, leaf, and each of them can be called an endmember. Similarly, roofs can contain metal chimneys, and pavements have cars, pedestrians, etc. For any macroscopic object, it may be decomposed into different levels of components, and each level can be seen as an endmember. By defining endmembers as high level objects (e.g. tree, roof, pavement), the nonlinear interaction of low level objects is treated as endmember variability within this object hence ignored in the LMM. Furthermore, the nonlinear interaction among these high level objects (e.g. multiple scattering) can also be treated as endmember variability. For example, consider a case where the light ray first hits the building or tree, then the ground, and finally received by the sensor. Since the resulting pixel on a fine enough image still corresponds to the ground, the effect of multiple scattering is ultimately incorporated into the formation of

the ground spectrum. Hence, considering endmember variability, we assume that these high level objects still contribute to a pixel spectrum in a linear way except that their spectra are altered either intrinsically or extrinsically.

By modeling endmember variability, the LMM follows equation:

$$\mathbf{y}_n = \sum_{j=1}^M \mathbf{m}_{nj} \alpha_{nj} + \mathbf{n}_n, n = 1, \dots, N \quad (1)$$

where  $\mathbf{y}_n \in \mathbb{R}^B$  is the spectrum of the  $n$ th pixel in the image (we use  $\mathbb{R}^B$  to denote the  $B$  dimensional vector space hence  $\mathbf{y}_n \in \mathbb{R}^B$  is short for  $\mathbf{y}_n$  being a  $B$  dimensional vector),  $B$  is the number of bands,  $N$  is the number of pixels,  $M$  is the number of endmembers.  $\mathbf{m}_{nj} \in \mathbb{R}^B$  is the  $j$ th endmember for the  $n$ th pixel.  $\alpha_{nj} \in \mathbb{R}$  is the abundance that usually satisfies the positivity and sum-to-one constraints, i.e.  $\alpha_{nj} \geq 0$ ,  $\sum_j \alpha_{nj} = 1$ . Finally, we have some additive noise  $\mathbf{n}_n$ . A direct observation of Eq. (1) is that the more sparse the abundances, the more confident we are about the validity of the model since in the extreme case when there is only 1 endmember in the pixel, the equation always holds ( $\alpha_{nk} = 1$  for some  $k$  and  $\alpha_{nj} = 0$  for  $j \neq k$ ).

When it comes to unmixing in terms of (1), we are referring to retrieving  $\{\mathbf{m}_{nj}, \alpha_{nj}\}$  from  $\{\mathbf{y}_n\}$ , or  $\{\alpha_{nj}\}$  from  $\{\mathbf{y}_n\}$  and a library of endmember spectra. The former can be called *unsupervised unmixing*, and this is a difficult problem. Studies that have worked to solve unsupervised unmixing usually require several assumptions, such as spatial smoothness of the abundances and the existence of contiguous pure pixels (Drumetz et al., 2016; Halimi et al., 2015; Zhou et al., 2018). However, endmember can be defined at any level in the object composition/category hierarchy. For example, suppose that there are different types of trees. We can either treat all different trees as one tree endmember or as different endmembers. The consequence is that for unsupervised unmixing, not only the solution is underdetermined, but also the validation is difficult to conduct. A more reasonable way is to define the endmember classes at one level and estimate the abundances given a spectral library at this level, which can be called *supervised unmixing*. Since true abundances can be obtained from co-registered high-resolution images, this problem has a unique solution and can be validated.

## 1.2. Previous work

Previous studies for this problem have used methods that can be categorized as *set-based* or *distribution-based* (Zare and Ho, 2014). Set-based methods treat the spectral library as an unordered set and try to pick the best combination of endmembers to model each pixel. A widely used set-based method is multiple endmember spectral mixture analysis (MESMA) (Roberts et al., 1998). The general idea of MESMA is to test every endmember combination and select the one with the smallest error within set thresholds that limit pixel complexity. Since it utilizes all the spectra in the library for unmixing, the accuracy can be excellent if the library is well derived. The success of MESMA in urban applications has been reported in (Franke et al., 2009; Powell et al., 2007; Rashed et al., 2003; Roberts et al., 2012; Wu et al., 2014).

There are many variations to MESMA. In multiple-endmember linear spectral unmixing model (MELSUM), the solution for abundances is obtained from directly solving the linear equations and discarding the negative values (Combe et al., 2008). In Bayesian spectral mixture analysis (BSMA), the final abundances are weighted sums of abundances from all combinations where the weights are proportional to the probability of endmembers in the library (Song, 2005). In automatic Monte Carlo unmixing (AutoMCU), pixels are unmixed using multiple sets of random combinations, with the mean fractional values assigned as abundances (Asner and Heidebrecht, 2002; Asner and Lobell, 2000). In alternate angle minimization (AAM), projection is iteratively used to find the spectrum index of one endmember given the other endmembers fixed (Heylen et al., 2016).

Besides MESMA variants, there is sparse unmixing that used a large spectral library with a sparsity constraint on the abundances, i.e. forcing the number of nonzero elements in the abundance vector to be small (Iordache et al., 2011; Tang et al., 2015). By using different norms on the abundance vector, it can control sparsity within each endmember class, or between endmember classes (Drumetz et al., 2019). A common way to optimize their objective function is through alternating direction method of multipliers (ADMM). ADMM is an iterative algorithm that gradually refines the abundances to approximate the optimal values. The sparsity constraint is handled by a proximal operator in the iteration.

Contrary to set-based methods, distribution-based methods assume that the endmembers for each pixel are sampled from probability distributions, hence the pixels as linear combinations of these endmembers also follow some distribution. It works by extracting parameters to represent these distributions, and unmixing the pixels based on the distribution parameters. Since it encodes the full library into only a few parameters, it can handle a large spectral library, which is particularly helpful to urban studies where the library derived from the image contains a large amount of spectra.

The most widely used distribution is Gaussian, and its application for spectral unmixing is known as the normal compositional model (NCM) (Eches et al., 2010a,b; Halimi et al., 2015; Stein, 2003; Zare and Gader, 2010; Zhang et al., 2014). The popularity of NCM comes from the fact that a linear combination of Gaussian random variables is also a Gaussian random variable whose mean and covariance matrix are linear combinations of the endmember means and covariance matrices. Hence, the resulting probability density function of the pixels has a simple analytical form. Fitting the actual pixel values to the pixel distribution, the abundances can be solved by several techniques, such as expectation maximization (Stein, 2003), sampling methods (Eches et al., 2010a,b; Halimi et al., 2015), and particle swarm optimization (Zhang et al., 2014).

Following this philosophy, some have worked to extend the idea to distributions beyond Gaussian. Du et al. (2014) proposed Beta distributions to model the spectral library. The benefit is that Beta distributions have a domain in the range 0 – 1, so are more suitable for the reflectance range, and the actual library may have a skewed mode in the distribution. They refer to the model as Beta compositional model (BCM) in correspondence to NCM. Zhou et al. (2018) further extended the idea to use Gaussian mixture models (GMM) for distributions. The rationale comes from the observation that library endmembers may have multiple modes, whose shape cannot be represented by a simple Gaussian or Beta distribution. Since GMM is more flexible, it can approximate any distribution found in the library.

### 1.3. Motivation

Supervised unmixing requires a spectral library, which can be extracted from the images (Franke et al., 2009) or laboratory (Kotthaus et al., 2014) or field measurements (Herold et al., 2004). Building a specific library for each study site is a time-consuming process. With existence of many spectral libraries from images collected at different spatial resolutions, for different cities, during different times, a more operational way is to combine them into a *universal spectral library* that can be applied to images from different locations, sensors and timings (Degerickx et al., 2017). Ideally, this universal spectral library should be large enough to capture all the variability foreseen. Given such a large library, it is necessary to reduce/prune it before application to a specific image using set-based methods. This shifts the focus of unmixing from the actual unmixing method to library reduction, leading to the emergence of a plethora of library pruning methods (Fan and Deng, 2014; García-Haro et al., 2005; Schaaf et al., 2011). Despite wide acceptability of this scheme, this 2-step approach is problematic: (i) library reduction may be laborious; (ii) the reduced library is targeted to a specific unmixing image hence the process needs to repeated for each dataset. We propose to solve the entire problem using a unified model that does not artificially separate it into 2 steps.

Potential candidates of this unified model are distribution-based methods since they innately accept large spectral libraries. However, they need to be evaluated more comprehensively since previous validation of these methods relies on comparisons to reference libraries, segmented images, or assessment of reconstruction error (Du et al., 2014; Zare et al., 2013; Zhou et al., 2018). These validation methods are not convincing enough, especially for urban imagery. First, different conditions (sensor, atmosphere, light source) during data collection will affect measured reflectances, making library comparison less ideal. Second, high spatial resolution hyperspectral images are primarily composed of pure pixels, and segmentation like abundance maps do not necessarily indicate good unmixing capability for mixed pixels. Third, reconstruction error is more related to model complexity than unmixing accuracy since small reconstruction error could be achieved by overfitting (Murphy, 2012).

Following the validation of MESMA, a better validation approach is to find high-resolution color images corresponding to the hyperspectral image, determine the abundances from these color images and validate the estimated abundance values (Franke et al., 2009; Powell et al., 2007; Roberts et al., 2012). Because deriving such a dataset is a labor-intensive and time-consuming process, to the best of our knowledge, distribution-based methods such as GMM and NCM have not been validated in this way in urban studies. The significance of the work is that if a method that uses the original spectral library has unmixing accuracy on par with or close to MESMA that uses the reduced library, then (i) manual work on library reduction

may be avoided in the case of a large spectral library, (ii) the method may be more stable to various datasets with this intermediate step removed, (iii) it will facilitate applications with a universal spectral library.

In this work, we applied several set-based and distribution-based unmixing methods to a highly validated, comprehensive dataset of 128 urban images with spatial resolutions of 4 m and 16 m. The dataset includes a wide range of urban landcover of different mixtures and a variety of materials, including different types of road, roof, vegetation, and soil (Section 2). Using this dataset, we investigated the abilities of distribution-based methods GMM, NCM, BCM (Section 3) and compared them to set-based methods MESMA, AAM on characterizing urban material abundances (Section 4).

## 2. Data

The study area includes the cities of Santa Barbara and Goleta as well as the land between them, near the California coast. Urban composition is typical of the southwestern United States, including man-made materials such as asphalt, concrete, metal, gravel, and brick, as well as vegetation in the forms of turfgrass, various tree species, and large areas of undeveloped land covered in senesced vegetation (Roberts et al., 2012).

We used two low-resolution images (16 m) and two high-resolution images (4 m) in this study. The low-resolution images were collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Vane et al., 1993) over Santa Barbara, CA, on August 29, 2014. The original data are downloadable from the AVIRIS data portal ([https://aviris.jpl.nasa.gov/alt\\_locator/](https://aviris.jpl.nasa.gov/alt_locator/)), where the flight names are “f140829t01p00r09” and “f140829t01p00r10”. The spatial resolutions are 15.6 m/pixel and 15.8 m/pixel. The spectral range measures wavelengths from 380 – 2500 nm with 224 bands of approximately 10 nm bandwidth. High-resolution images were collected on September 5 by AVIRIS-Next Generation (AVIRIS-NG) with 3.9 m/pixel and 3.6 m/pixel spatial resolutions. The spectral resolution is also higher, recording 432 bands of about 5 - 6 nm bandwidth across a similar spectral range as the 16 m dataset. We spectrally resampled the AVIRIS-NG imagery to 224 bands (by linear interpolation) to produce an image with identical spectral parameters to the 16 m AVIRIS image. We also removed certain bands due to atmospheric interference, reducing the number of bands to 164. Initial image processing was conducted by the Jet Propulsion Laboratory, with additional processing in the lab to reduce the effects of elevation change on pixel location.

### 2.1. Validation of abundance maps

Pixel-wise accuracy assessment of abundances is difficult since co-registration between the hyperspectral and reference images can be inaccurate and the signal of a pixel can be influenced by its neighboring pixels (Huang et al., 2002). A common way to mitigate this effect is to evaluate the total abundance in a larger spatial unit, e.g. polygon (Powell et al., 2007). We produced 64 polygons that represented the variety of landcover within the study area as described by Wetherley et al. (2017) (see Fig. 1). Each polygon was 180 m by 180 m in size, corresponding to 11-12 pixels wide in the 16 m images or 46 or 50 pixels wide in the 4 m images. Validation polygons were randomly distributed across the area with a minimum distance 400 m. If a polygon contained large areas of open water or an undetermined material, it was discarded and a new polygon was randomly generated (small areas of water, e.g. swimming pools, are present in some polygons, e.g. No. 37 and 58 in Fig. 1). Cover was determined within each polygon using a 1 m NAIP high-resolution image. We used a combination of image segmentation, using the ECognition software package (Baatz et al., 2004), and manual adjustments to classify the cover within each polygon as *turfgrass*, *tree*, *paved*, *roof*, *soil*, or *non-photosynthetic vegetation* (NPV). Cover was further confirmed by visually inspecting August 2014 Google Earth imagery. This is important for distinguishing between turfgrass and NPV as the former may have senesced to the latter by the time of the flights.

Fig. 2 shows a scatter plot of the 64 validated abundances when the 6 endmember classes are merged to 3 categories of *green vegetation* (GV), *impervious*, and non-vegetated *pervious*. Most polygons are dominated by a mixture of impervious and vegetation materials, which is an accurate reflection of the area. To improve the representation of less common mixtures in the scene, five polygons with high proportions of soil were intentionally added.



Figure 1: Validation polygons on the site (a) and all 4 m ROI images (b). The color images are created by extracting the bands closest to 650 nm, 540 nm and 470 nm. The two 16 m images are mosaicked by geographic coordinates.

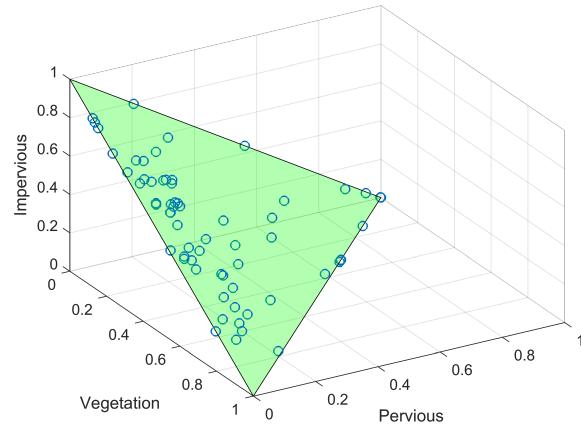


Figure 2: Scatter plot of validated total abundances in terms of 3 categories, green vegetation (turfgrass and tree), pervious (NPV and soil), and impervious (paved and roof). Most of them lie on the plane, which is in line with the selection of ROIs (almost all the pixels fall into the 6 endmember classes).

Table 1: Number of spectra for each endmember class in the libraries

	16 m		4 m	
	Full	Reduced	Full	Reduced
Turfgrass	537	10	1,468	5
NPV	884	14	3,465	7
Paved	299	6	2,902	17
Roof	435	17	2,941	16
Soil	262	3	1,442	5
Tree	870	11	3,208	45
Total	3,287	61	15,426	95

## 2.2. Spectral library building

We produced 240 polygons across the 4 m scene to extract pure spectra and build full spectral libraries. The polygons were intended to capture class material variability as much as possible, and so included multiple roof types, asphalt, concrete, trees, turfgrass, soil, and NPV, as well as less common materials like rubber, solar panels, tennis courts, and plastic tarps. These materials were then grouped into one of our 6 endmember classes: turfgrass, NPV, paved, roof, soil, and tree. The same polygons were used to extract spectra from the 16 m imagery, with necessary modifications as described in (Wetherley et al., 2017). Together, we produced a library of 16 m spectra and a library of 4 m spectra. After removing duplicate spectra, the final 16 m library was comprised of 3,287 spectra and the 4 m library contained 15,426 spectra.

These full spectral libraries were used to train the parameters of distribution-based methods. To use them with MESMA, we performed library reduction in two steps. First, iterative endmember selection (IES) (Schaaf et al., 2011) was used to automatically select a subset of spectra that represented the larger library. This is achieved iteratively, by gradually selecting the most representative spectra and evaluating their representativeness using a kappa coefficient. IES reduced the 16 m and 4 m library sizes to 226 and 187, respectively. Libraries were further reduced using iterative classification reduction (ICR), which uses MESMA as a classifier to identify and remove spectra that tend to map materials incorrectly (Wetherley et al., 2017). This step requires visually inspecting the results and manually removing spectra by referring to validated abundances. We carefully examined the results to maximize the performance of MESMA, resulting in reduced libraries of 61 spectra for 16 m images and 95 spectra for 4 m images. The spectra for each endmember class for all the cases are plotted in Fig 3, and their numbers are shown in Table 1.

## 3. Methods

We assess three distribution-based methods, GMM, NCM, BCM, and two set-based methods, MESMA, AAM on the 16 m and 4 m images with input of the same resolution spectral libraries.

### 3.1. Distribution-based methods

Distribution-based methods typically use the following generative process to model the observed spectra:

1. For each pixel,  $n = 1, 2, \dots, N$ :
  - 1.1. For each endmember,  $j = 1, 2, \dots, M$ ,  $\mathbf{m}_{nj} \sim p(\mathbf{m}_{nj} | \boldsymbol{\theta}_j)$ ,
  - 1.2.  $\mathbf{n}_n \sim p(\mathbf{n}_n | \boldsymbol{\theta}_0)$ ,
  - 1.3.  $\mathbf{y}_n \leftarrow \sum_{j=1}^M \mathbf{m}_{nj} \alpha_{nj} + \mathbf{n}_n$ .

Namely, suppose we have  $M$  endmember classes, a pixel spectrum  $\mathbf{y}_n$  can be assumed to be generated by randomly picking one spectrum  $\mathbf{m}_{nj}$  from each class, and linearly mixing them based on some abundances  $\{\alpha_{nj}\}$  plus some additive noise  $\mathbf{n}_n$ . Hence, if we use a probability distribution  $p(\mathbf{m}_{nj} | \boldsymbol{\theta}_j)$  to represent the spectral distribution, the actual endmembers  $\{\mathbf{m}_{nj}\}$  can be assumed to be sampled from this distribution. Given a spectral library, the distribution parameters  $\{\boldsymbol{\theta}_j : j = 1, \dots, M\}$  can be inferred from the library and our goal is to find  $\{\alpha_{nj}\}$  that produces the observed spectra.

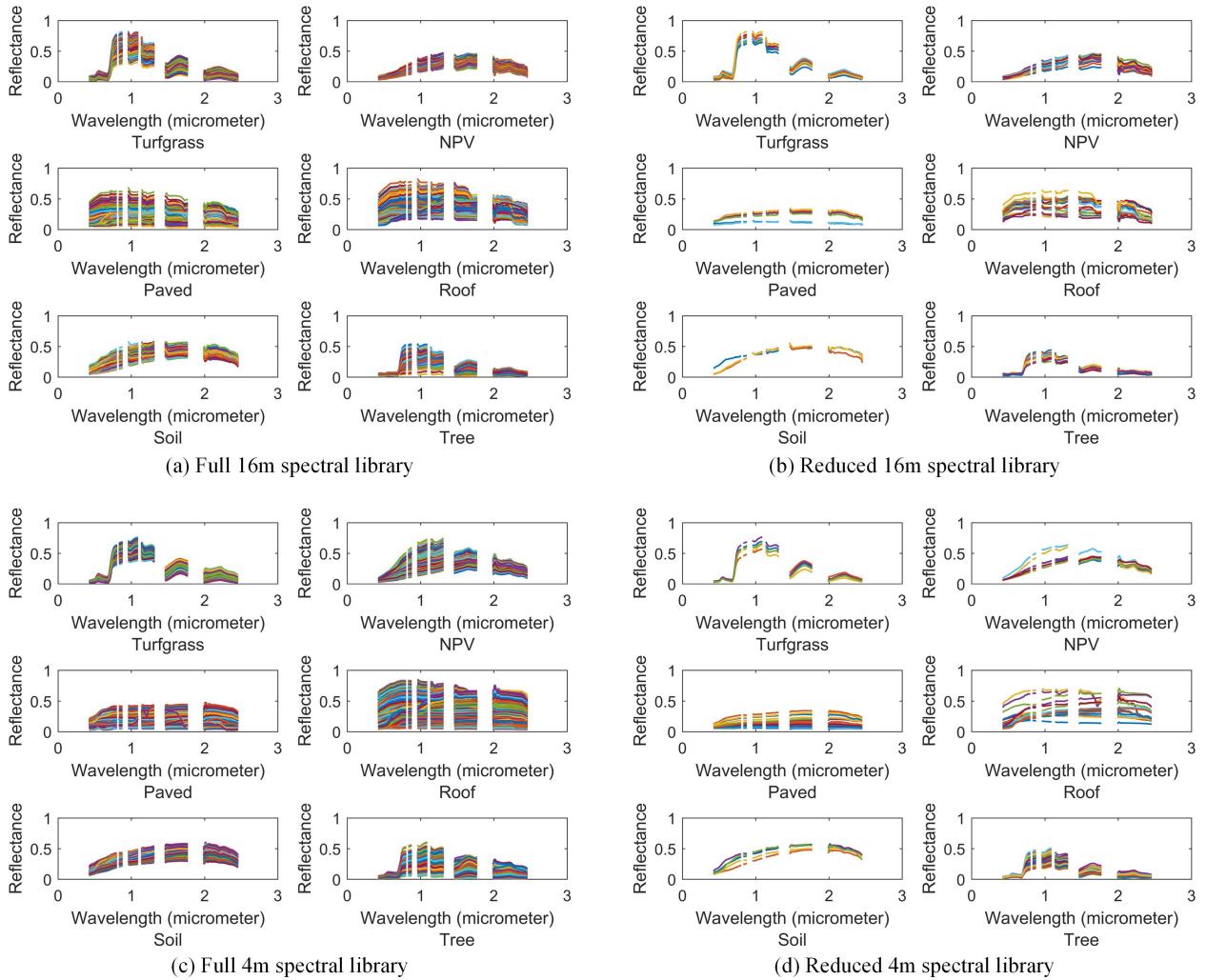


Figure 3: Original and reduced spectral libraries. The numbers of spectra in each category are shown in Table 1.

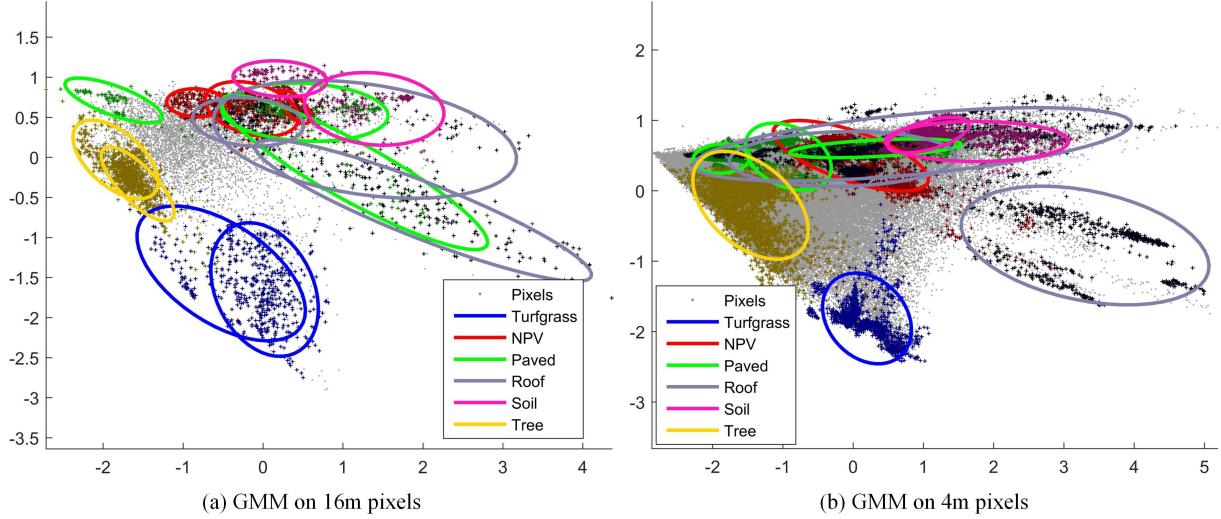


Figure 4: Scatter plot of GMM components on the pixels and library spectra. The projection is determined by performing PCA on all the spectra in the library. The pixels of 64 images for each scale are combined and denoted by gray dots. The colored dots show the spectra in the library for each endmember class. The ellipses represent the projected Gaussian components (twice the standard deviation along the major and minor axes, covering 86% of the total probability mass).

### 3.1.1. Gaussian mixture model

The GMM method uses a GMM distribution as  $p(\mathbf{m}_{nj}|\boldsymbol{\theta}_j)$  (Zhou et al., 2018). Specifically,

$$p(\mathbf{m}_{nj} | \boldsymbol{\theta}_j) = \sum_{k=1}^{K_j} \pi_{jk} \mathcal{N}(\mathbf{m}_{nj} | \boldsymbol{\mu}_{jk}, \boldsymbol{\Sigma}_{jk}), \quad (2)$$

where  $\mathcal{N}$  is the multivariate Gaussian density function with mean  $\mu_{jk}$  and covariance matrix  $\Sigma_{jk}$ ,  $\{\pi_{jk} : \pi_{jk} \geq 0, \sum_k \pi_{jk} = 1\}$  are the prior probabilities,  $K_j$  is the number of components for the  $j$ th endmember class. We refer to the GMM parameters  $\{\pi_{jk}, \mu_{jk}, \Sigma_{jk} : k = 1, \dots, K_j\}$  as  $\theta_j$ .

Given a spectral library,  $\{K_j\}$  can be estimated through *model selection* (McLachlan and Rathnayake, 2014) and  $\theta_j$  can be estimated through the standard *expectation maximization* (EM) algorithm (Bishop, 2006). In this study, we use cross-validation-based information criterion (CVIC) (Smyth, 2000) to select  $K_j$ . It works by calculating a log-likelihood value  $\mathcal{L}_{K_j}$  for each candidate value of  $K_j$  (e.g.  $K_j = 1, 2, 3, 4$ ) and picks the  $K_j$  that maximizes this log-likelihood. Specifically, the log-likelihood value is calculated by dividing the spectra into  $V$  subsets and adding up the log-likelihood values of these subsets given the remaining subsets for training. To reduce the number of components from a large library, we slightly modified this procedure by picking the smallest  $K_j$  such that  $|\mathcal{L}_{K_j} - \mathcal{L}'_j| \leq T_{CVIC} \mathcal{L}'_j$ , where  $\mathcal{L}'_j$  is the maximum value and  $T_{CVIC}$  is a threshold parameter. We use  $T_{CVIC} = 0.05$  for the 16 m data. For the 4 m data, since the spectral library is 5 times larger and the number of pixels to unmix is about 20 times more, we use a large  $T_{CVIC} = 0.2$  to select less components. For reproducibility, this entire procedure was repeated 15 times and the most frequent combination was selected. Fig. 4 shows the results by projecting all the pixels, library spectra, and Gaussian components to 2 dimensions. We see that the Gaussian components (ellipses) surround validation pixels at multiple positions on the edge of the pixel cloud. The pixels can be viewed as picking points within the ellipses and combining these points linearly by Eq. (1). Fig. 5 shows the Gaussian components from the wavelength-reflectance perspective, where the centers of Gaussian components and their variation patterns are shown as curves.

Following the distribution assumption on endmembers, the probability density function of pixels can be derived according to Eq. (1) and (2) using Theorem 1 in Zhou et al. (2018). Fitting the pixel spectra to this density function, the abundances can be estimated by *maximum likelihood estimation* (MLE), which maximizes

$$L(\{\alpha_{nj}\}) = \sum_{n=1}^N \log \sum_{\mathbf{k} \in \mathcal{K}} \pi_{\mathbf{k}} \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}}), \quad \text{s.t. } \alpha_{nj} \geq 0, \sum_{j=1}^M \alpha_{nj} = 1 \quad (3)$$

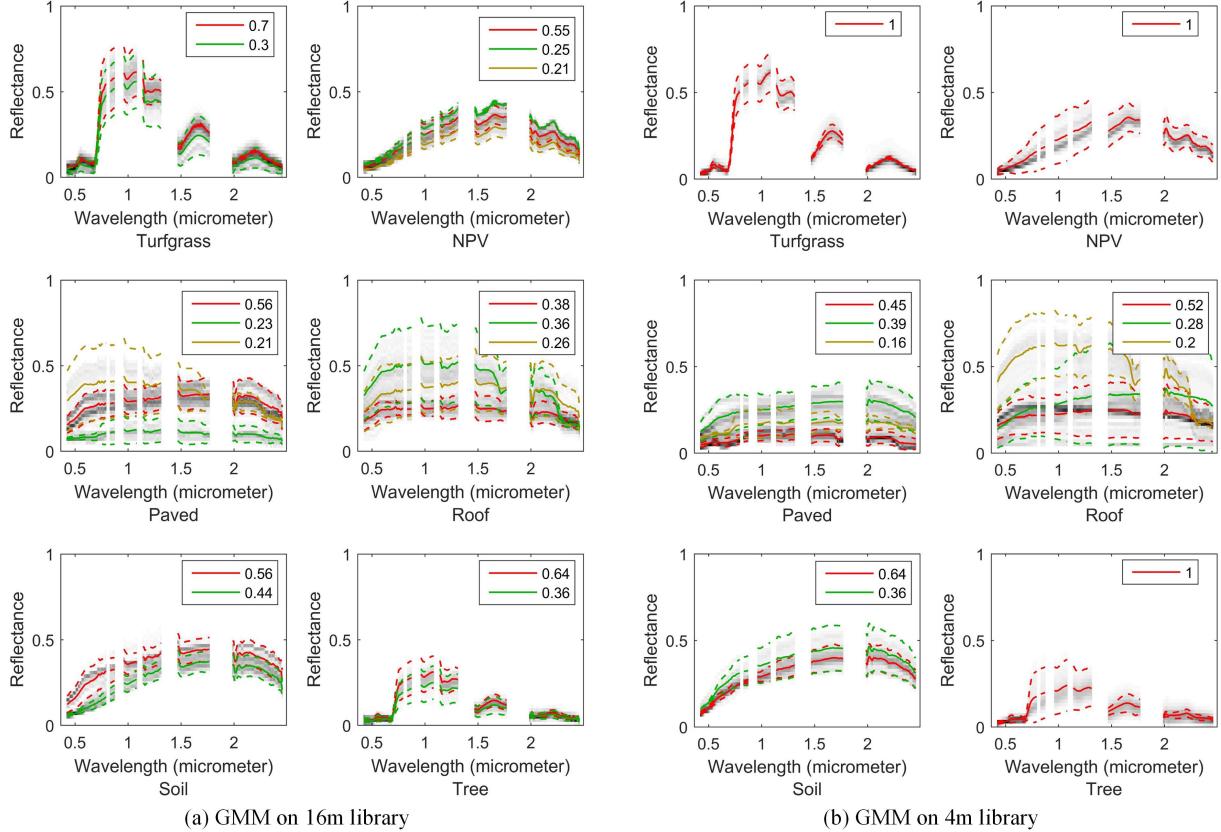


Figure 5: Wavelength-reflectance plot of GMM components on the library spectra. The spectra are put into 2-dimensional bins of wavelength-reflectance to form a histogram shown as gray scale background images. The center of each Gaussian component is shown as a solid curve. The center plus (minus) the largest eigenvector multiplied by twice the square root of the eigenvalue is shown as a dashed curve, which indicates the major variation pattern of a Gaussian component. The prior probabilities are shown in the legends.

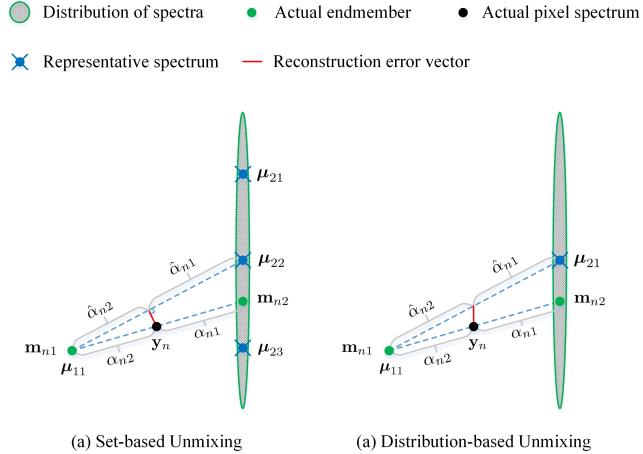


Figure 6: Rationale behind set-based methods (a) and GMM/NCM (b) for a simple example with 2 endmember classes. The first one has no endmember variability. The second one's variability is captured by the ellipse. Set-based methods try to find some representative spectra from the library such that their linear combination is closest to the pixel spectrum. GMM/NCM tries to find a linear combination of the centers such that the reconstruction error is in line with the variation in the library. Given ground truth  $\{\alpha_{nj}\}$ , the estimated abundances  $\{\hat{\alpha}_{nj}\}$  from GMM/NCM can exactly recover it ( $\hat{\alpha}_{n1}/\hat{\alpha}_{n2} = \alpha_{n1}/\alpha_{n2}$ ) while for set-based methods,  $\hat{\alpha}_{n1}/\hat{\alpha}_{n2} \neq \alpha_{n1}/\alpha_{n2}$ .

where  $\mathcal{K} = \{1, \dots, K_1\} \times \{1, \dots, K_2\} \times \dots \times \{1, \dots, K_M\}$  is the Cartesian product of the  $M$  index sets,  $\mathbf{k} = (k_1, \dots, k_M) \in \mathcal{K}$ , and  $\pi_{\mathbf{k}}$ ,  $\mu_{n\mathbf{k}}$ ,  $\Sigma_{n\mathbf{k}}$  are defined by

$$\pi_{\mathbf{k}} = \prod_{j=1}^M \pi_{jk_j}, \mu_{n\mathbf{k}} = \sum_{j=1}^M \alpha_{nj} \mu_{jk_j}, \Sigma_{n\mathbf{k}} = \sum_{j=1}^M \alpha_{nj}^2 \Sigma_{jk_j} + \epsilon \mathbf{I}. \quad (4)$$

$\epsilon$  is the noise variance that is usually negligible compared to the linear combination of  $\{\Sigma_{jk}\}$ . The objective function has an intuitive explanation. As seen in (4), every linear combination of  $\{\mu_{jk}\}$  given the abundances should be close to  $\mathbf{y}_n$ . However, this closeness is weighted in (3) by both the linear combination of  $\{\Sigma_{jk}\}$  (how endmembers scatter in a class) and the probability of picking a combination ( $\pi_k$ ).

Fig. 6(b) shows the rationale behind GMM for a simple example with 2 endmember classes. Using one Gaussian component, we have  $p(\mathbf{m}_{n1}|\boldsymbol{\theta}_1) = \mathcal{N}(\mathbf{m}_{n1}|\boldsymbol{\mu}_{11}, \epsilon\mathbf{I})$  and  $p(\mathbf{m}_{n2}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{m}_{n2}|\boldsymbol{\mu}_{21}, \boldsymbol{\Sigma}_{21})$  where  $\boldsymbol{\Sigma}_{21}$  is the covariance matrix corresponding to the ellipse. Since  $\epsilon\mathbf{I}$  is negligible compared to  $\boldsymbol{\Sigma}_{21}$ ,  $\boldsymbol{\Sigma}_{nk}$  in (4) is dominated by  $\alpha_{n2}^2 \boldsymbol{\Sigma}_{21}$  and the objective function (3) becomes mainly minimizing

$$\left( \mathbf{y}_n - \sum_{j=1}^2 \alpha_{nj} \boldsymbol{\mu}_{j1} \right)^T \boldsymbol{\Sigma}_{21}^{-1} \left( \mathbf{y}_n - \sum_{j=1}^2 \alpha_{nj} \boldsymbol{\mu}_{j1} \right). \quad (5)$$

Let  $\mathbf{v}_1$  and  $\mathbf{v}_2$  be the eigenvectors of  $\Sigma_{21}$  and  $\lambda_1$  and  $\lambda_2$  be the corresponding eigenvalues. Eq. (5) can be written as  $\sum_k \frac{1}{\lambda_k} \mathbf{v}_k^T (\mathbf{y}_n - \sum_j \alpha_{nj} \boldsymbol{\mu}_{j1})$ . Note that  $\mathbf{v}_1$  points to the direction of the large variation (vertical) and  $\mathbf{v}_2$  is perpendicular to  $\mathbf{v}_1$  (horizontal), and  $\lambda_1 \gg \lambda_2 > 0$ . Hence Eq. (5) can tolerate a large error along  $\mathbf{v}_1$  while only a little error along  $\mathbf{v}_2$ , resulting in estimated abundances  $\{\hat{\alpha}_{nj}\}$  shown in the figure (up to scaling such that  $\sum_j \hat{\alpha}_{nj} = 1$ ). Clearly,  $\hat{\alpha}_{n1}/\hat{\alpha}_{n2}$  is identical to the ground truth  $\alpha_{n1}/\alpha_{n2}$ . Moreover, even if we move  $\boldsymbol{\mu}_{21}$  in its range, as long as the covariance matrix is the same, we can still accurately retrieve the original abundances. This implies that GMM/NCM is more robust to changing representative spectra (Gaussian centers).

The optimization problem (3) can be solved by a generalized EM algorithm that alternates between an E step and an M step (Zhou et al., 2018). Because of the computational cost, it is implemented in a low dimensional subspace (10 dimensions) obtained by applying PCA on the original spectral data. Since Eq. (1) still holds if both the pixel spectra and endmember spectra are projected to a subspace, the estimated abundances for the projected spectra are the original abundances. To determine the projection direction, we selected an equal number of spectra for each endmember class in the library and concatenated them as

input for PCA. This ensures that the relative sizes of endmember classes do not affect the direction, and also ensures that the mean lies in the center.

### 3.1.2. Normal compositional model

Similar to GMM, NCM models the distribution of each endmember class as a Gaussian distribution and solves the MLE problem for the abundances. Since there are multiple implementations, we compared two NCM versions in this study. First, it is a special case of GMM when the number of components  $K_j = 1$  for all the endmember classes. Hence, we can use the same implementation of GMM while forcing  $K_j = 1$  as one type of NCM, referred to as GMM-1. Since it has the same implementation as GMM, it should reflect the true difference introduced by bringing multiple components.

Many variations of NCM utilize MCMC sampling as the optimization approach (Eches et al., 2010a,b; Halimi et al., 2015; Stein, 2003; Zare and Gader, 2010; Zhang et al., 2014). The second type of NCM uses a sampling algorithm in (Zare et al., 2013). The sampling method is the Metropolis-Hastings algorithm. It is an iterative algorithm that starts with an initial abundance value, then for each iteration, it samples abundance values from a Dirichlet distribution, and decides to accept the new value or not based on a random number. Specifically, the criterion to accept is to compare a uniformly generated random number in  $[0, 1]$  to the ratio of the likelihood values of the new abundance value and the old one. If the new likelihood is greater than the old one, i.e. the new abundance leads to a higher likelihood, the ratio is greater than 1 hence the new abundance is always accepted no matter what the random number is. Otherwise, the ratio is less than 1 and it becomes the probability to accept the new value. The algorithm lasts a predefined number of iterations and the last sample is kept as the final abundances. We tried 1000, 3000 iterations and picked the former after finding no significant improvement of accuracy with more iterations.

The difference between these two NCMs is not only the optimization technique, but also the number of dimensions of the input data. GMM-1 uses projected data with 10 dimensions, while NCM sampling uses the original data without dimensionality reduction. Hence, the latter will utilize more spectral information.

### 3.1.3. Beta compositional model

Beta composition model (BCM) was proposed to model the skewness in the endmember distribution (Du et al., 2014). It assumes that the endmembers are sampled from Beta distributions. Though a linear combination of Beta distribution random variables can have a complex distribution, it uses a Beta distribution to approximate this combined random variable. For each pixel, it uses K-means to find similar spectra and assumes that they have similar abundances. Then, from these spectra, the mean and variance of a Beta distribution can be inferred. On the other hand, the mean and variance of a combination of Beta distribution random variables can be obtained from the endmember distributions and the abundances. Equating these two sources of means and variances leads to an optimization problem that can be solved for the abundances.

There are two versions of BCM proposed, a spectral version and a spatial version with two kinds of optimization technique, quadratic programming and Metropolis-Hastings sampling. We chose the spectral version with quadratic programming, which minimizes the squared difference between the two sources of means. The code is available from Alina Zare's website. Comparing to the two NCMs above, a disadvantage of BCM is that it assumes that the bands are statistically independent, which is usually not the case in practice.

## 3.2. Set-based methods

Set-based methods aim to minimize the following objective function with respect to the abundances

$$L(\{\alpha_{nj}\}) = \sum_n \min_{\mathbf{k} \in \mathcal{K}} \|\mathbf{y}_n - \sum_{j=1}^M \alpha_{nj} \boldsymbol{\mu}_{jk_j}\|, \text{ s.t. } \alpha_{nj} \geq 0, \sum_{j=1}^M \alpha_{nj} = 1 \quad (6)$$

where  $\{\boldsymbol{\mu}_{jk} : j = 1, \dots, M, k = 1, \dots, K_j\}$  are the spectra in the library and  $K_j$  is the number of spectra for the  $j$ th class,  $\mathbf{k} = (k_1, \dots, k_M)$  and  $\mathcal{K}$  is again the Cartesian product of the  $M$  index sets  $\{1, \dots, K_j\}$ ,  $j = 1, \dots, M$ . Fig. 6(a) shows how this objective function works, which can be seen as finding a combination  $\mathbf{k}$  that is closest to  $\mathbf{y}_n$  in terms of projection distance. We reuse the notations in the GMM since theoretically, set-based methods can be seen as a special case of GMM where each spectrum in the library is a Gaussian component (see Appendix A for a proof). In practice, the difficulty of minimizing (6) comes from enumerating all the possible combinations and different methods take a different route to handle this.

Table 2: Characteristics of the comparing methods

Method	MESMA	AAM	GMM	GMM-1	NCM Sampling	BCM
Physical Assumption	LMM	LMM	LMM	LMM	LMM	LMM
Input Library	Reduced	Reduced	Full	Full	Full	Full
Objective Function	Reconstruction error	Reconstruction error	GMM likelihood	NCM likelihood	NCM likelihood	Moments matching
Operating Dimension	164	164	10	10	164	164
Optimization	Exhaustive search	Block coordinate descent	EM	EM	Metropolis Hasting	Quadratic programming
Maximal Endmembers Per Pixel	3	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
Inherent Shade Endmember	Yes	No	No	No	No	No

### 3.2.1. Multiple endmember spectral mixture analysis

The most widely used set-based method could be MESMA (Roberts et al., 1998). MESMA refers to each combination  $\mathbf{k}$  as a model and an  $P$ -endmember model means that there are only  $P$  nonzero  $\alpha_{nj}$  for each pixel. From one-endmember model to two-endmember model and etc., MESMA gradually increases the number of endmembers in a model if more endmembers improve the error by a threshold. MESMA also considers an additional shade endmember, which is set to be zero over all the bands. An additional threshold is set to prevent the abundance of shade from becoming too large.

The implementation was done by the original authors in IDL and provided as an extension Viper Tool to ENVI. We used the same parameters as in (Wetherley et al., 2017), i.e. maximum RMSE 2.5%, threshold RMSE 0.7%, abundances constrained between 0 and 1, maximum shade threshold 80%, and a maximum of three endmembers plus shade for each pixel. The obtained fractions were normalized to give the final abundances.

### 3.2.2. Alternate angle minimization

Unlike MESMA that iteratively tries each combination, alternate angle minimization (AAM) iteratively processes each endmember class by picking an endmember for this class while fixing the rest endmembers (Heylen et al., 2016). The theory is that given some fixed endmembers and the pixel spectrum, the reconstruction error of adding an endmember is proportional to the sine of the angle of two projections. Hence, finding the endmember that leads to the least error is equivalent to finding the one with the least angle. Since each search in the library for an endmember class will lower the error, iterative search over different classes will lead to an endmember set that minimizes the error. The purpose of this strategy is to reduce the computational cost of finding the best combination. It is shown that compared to the exponential time cost of MESMA, AAM has a linear time cost with respect to the library size (Heylen et al., 2016). The code was implemented in Matlab and downloaded from Rob Heylen's website.

Despite the same concept, AAM is different from MESMA in several ways. First, it may not find the global minimum because of its alternate optimization strategy. Second, it may find a pixel mixed by many endmembers instead of maximum three. Finally, it does not include a shade endmember to adjust for brightness differences between library endmembers and measured spectra.

To summarize, a comparison of the characteristics of all the methods is given in Table 2.

### 3.3. Validation strategy

Excluding MESMA, which was implemented in IDL, all methods were implemented in Matlab. Due to ENVI license availability, MESMA was run on a PC with Intel Core i7-2760QM CPU and 8 GB memory. The other methods were run on a PC with Intel Core i7-3820 CPU and 64 GB memory.

We used three metrics to measure the differences between the estimated and reference fractions: mean absolute error (MAE), root mean squared error (RMSE) and correlation coefficient ( $R$ ). They were calculated for each endmember class based on the 64 pairs of values. When comparing different methods for unmixing

Table 3: Comparison of error and correlation coefficient for the 16 m images

		Set-based		Distribution-based			
	MAE / $R^2$	MESMA	AAM	GMM	GMM-1	NCM Sampling	BCM
Individual	Turfgrass	<b>0.029 / 0.693</b>	<b>0.029 / 0.703</b>	0.045 / 0.632	0.041 / 0.629	0.193 / 0.117	0.042 / 0.610
	NPV	<b>0.069 / 0.830</b>	0.069 / 0.819	<b>0.064 / 0.805</b>	0.071 / 0.813	0.132 / 0.381	0.073 / 0.750
	Paved	0.093 / 0.588	<b>0.093 / 0.685</b>	<b>0.081 / 0.691</b>	0.093 / 0.768	0.196 / 0.058	0.096 / 0.538
	Roof	0.087 / 0.240	<b>0.079 / 0.241</b>	<b>0.078 / 0.279</b>	0.105 / 0.032	0.112 / 0.362	0.125 / 0.000
	Soil	0.069 / 0.773	0.080 / 0.768	0.071 / 0.834	<b>0.067 / 0.781</b>	0.071 / 0.000	<b>0.067 / 0.733</b>
	Tree	0.098 / 0.835	<b>0.065 / 0.855</b>	0.076 / 0.798	<b>0.071 / 0.849</b>	0.107 / 0.632	0.184 / 0.534
Merged	<b>Average<sup>a</sup></b>	<b>0.074 / 0.660</b>	<b>0.069 / 0.678</b>	<b>0.069 / 0.673</b>	<b>0.075 / 0.645</b>	<b>0.135 / 0.258</b>	<b>0.098 / 0.527</b>
	GV	0.088 / 0.909	<b>0.057 / 0.898</b>	<b>0.070 / 0.872</b>	0.073 / 0.913	0.156 / 0.515	0.175 / 0.701
	Pervious	0.082 / 0.836	0.084 / 0.870	<b>0.072 / 0.880</b>	0.075 / 0.880	0.145 / 0.515	<b>0.074 / 0.886</b>
	Impervious	0.102 / 0.762	<b>0.087 / 0.818</b>	<b>0.084 / 0.845</b>	0.088 / 0.819	0.199 / 0.181	0.175 / 0.722
	<b>Average<sup>b</sup></b>	<b>0.091 / 0.836</b>	<b>0.076 / 0.862</b>	<b>0.075 / 0.866</b>	<b>0.079 / 0.871</b>	<b>0.167 / 0.404</b>	<b>0.141 / 0.770</b>

<sup>a</sup> the average RMSEs in the individual case for the 6 methods are 0.1, 0.092, 0.093, 0.103, 0.184, 0.133 respectively.

<sup>b</sup> the average RMSEs in the merged case for the 6 methods are 0.117, 0.098, 0.097, 0.104, 0.226, 0.174 respectively.

<sup>c</sup> the entries in red denote the best two results in each category.

<sup>d</sup> the running time on all the images for the 6 methods is 18, 27, 83, 0.3, 220, 101 minutes respectively.

quality, we will mainly resort to MAE while also considering correlation coefficient and RMSE since we can do statistical tests using MAE. For RMSE, only averaged value over the 6 endmember classes is reported to save space. For  $R$ , it should be accompanied by slope and intercept for additional information. We visualize the results by scatter plots and Bland-Altman plots. Scatter plot shows the estimated value in the y-axis against the reference value in the x-axis. The slope and intercept are shown in the parenthesis below the scatter plot. Bland-Altman plot shows the difference of the estimated value and the reference value against the later (Bland and Altman, 1986). It also plots the mean of these differences with twice the standard deviation as an interval. We can expect that the error falls into this interval with 95% probability if the differences follow a Gaussian distribution.

## 4. Results

### 4.1. Accuracy and efficiency

#### 4.1.1. 16 m Case

Table 3 shows the MAE and correlation coefficient for the 16 m images (see Supplementary Fig. S1 - S64 for abundance maps). Original errors for 6 classes imply that GMM and AAM have the best accuracy, followed by MESMA. The difference comes from the paved, roof and tree classes, where GMM outperformed MESMA. In general, MESMA, AAM, GMM and GMM-1 had similar accuracy. Among all the distribution-based methods, GMM has the best performance overall, with fewest errors for NPV, paved and roof. Fig. 7 compares the estimated total abundances to validated abundances for each material in scatter plots. We can see that NCM sampling and BCM tend to ignore paved or roof when they have presence. The set-based methods and GMM appear to be better than the others.

We also summed up abundances for each category of GV, pervious and impervious. The errors for them will be referred to as merged errors and are also shown in Table 3. Merged errors show that GMM and AAM retain their higher accuracy, with MESMA falling further behind due to poor impervious accuracy. Since merged errors are differences between summed up quantities of similar materials, such as paved and roof, we may expect that merged errors will be less than the individual ones. However, the average errors in Table 3 show the opposite, which means that both of the similar materials are overestimated or underestimated.

We observe that though merging the abundances does not decrease the error, it significantly increases the correlation coefficient. We checked the scatter plot for merged abundances (see Supplementary Fig. S66). Compared to the scatter plot in Fig. 7 where many abundances are scattered close to the origin (e.g. turfgrass, roof, soil), the merged abundances are more dispersed over the full range 0 - 1. Note that correlation coefficients measure the linear correlation between two random variables. When they are close to the origin, the linear correlation will not be clear though the errors are small. Hence, we will mainly compare the MAE while keeping  $R^2$  for reference.

Fig. 8 shows the Bland-Altman plots for the 16 m data with dashed lines indicating the 95% limits of agreement (also in the parenthesis below). If we look at the boundary values of these intervals for the first 3 methods, the negative boundary for GMM is -0.22, which is slightly better than MESMA (-0.27) and AAM (-0.23). The positive boundary for GMM is 0.21, also slightly better than MESMA (0.26) and AAM (0.25).

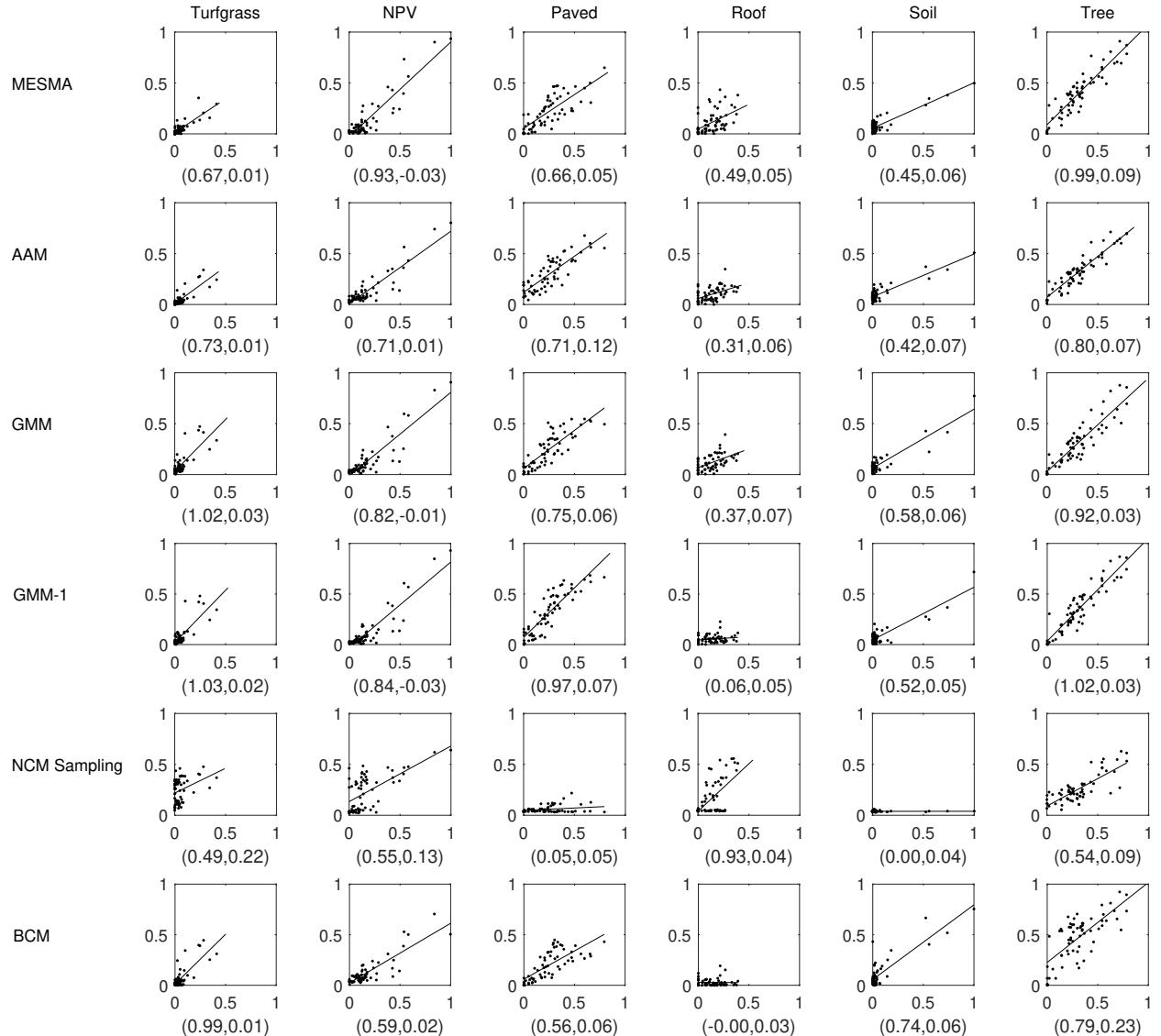


Figure 7: Scatter plots of 64 abundance values in 16 m for the ground truth (x-axis) and the estimated (y-axis). The parenthesis below the plot contains the slope and intercept.

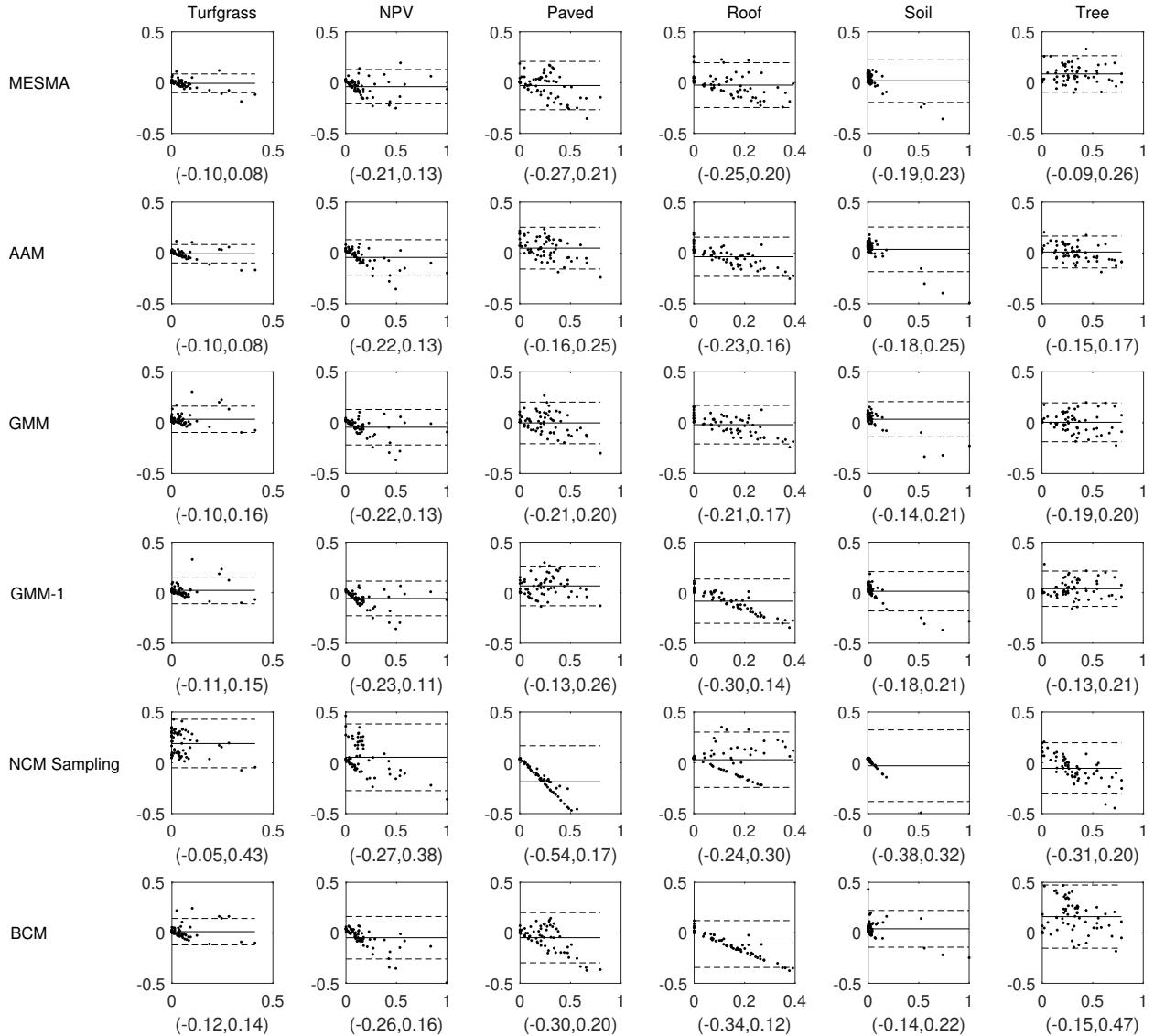


Figure 8: Bland-Altman plots of 64 abundance values for the 16 m data. The x-axis is the ground truth. The y-axis is the difference between the estimated value and the ground truth. The solid line is the mean of these differences while the dashed lines show the mean plus (minus) twice the standard deviation of these differences. The parenthesis below the plot contains the 95% limits of agreement. Dots outside the axes are not shown.

Table 4: Comparison of error and correlation coefficient for the 4 m images

		Set-based		Distribution-based			
	MAE / $R^2$	MESMA	AAM	GMM	GMM-1	NCM Sampling	BCM
Individual	Turfgrass	<b>0.028 / 0.740</b>	0.033 / 0.712	0.036 / 0.811	<b>0.027 / 0.811</b>	0.033 / 0.754	0.049 / 0.710
	NPV	<b>0.044 / 0.885</b>	0.059 / 0.874	<b>0.052 / 0.879</b>	0.052 / 0.867	0.068 / 0.872	0.074 / 0.822
	Paved	<b>0.057 / 0.859</b>	0.100 / 0.822	<b>0.061 / 0.834</b>	0.075 / 0.794	0.165 / 0.847	0.089 / 0.607
	Roof	<b>0.060 / 0.574</b>	0.078 / 0.287	<b>0.069 / 0.458</b>	0.100 / 0.250	0.199 / 0.655	0.105 / 0.181
	Soil	<b>0.040 / 0.941</b>	0.065 / 0.793	0.078 / 0.904	0.084 / 0.897	<b>0.065 / 0.867</b>	0.094 / 0.734
	Tree	0.049 / 0.910	0.055 / 0.918	<b>0.043 / 0.934</b>	<b>0.039 / 0.937</b>	0.057 / 0.926	0.102 / 0.682
Merged	Average <sup>a</sup>	<b>0.046 / 0.818</b>	<b>0.065 / 0.735</b>	<b>0.056 / 0.803</b>	<b>0.063 / 0.759</b>	<b>0.098 / 0.820</b>	<b>0.086 / 0.623</b>
	GV	<b>0.039 / 0.943</b>	0.045 / 0.931	0.043 / 0.943	<b>0.042 / 0.942</b>	0.058 / 0.932	0.120 / 0.823
	Pervious	<b>0.053 / 0.935</b>	0.091 / 0.858	<b>0.075 / 0.916</b>	0.082 / 0.906	0.087 / 0.886	0.078 / 0.897
	Impervious	<b>0.054 / 0.943</b>	0.072 / 0.817	<b>0.067 / 0.921</b>	0.075 / 0.911	0.092 / 0.895	0.151 / 0.791
	Average <sup>b</sup>	<b>0.049 / 0.941</b>	<b>0.069 / 0.868</b>	<b>0.062 / 0.927</b>	<b>0.067 / 0.919</b>	<b>0.079 / 0.904</b>	<b>0.116 / 0.837</b>

<sup>a</sup> the average RMSEs in the individual case for the 6 methods are 0.064, 0.092, 0.074, 0.082, 0.123, 0.113 respectively.

<sup>b</sup> the average RMSEs in the merged case for the 6 methods are 0.063, 0.107, 0.075, 0.08, 0.106, 0.142 respectively.

<sup>c</sup> the entries in red denote the best two results in each category.

<sup>d</sup> the running time on all the images for the 6 methods is 165, 499, 213, 8, 3153, 1347 minutes respectively.

#### 4.1.2. 4 m Case

The error statistics of 4 m data are shown in Table 4 (see Supplementary Fig. S1 - S64 for abundance maps). MESMA and GMM are the most accurate, with AAM following. Though MESMA and AAM take the same reduce library and aim for the same goal, they have some differences as shown in Table 2. On the other hand, AAM and GMM both place no limit on the number of endmembers nor use a shade endmember. Hence, the slightly better performance of GMM over AAM is more interesting since they use the same assumption. Similar to the 16 m case, merging the abundances in general does not decrease the error, except for NCM sampling. We observe that NCM sampling tends to have large errors for paved and roof. However, the sum of paved and roof, i.e. impervious, has a much lower error. This suggests that NCM sampling is confused by these two similar endmember classes.

Fig. 9 shows the Bland-Altman plots for this data (see Supplementary Fig. S65 for the scatter plots). Similar to the statistics in Table 4, GMM and the two set-based methods seem to work better. The other 3 methods all have some boundary value exceeding 0.25 (in absolute value). The advantage over them is mainly on paved and roof. Recall the large variation and overlap in these two endmember classes in the original library in Fig. 3, unimodal distributions may have confusion over them which leads to low accuracy. This time MESMA has the best negative boundary value (-0.18) compared to AAM (-0.25) and GMM (-0.20). It also has the best positive boundary value (0.15). Hence we can expect MESMA overestimates/underestimates the abundance by at most 0.18 with 95% probability. For comparison, this statistic is 0.2 for GMM and 0.25 for AAM.

#### 4.1.3. Test of difference significance

We used paired-sample  $t$ -test on the mean MAEs over all classes per image (64 samples) to determine if the difference is significant for two methods. Before proceeding, we need to verify that the error samples come from a Gaussian distribution. This can be visualized in a quantile-quantile plot, where the quantiles of the samples are plotted against the quantiles from the standard Normal distribution. If the pairs of quantiles fit a straight line, it means that the samples has a Gaussian distribution. Fig. 10 shows the quantile-quantile plots for all the methods on the 16 m data. We see that except some outliers, most of them fit the straight line well. Another more quantitative way is to use the chi-square goodness-of-fit test. It returns a test decision for the null hypothesis that the samples are from a Gaussian distribution with mean and variance estimated from the samples. If the  $p$ -value is less than a significance level (e.g.  $p < 0.05$ ), the null hypothesis is rejected, i.e. they are not from a Gaussian distribution. The  $p$ -values are shown below the plots in Fig. 10. We see that they are all above 0.05. For the 4 m data, the  $p$ -values are 0.16, 0.21, 0.14, 0.16, 0.31, 0.34 respectively, suggesting a weak evidence to reject the null hypothesis.

In the  $t$ -test for two sources of errors, we have a null hypothesis that the error samples come from normal distributions with equal means. We say two results are significantly different if the null hypothesis is rejected at the 5% significance level ( $p < 0.05$ ). Table 5 shows the  $p$ -values for pairwise comparison. We see that in the 16 m case, MESMA, AAM and GMM all have no significant difference between each other. Given the errors in Table 3, the ranking in terms of accuracy is MEMSA  $\approx$  AAM  $\approx$  GMM  $>$  GMM-1  $>$  BCM  $>$

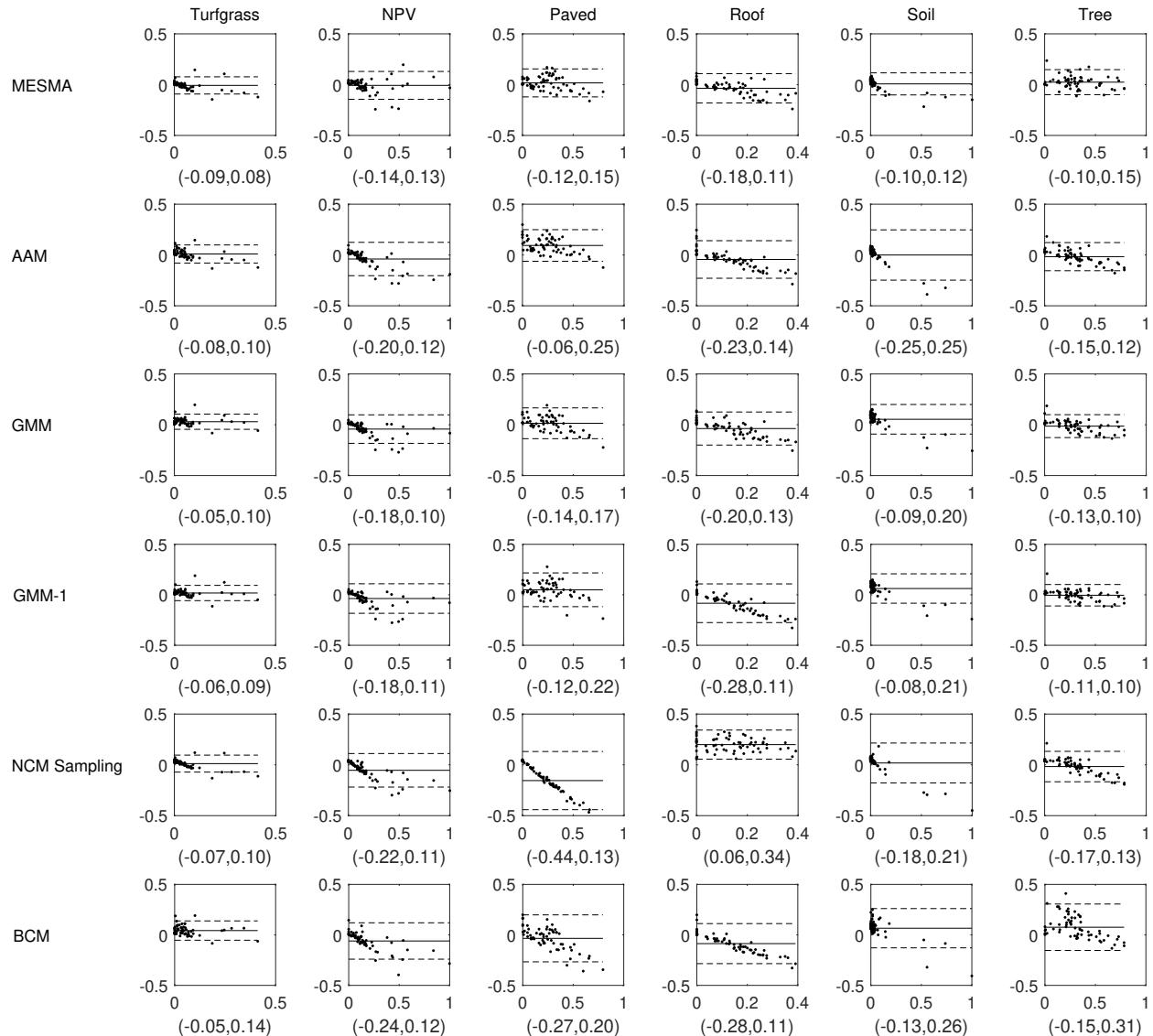


Figure 9: Bland-Altman plots of 64 abundance values for the 4 m data. The parenthesis below the plot contains the 95% limits of agreement. Dots outside the axes are not shown.

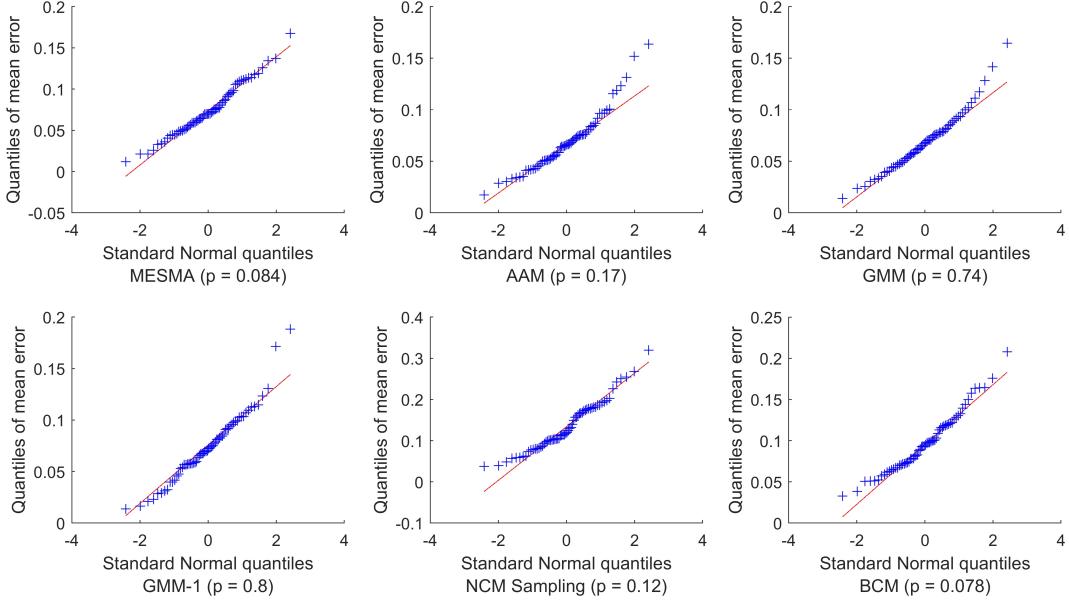


Figure 10: Quantile-quantile plots for all the methods on the 16 m data.

Table 5:  $P$ -values for significance of difference between any two methods

$p$ -value	Method	MESMA	AAM	GMM	GMM-1	NCM	BCM
16 m	MESMA	-	<b>0.16</b>	<b>0.25</b>	<b>0.95</b>	2e-12	7e-6
	AAM	<b>0.16</b>	-	<b>0.98</b>	<b>0.07</b>	4e-12	3e-7
	GMM	<b>0.25</b>	<b>0.98</b>	-	0.04	5e-11	3e-7
	GMM-1	<b>0.95</b>	<b>0.07</b>	0.04	-	3e-9	4e-5
	NCM	2e-12	4e-12	5e-11	3e-9	-	2e-5
	BCM	7e-6	3e-7	3e-7	4e-5	2e-5	-
4 m	MESMA	-	8e-6	7e-5	3e-8	3e-17	2e-14
	AAM	8e-6	-	0.02	<b>0.57</b>	2e-7	1e-4
	GMM	7e-5	0.02	-	1e-4	3e-14	2e-12
	GMM-1	3e-8	<b>0.57</b>	1e-4	-	7e-10	1e-7
	NCM	3e-17	2e-7	3e-14	7e-10	-	5e-4
	BCM	2e-14	1e-4	2e-12	1e-7	5e-4	-

<sup>a</sup> the entries in bold denote no significant difference between the two comparing methods ( $p > 0.05$ ).

NCM sampling. In the 4 m case, only GMM-1 and AAM have no significant difference, hence the ranking is MESMA > GMM > GMM-1  $\approx$  AAM > BCM > NCM sampling (Table 4).

#### 4.1.4. Efficiency

Since they were run on different machines with different implementation (MESMA and NCM sampling have multiple threads), the time costs are not for comparison, but for reference. In general, all the methods run in a few hours. The fastest method is GMM-1, which is our implemented GMM with only one combination. NCM sampling turns out to be the slowest one. It is expected since sampling algorithms are usually slower than deterministic algorithms. The time costs on the 4 m dataset are usually more than 10 times slower than those on the 16 m dataset. This is because the image size of the former is 16-19 times larger than the latter and the library size is also larger. The least gap comes from GMM because of a changed  $T_{CVIC}$  leading to a significantly less number of combinations. Heylen et al. (2016) showed that with the same implementation AAM was much faster than MESMA, but here the result is converse. One reason is that the parameters of MESMA force it to pick at most 3 endmembers for a pixel instead of all the combinations. Also, multi-threading and implementation techniques impact the real world time costs significantly.

#### 4.2. Extend to semi-realistic images

We extended the experiments to semi-realistic images to check if the method or library reduction was overfitted to this particular dataset. To this end, we tested the methods on another batch of synthetic images



Figure 11: Simulated 4 m ROI images. They are very similar to the real images in Fig. 1.

Table 6: Comparison of error and correlation coefficient for the synthetic images

		Set-based		Distribution-based				
	MAE / $R^2$	MESMA	AAM	GMM	GMM-1	NCM Sampling	BCM	
16 m	Turfgrass	0.044 / <b>0.791</b>	0.050 / <b>0.837</b>	<b>0.021 / 0.913</b>	0.025 / 0.916	<b>0.016 / 0.954</b>	0.042 / 0.827	
	NPV	0.039 / <b>0.948</b>	0.027 / <b>0.940</b>	<b>0.023 / 0.963</b>	0.026 / 0.956	<b>0.018 / 0.978</b>	0.041 / 0.848	
	Paved	<b>0.074 / 0.512</b>	0.115 / 0.563	0.078 / 0.745	0.116 / 0.598	<b>0.038 / 0.927</b>	0.101 / 0.307	
	Roof	0.063 / 0.441	0.063 / 0.569	<b>0.054 / 0.742</b>	0.118 / 0.259	<b>0.027 / 0.894</b>	0.129 / 0.170	
	Soil	0.038 / 0.643	0.033 / 0.723	0.028 / 0.738	<b>0.028 / 0.781</b>	<b>0.016 / 0.956</b>	0.067 / 0.604	
	Tree	0.090 / 0.882	0.058 / 0.822	<b>0.051 / 0.880</b>	0.060 / 0.923	<b>0.029 / 0.960</b>	0.179 / 0.669	
	<b>Average<sup>a</sup></b>	<b>0.058 / 0.703</b>	<b>0.058 / 0.742</b>	<b>0.042 / 0.830</b>	<b>0.062 / 0.739</b>	<b>0.024 / 0.945</b>	<b>0.093 / 0.571</b>	
4 m	Turfgrass	0.032 / 0.838	0.019 / 0.912	0.021 / 0.921	<b>0.015 / 0.921</b>	<b>0.014 / 0.919</b>	0.019 / 0.896	
	NPV	0.040 / 0.955	0.016 / 0.965	<b>0.012 / 0.990</b>	<b>0.013 / 0.991</b>	0.019 / 0.995	0.027 / 0.955	
	Paved	0.062 / 0.836	0.121 / 0.819	0.070 / 0.773	0.103 / 0.742	<b>0.041 / 0.890</b>	<b>0.053 / 0.723</b>	
	Roof	0.048 / 0.478	0.051 / 0.537	<b>0.046 / 0.472</b>	0.087 / 0.254	<b>0.035 / 0.617</b>	0.088 / 0.222	
	Soil	0.043 / 0.661	0.042 / 0.845	<b>0.026 / 0.840</b>	<b>0.028 / 0.813</b>	0.036 / 0.953	0.041 / 0.766	
	Tree	0.050 / 0.906	<b>0.035 / 0.911</b>	0.055 / 0.864	0.045 / 0.880	<b>0.033 / 0.951</b>	0.099 / 0.748	
	<b>Average<sup>b</sup></b>	<b>0.046 / 0.779</b>	<b>0.047 / 0.832</b>	<b>0.038 / 0.810</b>	<b>0.049 / 0.767</b>	<b>0.030 / 0.887</b>	<b>0.054 / 0.718</b>	

<sup>a</sup> the average RMSEs in 16 m for the 6 methods are 0.080, 0.077, 0.060, 0.084, 0.031, 0.12 respectively.

<sup>b</sup> the average RMSEs in 4 m for the 6 methods are 0.065, 0.063, 0.054, 0.066, 0.039, 0.071 respectively.

<sup>c</sup> the entries in red denote the best two results in each category.

generated by the library spectra. Since all the methods assumed that a pixel was a linear combination of endmember spectra from the library, the creation of this synthetic dataset follows this assumption.

We created this dataset following the literature that emphasizes realistic simulation (Gao et al., 2013; Hao et al., 2015). For each ROI in the original dataset, we randomly sampled spectra from the full library with the number of spectra for each class being the same as that in the reduced library. Then we used AAM to unmix the pixels with these sampled spectra. The obtained abundances were sorted to keep the largest three while the other were set to 0, and rescaled such that their summation was one. This is to conform with the assumption of MESMA. The endmembers and abundances were combined according to the LMM to generate synthetic pixels. In this way, we can create a dataset where the endmembers are randomly picked from the full library, and the spatial distribution of abundances looks similar to the original one. Fig. 11 shows all the 4 m synthetic images generated in this way. Comparing it with Fig. 1, we can see its similarity. But inherently, the synthetic images follow exactly the LMM with at most 3 endmembers for each pixel and they are randomly picked from the full library.

We validated all the methods on this simulated dataset. Table 6 shows the unmixing results on this dataset. We see that GMM and NCM sampling turn out to be the best two methods. The superior performance of NCM sampling contrasts sharply to its worst result in Table 3. Since we evaluate the difference between total abundances in an area for a class, it is possible that the relatively large pixel abundance error is mitigated by averaging them. This is more possible for sampling algorithms because statistically they tend to sample values around the expected value.

Compared to the superior results from set-based methods in Section 4.1, the advantage of distribution-

based methods becomes obvious. Since the endmembers were randomly sampled from the big library, set-based methods showed insufficiency to unmix the pixels using a reduced library derived from another dataset. It is possible that a different reduced library based on this simulated dataset may lead to better results for set-based methods. However, that means, to maximize the performance, they need repetitive library reduction when applied to different sites.

## 5. Discussion

### 5.1. Sources of errors and difficulty in urban application

The reported errors come from three sources: dataset, model and method. First, the dataset was well developed on various scenes, but the validated abundances have intrinsic errors from the same UTM coordinates that were applied to the two kinds of images for region correspondence. Because these airborne images are collected from an unstable process, the coordinates, though spatially calibrated, may still not correspond to the precise locations accurately. Using a polygon and validating its total abundances mitigate this effect, but still the region for reference could have a small shift compared to the region for unmixing. This is more likely in the 16 m data, which may explain partly the larger overall errors in this data. This also supports our choice of MAE as the error measure over RMSE because the  $L_2$  norm of RMSE is more sensitive to outliers than the  $L_1$  norm of MAE. A possible route to reduce this effect is to find a pixel-wise correspondence by image-based co-registration (Zhou et al., 2017b).

Second, the LMM may not hold in real datasets even by assuming endmember variability. The LMM assumption comes from the checkerboard interpretation, where the area of each material contributes to the formation as coefficients. This holds when the areas of different materials are segregated and there is no interaction between different materials. If the spatial structure is not flat such that the light ray undergoes multiple reflections among these materials, simple linear mixing is violated and a more complicated nonlinear model may be used (Ray and Murray, 1996). Allowing the endmember spectra to vary per pixel based on a proper spectral library could mitigate this effect. But the modeling error still exists to some extent and it is common in urban applications where buildings and tree canopies easily create multiple reflections.

Moreover, the application of LMM is also undermined by the spectral ambiguity among different urban categories (as seen in Fig. 3). For example, the endmember classes of roof and paved sometimes use the same actual material, such as bitumen. Also, shadow along buildings can reduce the spectral reflectance when the coarse pixels contain the area. These shadowed roof and paved materials have further resemblance to shadowed trees. On the other hand, cars with large albedo on the road can increase the reflectance of paved, making it similar to some roof material. In the Results section, the poor performance on roof and paved from unimodal distribution-based methods could be caused by this library overlap. Note that its influence on the set-based methods is not that large due to spectral separability from the reduced library. Since this ambiguity reflects the fact that different classes share some common spectral signatures, we kept it for distribution-based methods.

Lastly, a method has intrinsic errors from itself. For set-based methods, since they can explore the small combination space to find the global optimal solution, the errors are more likely to be from its library reduction scheme. Hence it is critical to find a reduced subset that conforms to the LMM on the unmixing pixels. As the dataset grows large, it is more and more difficult to find a reduced subset that is suitable for increasing pixels. For distribution-based methods, the problem is solved by iterative numerical algorithms, thus they may only approach a local optimal solution in the continuous solution space. They also depend on the library as they assume the endmembers follow a distribution on the library. If the distribution does not reflect the actual endmembers in the image, bias will exist in the estimation. Conceptually, this issue is equivalent to the library reduction problem in set-based methods if we view the latter as a special case of GMM.

### 5.2. Comparison with respect to spatial resolution

The 16 m dataset has large overall errors compared to the 4 m counterpart. This phenomenon is commonly seen in the previous research (Okujeni et al., 2015, 2013; Wetherley et al., 2017). There are many reasons behind it. One is the possible large co-registration error in the 16 m data. Another reason is that a large polygon size in terms of pixel could mitigate co-registration error and average overestimated and

Table 7: Average pixel complexity ( $\mathcal{C}$ ) and diversity ( $\mathcal{D}$ )

		MESMA	AAM	GMM	GMM-1	NCM Sampling	BCM
$\mathcal{C}$	16 m	2.07	4.13	3.90	3.53	5.88	3.26
	4 m	1.80	3.86	3.90	3.94	5.90	2.84
$\mathcal{D}$	16 m	1.85	2.94	2.57	2.28	2.33	2.39
	4 m	1.56	2.45	2.35	2.24	2.54	2.15

underestimated abundances (Powell et al., 2007). Since the physical polygon size is 180 m by 180 m (11 or 12 pixels per side for 16 m and 46 or 50 pixels per side for 4 m), the more pixels for 4 m will lead to less error. Finally, a large pixel size implies complicated mixing involving more endmembers, which makes unmixing more difficult. A common route to evaluate how many endmembers exist in a pixel is to count the number of abundances greater than 0, which is called *pixel complexity* ( $\mathcal{C}$ ). Table 7 shows the average pixel complexity for each method. According to MESMA, 24% pixels comprise one endmember, 44% pixels comprise two endmembers for 16 m data, while for 4 m data the proportions are 54% and 10% respectively. Considering this effect, the comparison within the paper or beyond will focus on similar spatial resolutions.

A problem with the calculation of pixel complexity is that if a method has abundances close to zero but not exactly equal to zero, the complexity will be quite different from another with similar abundances and exactly zero values. In other words, a slight perturbation to the abundances from MESMA will lead to a very different complexity though the changes are small. This is seen in Table 7 where the other methods all have a large pixel complexity since they use numerical algorithms without any constraint on the maximal number of endmembers. To measure the pixel composition more consistently, we propose to use another measure, *diversity* ( $\mathcal{D}$ ), which is commonly used in information theory, ecology and physics (Jost, 2006). The definition of diversity is the exponential of the entropy of the abundance distribution,  $\exp\left(\sum_j \alpha_{nj} \log \alpha_{nj}\right)$ . Its connection to complexity is that if the abundances are equally distributed, then the diversity is equal to the number of endmembers (nonzero abundances). The advantage of diversity is that when the abundances transition from  $M - 1$  equally distributed nonzero values to  $M$  equally distributed nonzero values, the transition of diversity is a smooth curve connecting  $M - 1$  and  $M$ . Hence, it is more suitable for a wide range of methods. Table 7 also shows the diversity of different results. We see that in average, a pixel contains 2 - 3 endmembers for most of the methods. The similar values on pixel diversity and complexity from MESMA suggest that the new measure has a similar interpretation of the pixel composition. Under this measure, the interpretation from distribution-based methods is similar to MESMA.

### 5.3. Comparison between synthetic and real datasets

Comparing the real dataset and the synthetic dataset, the accuracy difference for MESMA is quite small compared to the other methods. MESMA achieves average MAE 0.074 (real data) versus 0.058 (synthetic data) for the 16 m case, the same 0.046 for the 4 m case respectively. Considering the source of error analysis in Section 5.1, the synthetic dataset eliminates errors introduced by the LMM assumption and the co-registration error. No improvement for the 4 m data from MESMA suggests an underlying flaw to the current 2-step unmixing workflow. That is, the reduced library may be overfitted to the specific unmixing dataset. This is very likely as ICR uses MESMA as a classifier in the reduction procedure. The strongest evidence is that for all the other methods, the errors are reduced in the synthetic case. For example, this comparison for AAM is MAE 0.069 vs. 0.058 (16 m), 0.065 vs. 0.047 (4 m), for GMM is MAE 0.069 vs. 0.042 (16 m), 0.056 vs. 0.038 (4 m), for GMM-1 is MAE 0.075 vs. 0.062 (16 m), 0.063 vs. 0.049 (4 m).

Another evidence is that in the real data experiments, MESMA and AAM have no significant difference in the 16 m data ( $p = 0.16$ ) while MESMA is significantly better than AAM in the 4 m data (MAE = 0.046 vs. MAE = 0.065,  $p = 8e - 6$ ). Recall that AAM also uses the reduced library and tries to find the optimal combination. Its difference compared to MESMA lies in the details, including no inherent shade endmember, a different optimization approach, and no constraint on the maximal number of endmembers in the model. These are all possible reasons that may lead to the success of MESMA in the 4 m case. However, we may also explain that the 4 m reduced library is especially designed to work best under MESMA for the specific 4 m data. This conjecture is supported by the additional fact that in the synthetic data experiment, MESMA and AAM have similar accuracy for both scales data (MAE = 0.058 vs. MAE = 0.058 for 16 m,

$\text{MAE} = 0.046$  vs.  $\text{MAE} = 0.047$  for 4 m). We also find no significant difference between them on this dataset ( $p = 0.89$  for 16 m and  $p = 0.59$  for 4 m). This suggests that MESMA and AAM may work equally well when we change to another batch of images given the same reduced library.

Based on this analysis, the current unmixing accuracy is highly dependent on the reduced library. This shifts the focus from the actual unmixing methodology to finding a reduced library that fits the unmixing method. Given a universal spectral library, a direct consequence is that when we apply unmixing to different cities, we need to repeatedly reduce the large library to ensure good performance. This complicates the process and also adds uncertainty to the results, especially for reduction methods that require manual guidance.

#### 5.4. Comparison with respect to different methods

Among the set-based methods, MESMA is slightly inferior to AAM for the 16 m data ( $\text{MAE} = 0.074$  vs.  $\text{MAE} = 0.069$ ,  $p = 0.16$ ) while significantly better for the 4 m data ( $\text{MAE} = 0.046$  vs.  $\text{MAE} = 0.065$ ,  $p = 8e - 6$ ). Hence MESMA can represent the best performance for set-based methods in this paper. Among the distribution-based methods, GMM performed best for both the 16 m and 4 m data, with the least average error ( $\text{MAE} = 0.069$  for 16 m,  $\text{MAE} = 0.056$  for 4 m). The following best method is GMM-1 ( $\text{MAE} = 0.075$  for 16 m and  $\text{MAE} = 0.063$  for 4 m), which is our NCM with the same implementation as GMM. The difference between GMM and GMM-1 is significant for both scales data ( $p < 0.05$ ). For the remaining two methods, BCM has better accuracy than NCM sampling, while their errors are noticeable larger than the former two ( $p < 0.05$  when against them on both real data). It is unsurprising to see that GMM has a superior performance as the other distribution-based methods can not model multiple modes in the distribution of the library spectra (Zhou et al., 2018) (GMM vs. NCM/BCM is analogous to MESMA vs. traditional unmixing with a fixed endmember combination). We will use GMM to represent the best method for distribution-based methods here.

Comparing the two exemplary methods on the real dataset, GMM is slightly better than MESMA for the 16 m data ( $\text{MAE} = 0.069$  vs.  $\text{MAE} = 0.074$ ,  $p = 0.25$ ) while inferior for the 4 m data ( $\text{MAE} = 0.056$  vs.  $\text{MAE} = 0.046$ ,  $p = 7e - 5$ ). At first glance we may conclude that GMM and MESMA work equally well for the 16 m data while MESMA is better at handling high-resolution data. However, Table 4 also shows that GMM is better than AAM for the 4 m data ( $\text{MAE} = 0.056$  vs.  $\text{MAE} = 0.065$ ,  $p = 0.02$ ). According to the analysis in Section 5.3, the reduced library may be overfitted to MESMA, the error from MESMA could be the extreme limit for this data since the error could not be reduced in the ideal case (synthetic data). A further hypothesis suggests that MESMA and AAM may work equally well on other datasets. Under this assumption, the superior performance of GMM over AAM for 4 m implies an opposite direction to our initial conclusion, i.e. GMM may have better unmixing capability for fine scale data than MESMA with a routinely reduced library. Whether this argument is true depends on many conditions (library reduction, parameter tuning, etc.), but the initial attempt on the synthetic dataset has shown this trend ( $\text{MAE} = 0.042$  vs.  $\text{MAE} = 0.058$  and  $p = 2e - 4$  for 16 m,  $\text{MAE} = 0.038$  vs.  $\text{MAE} = 0.046$  and  $p = 0.01$  for 4 m).

#### 5.5. Comparison with results reported in other literature

Here we give some remarks on the comparison with results reported in other work. First, the results of MESMA are slightly different from those reported in (Wetherley et al., 2017), which used a similar dataset. No MAE was reported in (Wetherley et al., 2017), but they calculated the average  $R^2$  values, which are 0.642 (individual) and 0.867 (merged) for 18 m, 0.811 (individual) and 0.923 (merged) for 4 m. Based on the small difference in  $R^2$  value by comparing them to Table 3 and Table 4 ( $R^2 = 0.660$  and 0.836 for 16 m,  $R^2 = 0.818$  and 0.941 for 4 m), we conclude that the two MESMAs have similar results hence the comparison here can be extended to that work. Note that we used Matlab to extract the polygon ROIs directly from the original images while they resampled the images to a uniform spatial resolution of 18 m. Additionally, that study used 178 spectral bands in contrast to 164 bands used here. The less bands are derived from removing all the bad bands from all the validation polygons and spectral libraries.

Second, the set-based methods can achieve better accuracy if the reduced library is well constructed such that it contains potential spectra that conform to the LMM for each pixel. Wetherley et al. (2017) showed that when the original library contained spectra from different resolutions, the results were better ( $R^2 = 0.760$  for 18 m,  $R^2 = 0.837$  for 4 m) than those from single-resolution libraries. In (Degerickx et al., 2017), unmixing was performed on Berlin images with three sources of spectral library, Berlin, Bonn,

Brussels. It turns out that when using Berlin spectra mixed with a small portion of non-Berlin spectra, the result is slightly better than using pure Berlin spectra (RMSE = 0.13 vs RMSE = 0.14). This is expected as the larger the original library, the more possible that we may find a subset that fits the dataset under the LMM assumption.

Finally, the accuracy of GMM exceeds other work with different datasets in terms of error. In (Okujeni et al., 2013), support vector regression (SVR) was used with synthetically generated training data to unmix hyperspectral images over Berlin, which have 92 polygons with 128 bands ranging from 440 to 2500 nm and 3.6 m spatial resolution. The reported accuracy for SVR is MAE = 0.084, 0.128, 0.064, 0.068,  $R^2$  = 0.86, 0.58, 0.81, 0.85 for roof, pavement, grass, tree respectively. In comparison, GMM for 4 m data in this study has accuracy MAE = 0.069, 0.061, 0.036, 0.043,  $R^2$  = 0.46, 0.83, 0.81, 0.93 for the same materials. The same SVR framework is applied to 80 polygons with 9 m spatial resolution in a further study (Okujeni et al., 2015). The reported accuracy is MAE = 0.107, 0.203, 0.063, 0.146,  $R^2$  = 0.74, 0.73, 0.88, 0.70 for roof, pavement, vegetation and tree respectively. Compared to the 16 m results in Table 3, our error for GMM is less (MAE = 0.078, 0.081, 0.045, 0.076) though the correlation coefficient is lower ( $R^2$  = 0.279, 0.691, 0.632, 0.798). In a recent work focusing on discovering a better library reduction algorithm (72 polygons with 9 m spatial resolution), the best results using MESMA and mixed spectral libraries have RMSE = 0.11, 0.17, 0.12, 0.16, 0.1 for roof, pavement, tree, grass, soil respectively (Degerickx et al., 2017). The corresponding RMSEs from GMM in 16 m are 0.096, 0.102, 0.095, 0.072, 0.092. Note that the  $R^2$  values are highly dependent on the dispersion of the abundance values. If we only look at the error comparison, GMM turns out to be better in all cases. Note that the performance of MESMA in our study is comparable to GMM on coarse-scale data and even slightly better on fine-scale data. Comparing it to MESMA in other work, it further verifies our previous conjecture that the reduced library is overfitted to our dataset and GMM may be better than MESMA with a routinely reduced library.

### *5.6. Implications for a universal spectral library*

Distribution-based methods have profound applications in terms of a universal spectral library. With the existence of spectral libraries from different sources, a universal spectral library can be established by combining these spectra from different locations and seasons. Given such a large spectral library, GMM is ideal to encode the library without losing much information. According to the implementation of GMM, it works by initializing the abundances by finding the combination of Gaussian centers that matches the pixel, followed by an EM algorithm that refines the abundances (Zhou et al., 2018). This procedure is similar to the current 2-step workflow that first reduces the library, then uses MESMA for unmixing, with the Gaussian centers corresponding to the reduced library, the initialized abundances corresponding to the MESMA unmixed abundances, except an additional abundance refinement step. However, conceptually, GMM uses a single model to describe the unmixing task, and unifies the current workflow in the implementation.

Applying GMM to a universal spectral library is straightforward. The simplest way is to use a separate Gaussian component for the spectra from each source, though the error will probably be larger than using the spectra from the same location with a similar spatial resolution. The advantage over the current approach lies in the final EM refinement. Since library information is encoded in the distribution parameters, calculation of the abundances is not merely determined by the representing spectra from the library (Gaussian centers), but also by the covariance matrices, which encode how spectra change in the library (Fig. 6). This additional information is critical to the success of GMM over AAM in the 4 m data, since GMM uses the full library information while AAM uses a reduced library designed for MESMA. Another advantage of applying GMM is that it simplifies the library reduction procedure for each dataset (if finding the Gaussian centers is a kind of library reduction), making it become picking related spectra and estimating GMM parameters, the latter of which is a well studied problem (McLachlan and Rathnayake, 2014).

Finally, we discuss some empirical criteria for picking related spectra in a large spectral library. We used the 4 m library on the 16 m data for MESMA and GMM. But the results are as good as using the same resolution library, both achieving average MAE 0.099 and  $R^2$  below 0.6. Note that the collection interval between the two kinds of data is only about 1 week, but the 16 m data were collected by AVIRIS from 20 km above the ground while the 4 m data were collected by AVIRIS-NG from a 5 km altitude. A similar result is found elsewhere (Degerickx et al., 2017; Wetherley et al., 2017). Especially, in (Degerickx et al., 2017) even if the mixed library comes from several sources, as long as it does not contain the original location related spectra, the result is worse than using a single library. This implies that when applying a universal

spectral library, the spectra related to the task (same location, same spatial resolution, similar time) should be selected as a major source.

### 5.7. Significance for environmental applications

Before conclusion, we discuss the limitations and significance in terms of higher level environmental applications. Note that hyperspectral imagery is not the only source to obtain land cover fractions (Zhang and Weng, 2016). It has following limitations. First, hyperspectral imagery typically has a large number of mixed pixels from a variety of different surface materials. Second, the spectral ambiguity between land cover types often confuses the analyzing method. This ambiguity comes from both the intra-class variability due to changing illumination condition, shadow and object composition, and inter-class similarity due to the same material applied to different surface types. Third, vertical structures in urban areas cause problems such as obscured objects and complex physical models (Gastellu-Etchegorry et al., 2015). However, these limitations apply more or less to all remote sensing techniques.

In comparison, hyperspectral imagery still has an advantage of a wide spectral range and a high resolution that provide additional details to the observed surface, which leads to many urban environmental applications (Van der Linden et al., 2019), including land cover mapping (Roberts et al., 1998), assessing vegetation condition (Degerickx et al., 2018), modeling surface temperature (Deng and Wu, 2013). These applications are associated with more significant environmental problems such as global climate change (Feddeema et al., 2005; Kalnay and Cai, 2003). Hence, an imaging spectrometer has become an important instrument for monitoring environment and climate change, as mentioned in the 2007 National Research Council (NRC) Decadal Survey: “A hyperspectral sensor combined with a multispectral thermal sensor in low Earth orbit (LEO) is part of an integrated mission concept described in Parts I and II that is relevant to several panels, especially the climate variability panel” (Board et al., 2007). With the advent of orbital spectrometers (e.g. HyspIRI, EnMAP) that feature global area coverage and high temporal frequency, real-time monitoring of urban environment using hyperspectral imagery is becoming a reality (Lee et al., 2015; Guanter et al., 2015).

Currently, extensive works have been done by using high-resolution (less than 4 m) airborne hyperspectral images for classification of urban land cover types (Demarchi et al., 2014; Franke et al., 2009; Melgani and Bruzzone, 2004). In this case, a high spatial resolution is critical to success (Herold et al., 2003). However, even at a 4 m spatial resolution, half of the pixels may consist of two or more endmembers (Wetherley et al., 2017). Coarse-scale hyperspectral images further increase the number of mixed pixels as well as the number of endmembers in each pixel. This brings many challenges to the future of orbital spectrometers since their spatial resolution is even worse than airborne data (e.g. 30 m). This makes unmixing an important preprocessing step to not only existing, but also future spectrometer data.

Considering the amount of data generated in the future, an efficient way to unmix the hyperspectral imagery repeatedly is to use MESMA with a universal spectral library. The accuracy improvement of GMM compared to MESMA may be small according to Table 3 and Table 4. However, the results of MESMA in this work are not repeatable on continuous processing of spaceborne imagery since careful human intervention is required to achieve them (see Section 2.2). Hence, the first and direct significance of distribution-based methods is that they feature a fully automatic processing pipeline where the computational complexity does not scale with the library size. This benefit is especially useful for orbital spectrometers. Since they are collecting data endlessly, GMM enables fully automatic real-time analysis of these data.

Besides the stable performance and computational efficiency, there is another implication of GMM. In practice, for accurate land cover mapping, hyperspectral imagery may not be used alone. Instead, multiple sources can be combined to improve the accuracy. One source is the multispectral image, which can be combined with a hyperspectral image to produce a spatially and spectrally high-resolution image, a.k.a. hyperspectral-multispectral fusion (Zhou et al., 2019; Yokoya et al., 2017). The combined high-resolution pixels can be classified to provide accurate mapping of urban areas. Another source is the lidar data, which provides height information above the ground. Combining lidar and hyperspectral images has been shown to improve the performance (Degerickx et al., 2019; Chen et al., 2018; Priem and Canters, 2016). Despite these new directions that may be more useful in practice, they still require the hyperspectral imagery and the underlying assumption is still the LMM. For example, many hyperspectral-multispectral fusion approaches rely on the LMM with a fixed set of endmembers (Zhukov et al., 1999; Yokoya et al., 2012; Wei et al., 2015), combining lidar and hyperspectral images for land cover mapping still relies on MESMA (Degerickx et al., 2019). Hence, if GMM provides more stable and accurate unmixing results, these new directions may also

benefit by replacing the LMM with the probabilistic formulation. For example, height information derived from lidar data serves as an endmember model constraint and a fraction constrain for MESMA in Degerickx et al. (2019). The same constraints can be unified as an abundance constraint and replace the positivity and sum-to-one constraint in Eq. (3) for GMM. Therefore, the significance of this work is not restricted to improvement in terms of unmixing accuracy, but also includes replacement of the underpinning of many hyperspectral imagery-based applications.

## 6. Conclusion

This study aims to compare the unmixing capabilities of distribution-based methods and set-based methods by applying them on a comprehensive dataset consisting of 128 images across 2 scales, with reference abundances obtained by inspecting high-resolution images. The results show that with large libraries, the best distribution-based method (GMM) achieves comparable accuracy to MESMA on the 16 m data without need for library reduction (MAE = 0.069 vs. MAE = 0.074,  $p = 0.25$ ). For the 4 m data, though GMM is not as accurate as MESMA (MAE = 0.056 vs. MAE = 0.046,  $p = 7e-5$ ), it is better than another set-based method AAM that claims to be a re-implementation of MESMA (MAE = 0.056 vs. MAE = 0.065,  $p = 0.02$ ). Moreover, further experiments on a synthetic simulation imply that the reduced library may be overfitted to the real data for MESMA and GMM is significantly better than both the set-based methods. These results suggest that GMM is at least comparable to set-based methods on coarse-scale data and potentially more stable across datasets. Considering that land cover mapping using fine-scale data can be achieved by classification instead of unmixing, the comparison results on coarse-scale data could be more significant in practice.

To summarize, distribution-based methods have the benefit of keeping the variation patterns in the spectral library. Since they keep more information from the library, theoretically they should work better. In practice, for a small dataset, we can always derive a subset from the library that fits the unmixing task. Due to simplicity of the set-based methods, they can behave very well given this reduced library. However, the derivation of this subset requires too much labor work and the derived reduced library is also hard to be generalized to other datasets. This study confirms that the actual performance of the distribution-based method GMM is in line with its theoretical advantage. Hence, it could have more potential for unmixing different sites with a large spectral library and facilitate the automatic processing of a wealth of spaceborne imaging spectroscopy data in the future.

## Acknowledgments

The authors wish to thank Alina Zare and Sheng Zou for providing the NCM sampling code. We wish to acknowledge funding from the NASA Earth and Space Science Fellowship Program and the Belgian Science Policy Office in the framework of the STEREO III Program—Project UrbanEARS (SR/00/307). We also thank the Jet Propulsion Laboratory for providing radiometrically calibrated, orthorectified AVIRIS imagery.

## Appendix A. Relation between MESMA and GMM

The connection between set-based methods and distribution-based methods is that the former can be seen as a special case of the latter. Specifically, given a library of spectra  $\{\boldsymbol{\mu}_{jk} : j = 1, \dots, M, k = 1, \dots, K_j\}$ , the minimization problem in Eq. (6) can be seen as a special case of GMM by choosing  $p(\mathbf{m}_{nj}|\boldsymbol{\theta}_j) = \sum_{k=1}^{K_j} \pi_{jk} \mathcal{N}(\mathbf{m}_{nj}|\boldsymbol{\mu}_{jk}, \boldsymbol{\Sigma}_{jk})$  where  $\pi_{jk} = 1/K_j$ ,  $\boldsymbol{\Sigma}_{jk} = \epsilon \mathbf{I}$  with  $\epsilon \rightarrow 0$ , and  $p(\mathbf{n}_n|\boldsymbol{\theta}_0) = \mathcal{N}(\mathbf{n}_n|\mathbf{0}, C\epsilon \mathbf{I})$  where  $C$  is a large constant, e.g. 1e4. Intuitively, it means that each spectrum in the library is a Gaussian component, each component has an equal probability to be chosen, and the center of the component is always chosen as  $\mathbf{m}_{nj}$ . To see why this leads to (6), recall that the MLE objective function becomes

$$L(\{\alpha_{nj}\}) = \sum_{n=1}^N \log \sum_{\mathbf{k} \in \mathcal{K}} \pi_{\mathbf{k}} \mathcal{N}(\mathbf{y}_n|\boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}})$$

where

$$\pi_{\mathbf{k}} = \prod_{j=1}^M \frac{1}{K_j}, \boldsymbol{\mu}_{n\mathbf{k}} = \sum_{j=1}^M \alpha_{nj} \boldsymbol{\mu}_{jk_j}, \boldsymbol{\Sigma}_{n\mathbf{k}} = \left( \sum_{j=1}^M \alpha_{nj}^2 + C \right) \epsilon \mathbf{I}.$$

Around the optimal  $\{\alpha_{nj}\}$ , the objective function will be dominated by the  $\mathbf{k}$  that maximizes  $\mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}})$  since the exponential term  $\exp\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})\}$  in the Gaussian density function will decay fast given a small covariance matrix, which makes the contribution from the other  $\mathbf{k}$ 's negligible and the summation operator become the maximum operator. Then the objective function becomes

$$L(\{\alpha_{nj}\}) = \sum_{n=1}^N \log \max_{\mathbf{k} \in \mathcal{K}} \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}}), \quad (\text{A.1})$$

with the constant term  $N \log \prod_{j=1}^M \frac{1}{K_j}$  ignored.

Moving the logarithm operator inside,  $-\log \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}})$  can be expanded (up to a constant term) as

$$\begin{aligned} & \log |\boldsymbol{\Sigma}_{n\mathbf{k}}| + (\mathbf{y}_n - \boldsymbol{\mu}_{n\mathbf{k}})^T \boldsymbol{\Sigma}_{n\mathbf{k}}^{-1} (\mathbf{y}_n - \boldsymbol{\mu}_{n\mathbf{k}}) \\ &= B \log \left( \sum_{j=1}^M \alpha_{nj}^2 + C \right) \epsilon + \frac{\|\mathbf{y}_n - \boldsymbol{\mu}_{n\mathbf{k}}\|^2}{\left( \sum_{j=1}^M \alpha_{nj}^2 + C \right) \epsilon}, \end{aligned}$$

which makes Eq. (A.1) almost equivalent to (6) except the covariance term. Note that  $\sum_{j=1}^M \alpha_{nj}^2$  is bounded ( $1/M \leq \sum_{j=1}^M \alpha_{nj}^2 \leq 1$ ) hence is negligible to a large  $C$ . Replacing  $\sum_{j=1}^M \alpha_{nj}^2 + C$  with  $C$  and  $\max_{\mathbf{k}} \log \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}})$  with  $\min_{\mathbf{k}} [-\log \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_{n\mathbf{k}}, \boldsymbol{\Sigma}_{n\mathbf{k}})]$ , Eq. (A.1) becomes a scaled version of (6) plus some constant terms and the original maximization problem becomes a minimization problem.

## References

- Alonso, M., Roth, K., Roberts, D., 2013. Identifying Santa Barbara's urban tree species from AVIRIS imagery using canonical discriminant analysis. *Remote sensing letters* 4, 513–521.
- Asner, G.P., Heidebrecht, K.B., 2002. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. *International Journal of Remote Sensing* 23, 3939–3958.
- Asner, G.P., Lobell, D.B., 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote sensing of environment* 74, 99–112.
- Baatz, M., Benz, U., Dehghani, S., Heynen, M., Höltje, A., Hofmann, P., Lingenfelder, I., Mimler, M., Sohlbach, M., Weber, M., et al., 2004. eCognition user guide. Definiens Imaging, Munich .
- Bishop, C.M., 2006. Pattern recognition and machine learning. Springer New York.
- Bland, J.M., Altman, D., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 327, 307–310.
- Board, S.S., Council, N.R., et al., 2007. Earth science and applications from space: national imperatives for the next decade and beyond. National Academies Press.
- Chen, J., Du, P., Wu, C., Xia, J., Chanussot, J., 2018. Mapping urban land cover of a large area using multiple sensors multiple features. *Remote Sensing* 10, 872.
- Combe, J.P., Le Mouelic, S., Sotin, C., Gendrin, A., Mustard, J., Le Deit, L., Launeau, P., Bibring, J.P., Gondet, B., Langevin, Y., et al., 2008. Analysis of OMEGA/Mars express data hyperspectral data using a multiple-endmember linear spectral unmixing model (MELSUM): Methodology and first results. *Planetary and Space Science* 56, 951–975.

- De Ridder, K., Lauwaet, D., Maiheu, B., 2015. UrbClim–A fast urban boundary layer climate model. *Urban Climate* 12, 21–48.
- Degerickx, J., Okujeni, A., Iordache, M.D., Hermy, M., van der Linden, S., Somers, B., 2017. A novel spectral library pruning technique for spectral unmixing of urban land cover. *Remote Sensing* 9, 565.
- Degerickx, J., Roberts, D.A., McFadden, J.P., Hermy, M., Somers, B., 2018. Urban tree health assessment using airborne hyperspectral and lidar imagery. *International journal of applied earth observation and geoinformation* 73, 26–38.
- Degerickx, J., Roberts, D.A., Somers, B., 2019. Enhancing the performance of Multiple Endmember Spectral Mixture Analysis (MESMA) for urban land cover mapping using airborne lidar data and band selection. *Remote sensing of environment* 221, 260–273.
- Demarchi, L., Canters, F., Cariou, C., Licciardi, G., Chan, J.C.W., 2014. Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 166–179.
- Deng, C., Wu, C., 2013. Examining the impacts of urban biophysical compositions on surface urban heat island: A spectral unmixing and thermal mixing approach. *Remote Sensing of Environment* 131, 262–274.
- Drumetz, L., Meyer, T.R., Chanussot, J., Bertozzi, A.L., Jutten, C., 2019. Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms. *IEEE Transactions on Image Processing*, 1–1.
- Drumetz, L., Veganzones, M.A., Henrot, S., Phlypo, R., Chanussot, J., Jutten, C., 2016. Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. *IEEE Transactions on Image Processing* 25, 3890–3905.
- Du, X., Zare, A., Gader, P.D., Dranishnikov, D., 2014. Spatial and spectral unmixing using the Beta compositional model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 1994–2003.
- Eches, O., Dobigeon, N., Mailhes, C., Tourneret, J.Y., 2010a. Bayesian estimation of linear mixtures using the normal compositional model: Application to hyperspectral imagery. *IEEE Trans. on Image Processing* 19, 1403–1413.
- Eches, O., Dobigeon, N., Tourneret, J.Y., 2010b. Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm. *IEEE Journal of Selected Topics in Signal Processing* 4, 582–591.
- Fan, F., Deng, Y., 2014. Enhancing endmember selection in multiple endmember spectral mixture analysis (MESMA) for urban impervious surface area mapping using spectral angle and spectral distance parameters. *International Journal of Applied Earth Observation and Geoinformation* 33, 290–301.
- Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., Washington, W.M., 2005. The importance of land-cover change in simulating future climates. *Science* 310, 1674–1678.
- Franke, J., Roberts, D.A., Halligan, K., Menz, G., 2009. Hierarchical multiple endmember spectral mixture analysis (mesma) of hyperspectral imagery for urban environments. *Remote Sensing of Environment* 113, 1712–1723.
- Gao, L., Du, Q., Zhang, B., Yang, W., Wu, Y., 2013. A comparative study on linear regression-based noise estimation for hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6, 488–498.
- García-Haro, F., Sommer, S., Kemper, T., 2005. A new tool for variable multiple endmember spectral mixture analysis (VMESMA). *International Journal of Remote Sensing* 26, 2135–2162.

- Gastellu-Etchegorry, J.P., Yin, T., Lauret, N., Cajgfinger, T., Gregoire, T., Grau, E., Feret, J.B., Lopes, M., Guilleux, J., Dedieu, G., et al., 2015. Discrete anisotropic radiative transfer (DART 5) for modeling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes. *Remote Sensing* 7, 1667–1701.
- Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., et al., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing* 7, 8830–8857.
- Halimi, A., Dobigeon, N., Tourneret, J.Y., 2015. Unsupervised unmixing of hyperspectral images accounting for endmember variability. *IEEE Trans. on Image Processing* 24, 4904–4917.
- Hao, Z., Berman, M., Guo, Y., Stone, G., Johnstone, I., 2015. Semi-realistic simulations of natural hyperspectral scenes, in: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE. pp. 1004–1007.
- Herold, M., Gardner, M.E., Roberts, D.A., 2003. Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and remote sensing* 41, 1907–1919.
- Herold, M., Roberts, D., 2005. Spectral characteristics of asphalt road aging and deterioration: implications for remote-sensing applications. *Applied Optics* 44, 4327–4334.
- Herold, M., Roberts, D.A., Gardner, M.E., Dennison, P.E., 2004. Spectrometry for urban area remote sensing - development and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment* 91, 304–319.
- Heylen, R., Parente, M., Gader, P.D., 2014. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 1844–1868.
- Heylen, R., Zare, A., Gader, P., Scheunders, P., 2016. Hyperspectral unmixing with endmember variability via alternating angle minimization. *IEEE Transactions on Geoscience and Remote Sensing* 54, 4983–4993.
- Huang, C., Townshend, J.R., Liang, S., Kalluri, S.N., DeFries, R.S., 2002. Impact of sensor's point spread function on land cover characterization: assessment and deconvolution. *Remote Sensing of Environment* 80, 203–212.
- Iordache, M.D., Bioucas-Dias, J.M., Plaza, A., 2011. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 49, 2014–2039.
- Jost, L., 2006. Entropy and diversity. *Oikos* 113, 363–375.
- Kalnay, E., Cai, M., 2003. Impact of urbanization and land-use change on climate. *Nature* 423, 528–531.
- Kotthaus, S., Smith, T.E., Wooster, M.J., Grimmond, C., 2014. Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. *ISPRS journal of photogrammetry and remote sensing* 94, 194–212.
- 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 (HyspIRI) mission and preparatory activities. *Remote Sensing of Environment* 167, 6–19.
- Van der Linden, S., Okujeni, A., Canters, F., Degerickx, J., Heiden, U., Hostert, P., Priem, F., Somers, B., Thiel, F., 2019. Imaging spectroscopy of urban environments. *Surveys in Geophysics* 40, 471–488.
- McLachlan, G.J., Rathnayake, S., 2014. On the number of components in a Gaussian mixture model. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 4, 341–355.
- Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing* 42, 1778–1790.
- Murphy, K.P., 2012. Machine learning: a probabilistic perspective. MIT press.

- Okujeni, A., van der Linden, S., Hostert, P., 2015. Extending the vegetation–impervious–soil model using simulated enmap data and machine learning. *Remote Sensing of Environment* 158, 69–80.
- 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 Sensing of Environment* 137, 184–197.
- Powell, R.L., Roberts, D.A., Dennison, P.E., Hess, L.L., 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment* 106, 253–267.
- Priem, F., Canters, F., 2016. Synergistic use of LiDAR and APEX hyperspectral data for high-resolution urban land cover mapping. *Remote sensing* 8, 787.
- Rashed, T., Weeks, J.R., Roberts, D., Rogan, J., Powell, R., 2003. Measuring the physical composition of urban morphology using multiple endmember spectral mixture models. *Photogrammetric Engineering & Remote Sensing* 69, 1011–1020.
- Ray, T.W., Murray, B.C., 1996. Nonlinear spectral mixing in desert vegetation. *Remote sensing of environment* 55, 59–64.
- 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 Sensing of Environment* 65, 267–279.
- Roberts, D.A., Quattrochi, D.A., Hulley, G.C., Hook, S.J., Green, R.O., 2012. Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the hyperspectral infrared imager (HyspIRI) decadal survey mission. *Remote Sensing of Environment* 117, 83–101.
- Schaaf, A.N., Dennison, P.E., Fryer, G.K., Roth, K.L., Roberts, D.A., 2011. Mapping plant functional types at multiple spatial resolutions using imaging spectrometer data. *GIScience & Remote Sensing* 48, 324–344.
- Settle, J., Drake, N., 1993. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing* 14, 1159–1177.
- Small, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis. *International journal of remote sensing* 22, 1305–1334.
- Smyth, P., 2000. Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing* 10, 63–72.
- Somers, B., Asner, G.P., Tits, L., Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment* 115, 1603–1616.
- Song, C., 2005. Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability? *Remote Sensing of Environment* 95, 248–263.
- Stein, D., 2003. Application of the normal compositional model to the analysis of hyperspectral imagery, in: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, pp. 44–51.
- Tang, W., Shi, Z., Wu, Y., Zhang, C., 2015. Sparse unmixing of hyperspectral data using spectral a priori information. *IEEE Transactions on Geoscience and Remote Sensing* 53, 770–783.
- Vane, G., Green, R.O., Chrien, T.G., Enmark, H.T., Hansen, E.G., Porter, W.M., 1993. The airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment* 44, 127–143.
- Wang, Z.M., Batelaan, O., De Smedt, F., 1996. A distributed model for water and energy transfer between soil, plants and atmosphere (wetspa). *Physics and Chemistry of the Earth* 21, 189–193.
- Wei, Q., Dobigeon, N., Tourneret, J.Y., 2015. Fast fusion of multi-band images based on solving a sylvester equation. *IEEE Transactions on Image Processing* 24, 4109–4121.

- Wetherley, E.B., Roberts, D.A., McFadden, J.P., 2017. Mapping spectrally similar urban materials at sub-pixel scales. *Remote Sensing of Environment* 195, 170–183.
- Wu, C., Deng, C., Jia, X., 2014. Spatially constrained multiple endmember spectral mixture analysis for quantifying subpixel urban impervious surfaces. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 1976–1984.
- Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine* 5, 29–56.
- Yokoya, N., Yairi, T., Iwasaki, A., 2012. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing* 50, 528–537.
- Zare, A., Gader, P., Casella, G., 2013. Sampling piecewise convex unmixing and endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing* 51, 1655–1665.
- Zare, A., Gader, P.D., 2010. PCE: Piecewise convex endmember detection. *IEEE Trans. on Geoscience and Remote Sensing* 48, 2620–2632.
- Zare, A., Ho, K., 2014. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine* 31, 95–104.
- Zhang, B., Zhuang, L., Gao, L., Luo, W., Ran, Q., Du, Q., 2014. PSO-EM: A hyperspectral unmixing algorithm based on normal compositional model. *IEEE Trans. on Geoscience and Remote Sensing* 52, 7782–7792.
- Zhang, L., Weng, Q., 2016. Annual dynamics of impervious surface in the Pearl River Delta, China, from 1988 to 2013, using time series Landsat imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 113, 86–96.
- Zhou, B., Rybski, D., Kropp, J.P., 2017a. The role of city size and urban form in the surface urban heat island. *Scientific reports* 7, 1–9.
- Zhou, Y., Rangarajan, A., Gader, P.D., 2017b. Nonrigid registration of hyperspectral and color images with vastly different spatial and spectral resolutions for spectral unmixing and pansharpening, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1571–1579.
- Zhou, Y., Rangarajan, A., Gader, P.D., 2018. A Gaussian mixture model representation of endmember variability in hyperspectral unmixing. *IEEE Transactions on Image Processing* 27, 2242–2256.
- Zhou, Y., Rangarajan, A., Gader, P.D., 2019. An integrated approach to registration and fusion of hyperspectral and multispectral images. *IEEE Transactions on Geoscience and Remote Sensing* .
- Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. *IEEE Transactions on Geoscience and Remote Sensing* 37, 1212–1226.

# Unmixing urban hyperspectral imagery using probability distributions to represent endmember variability

## Supporting Information

Yuan Zhou<sup>a,1,\*</sup>, Erin B. Wetherley<sup>b</sup>, Paul D. Gader<sup>a</sup>

<sup>a</sup>*Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA*

<sup>b</sup>*Department of Geography, University of California Santa Barbara, Santa Barbara, CA, USA*

---

### Abstract

This supporting information contains additional results to the main paper. Section 1 presents all the 64 polygons with details on the hyperspectral images, ground truth abundances, abundance maps from all the algorithms, and average errors. Section 2 presents additional scatter plots and Bland-Altman plots for the real data.

---

### 1. Abundance Maps

We show all the abundance maps here (Fig. S1 - Fig. S64). The color images are derived by extracting the bands corresponding to 650 nm, 540 nm, 470 nm. The reference abundances are derived from corresponding high-resolution images. The abundance values in the bar plot are in the same order of abundance maps. The last column shows the average MAE over all the classes.

One possible example of regions with slight shift is polygon No. 9 (Fig. S9). In its 4 m image, we see there may be a sufficient amount of trees, but the validated abundances show very little tree presence. Also, all the methods give very bad unmixing in this polygon. Hence, the actual region corresponding to the ground truth abundances may be slightly shifted due to inaccurate UTM coordinates.

In general, the top 4 methods give similar abundance maps. MESMA has sharp abundance maps compared to AAM due to its search order that starts with one endmember. This could be beneficial in the 4 m case but may not help much in 16 m. We see most of the time GMM and GMM-1 are similar. But the estimated abundances for roof is slightly higher in GMM (e.g. Fig. S2), which could result from using a multi-modal distribution for roof in GMM. NCM Sampling sometimes gives very strange abundance maps in the 16 m case, but does not in 4 m. We conjecture that it is the actual spectra in 16 m and the library that confused NCM Sampling.

---

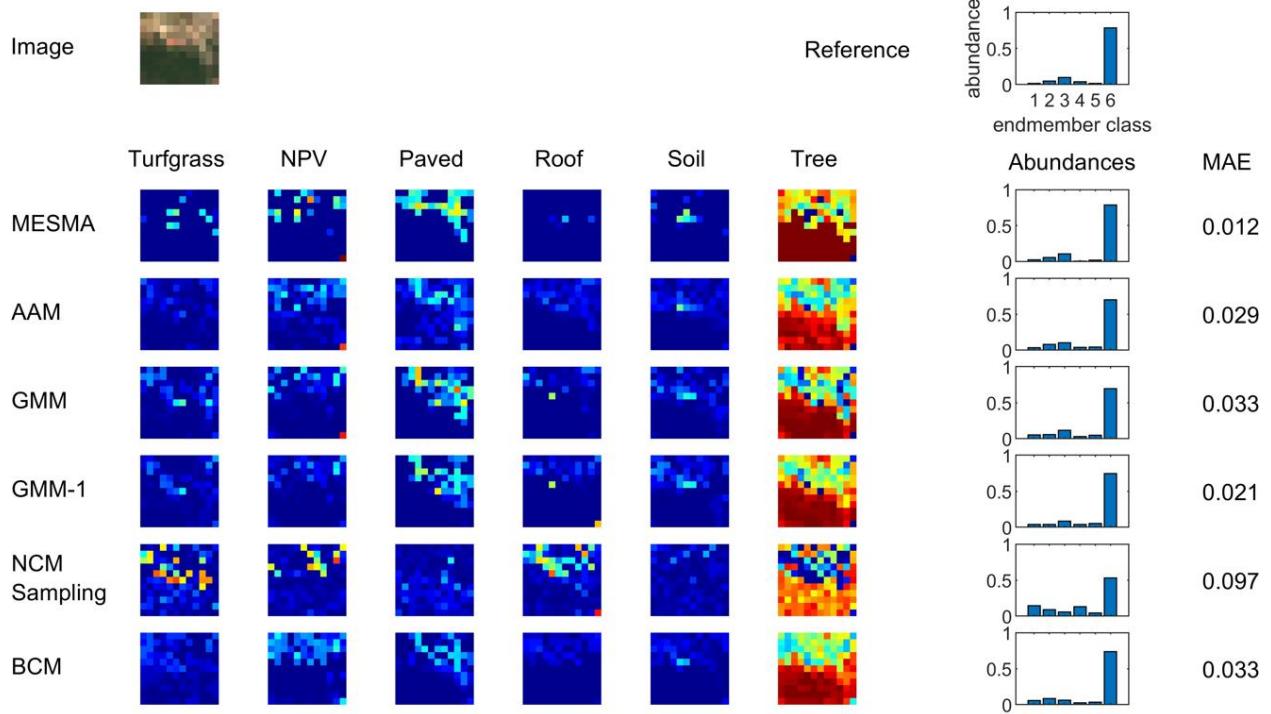
\*Corresponding author

Email addresses: zhouyuanzxcv@gmail.com (Yuan Zhou), wetherley@umail.ucsb.edu (Erin B. Wetherley), pgader@cise.ufl.edu (Paul D. Gader)

<sup>1</sup>Present address: Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA

Comparing MESMA and GMM, they are equally capable in 16 m, while MESMA has better abundance maps in 4 m due to the well reduced library. Sometimes GMM gives better abundance maps in 4 m. For example, in Fig. S18, S19 and S46, MESMA underestimated grass while GMM gives a more faithful estimate of grass. We conjecture that it is the reduced library that leads to this underestimation from MESMA. Since GMM utilizes the full library, it is not influenced by this reduction step. Note that turfgrass is under-represented in this dataset (average abundance 0.055 compared to 0.175, 0.246, 0.138, 0.068, 0.313 from NPV, paved, roof, soil, tree). We may expect that GMM achieves better results if turfgrass has more presence in the dataset.

16 m



4 m

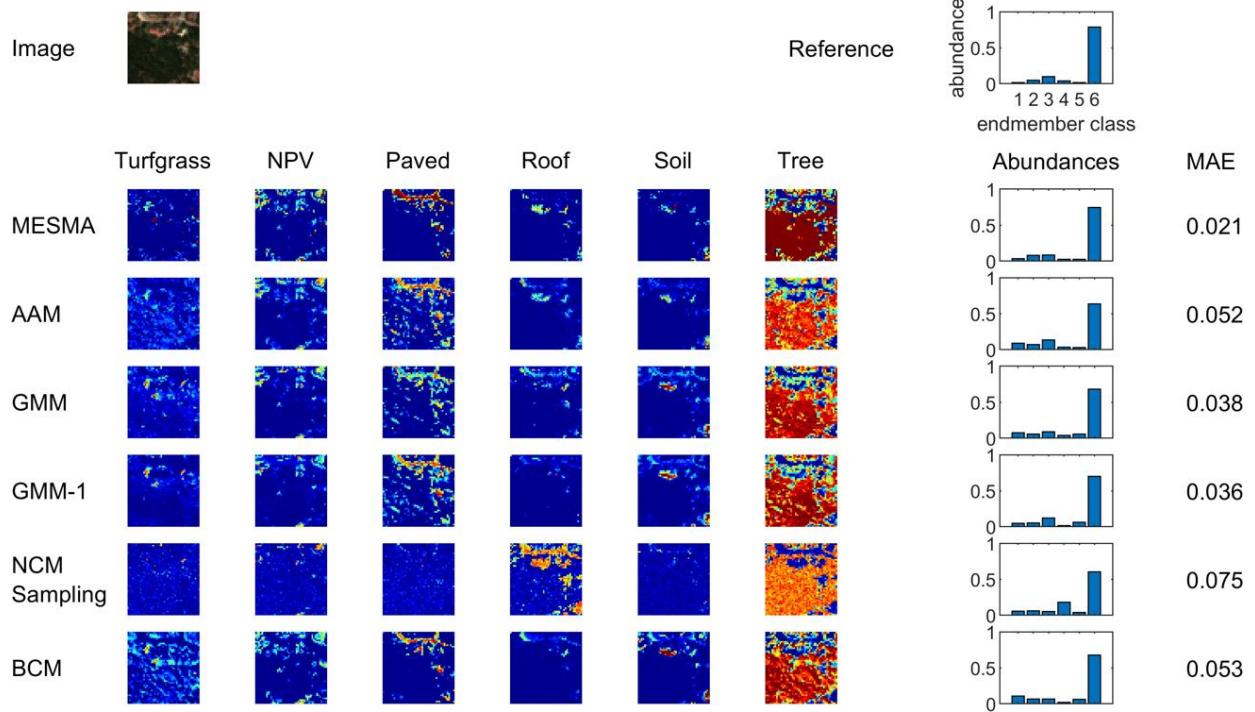
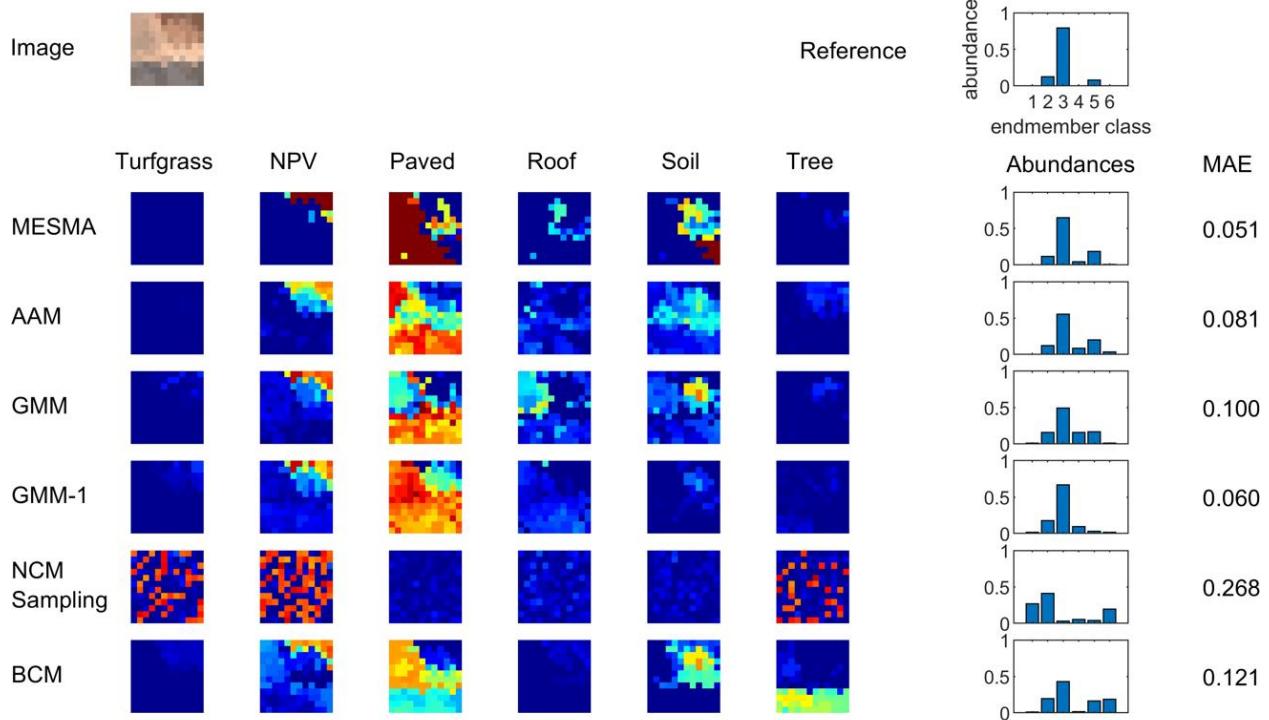


Figure S1: Abundance maps for polygon No. 1.

16 m



4 m

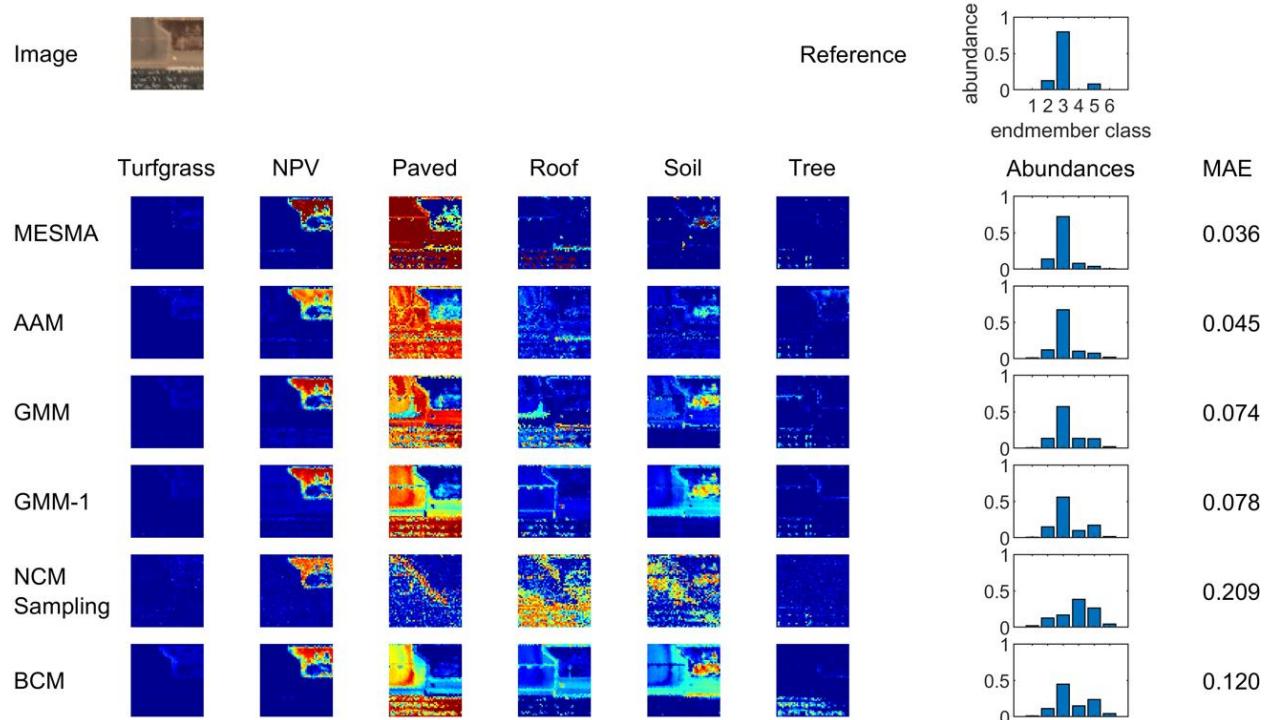
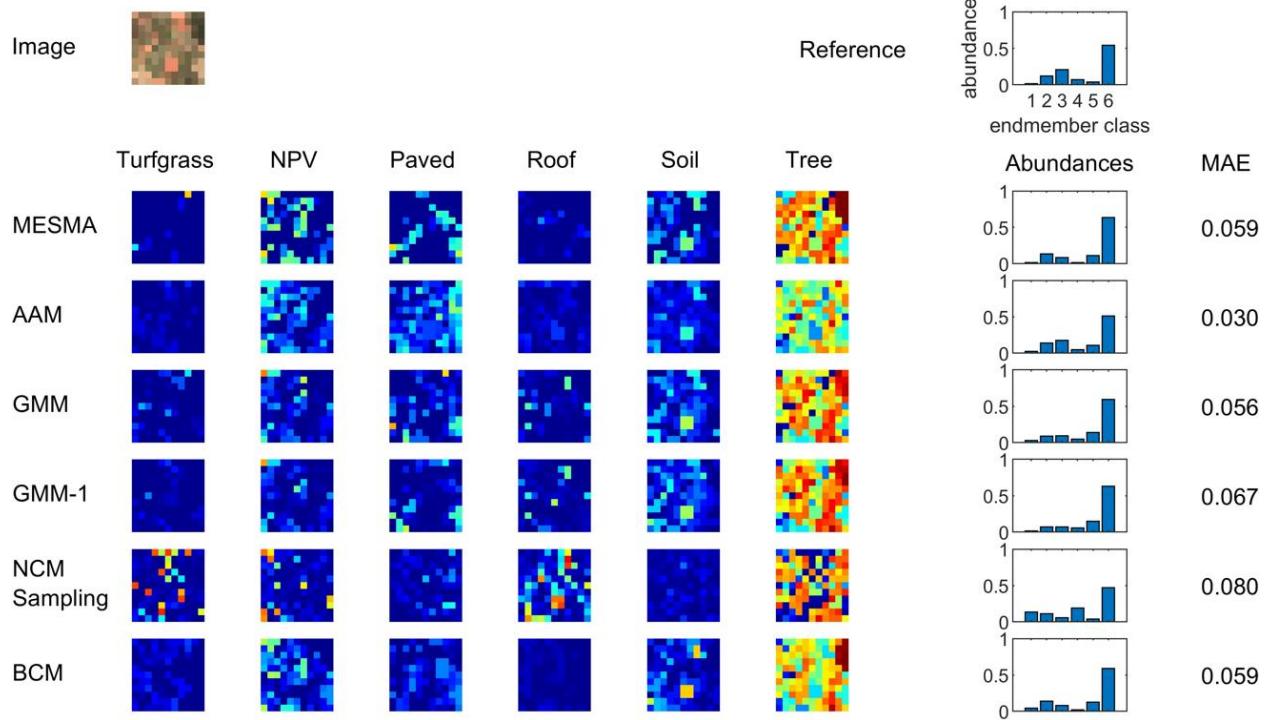


Figure S2: Abundance maps for polygon No. 2.

16 m



4 m

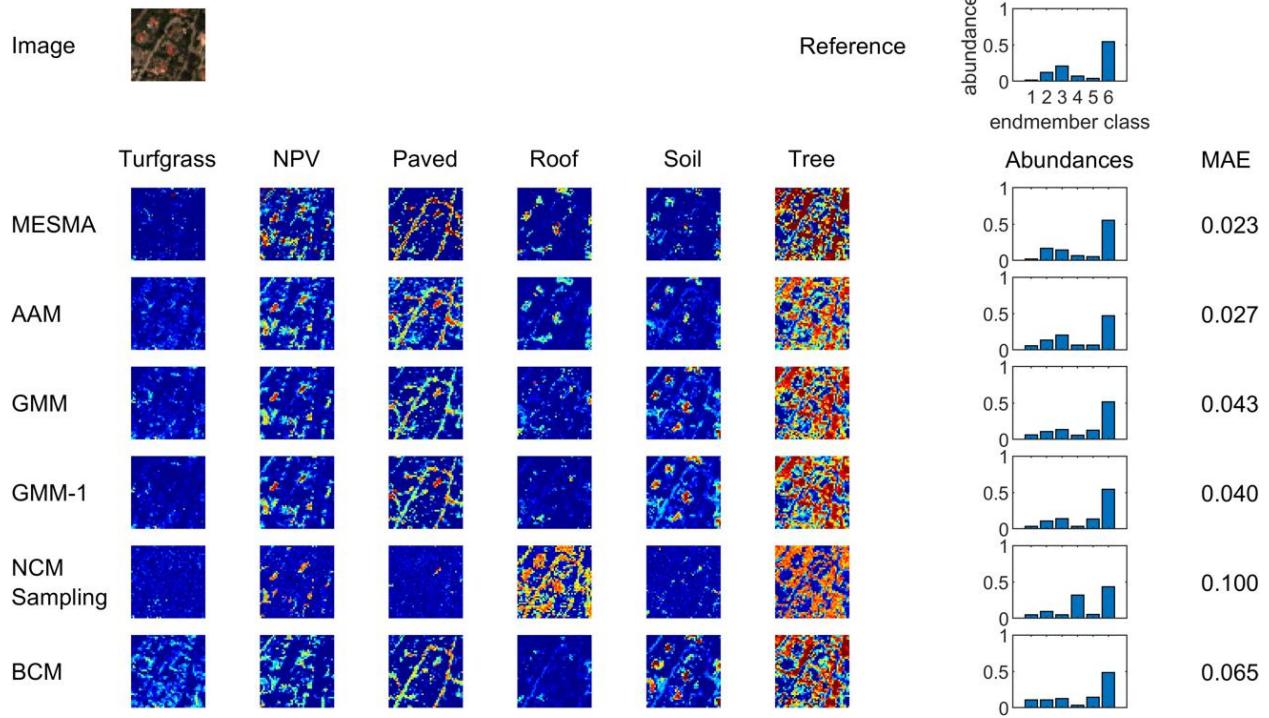
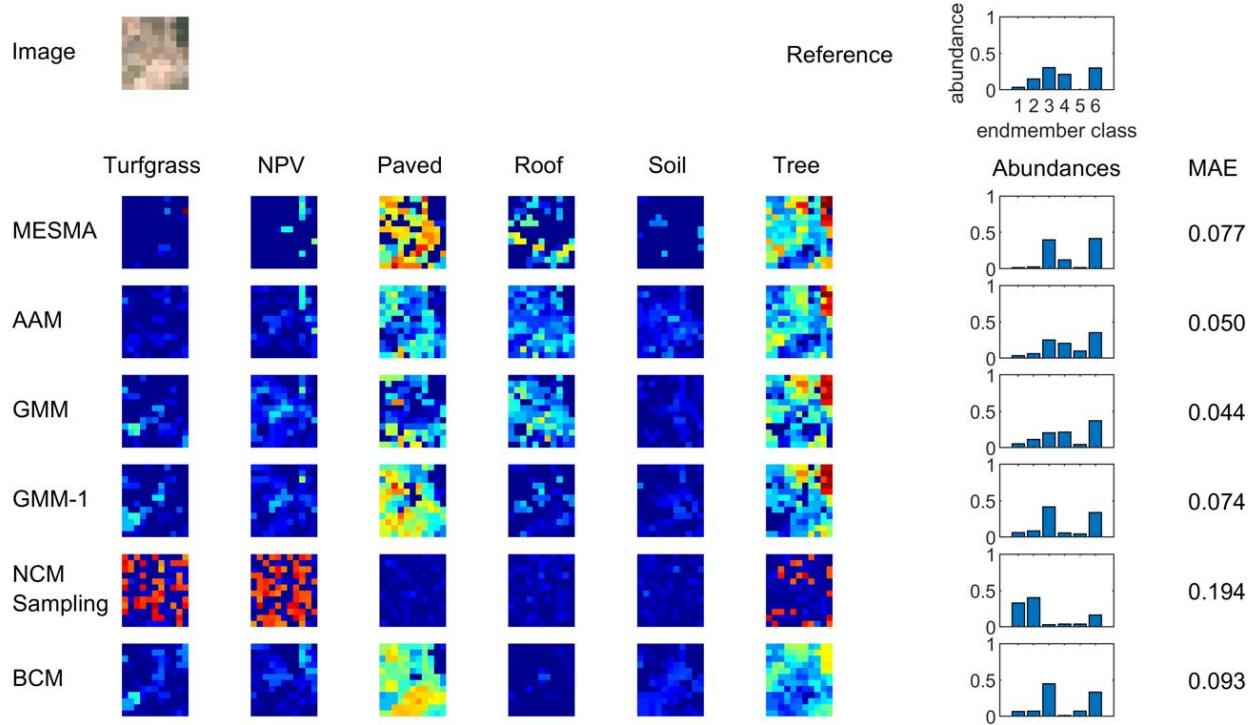


Figure S3: Abundance maps for polygon No. 3.

16 m



4 m

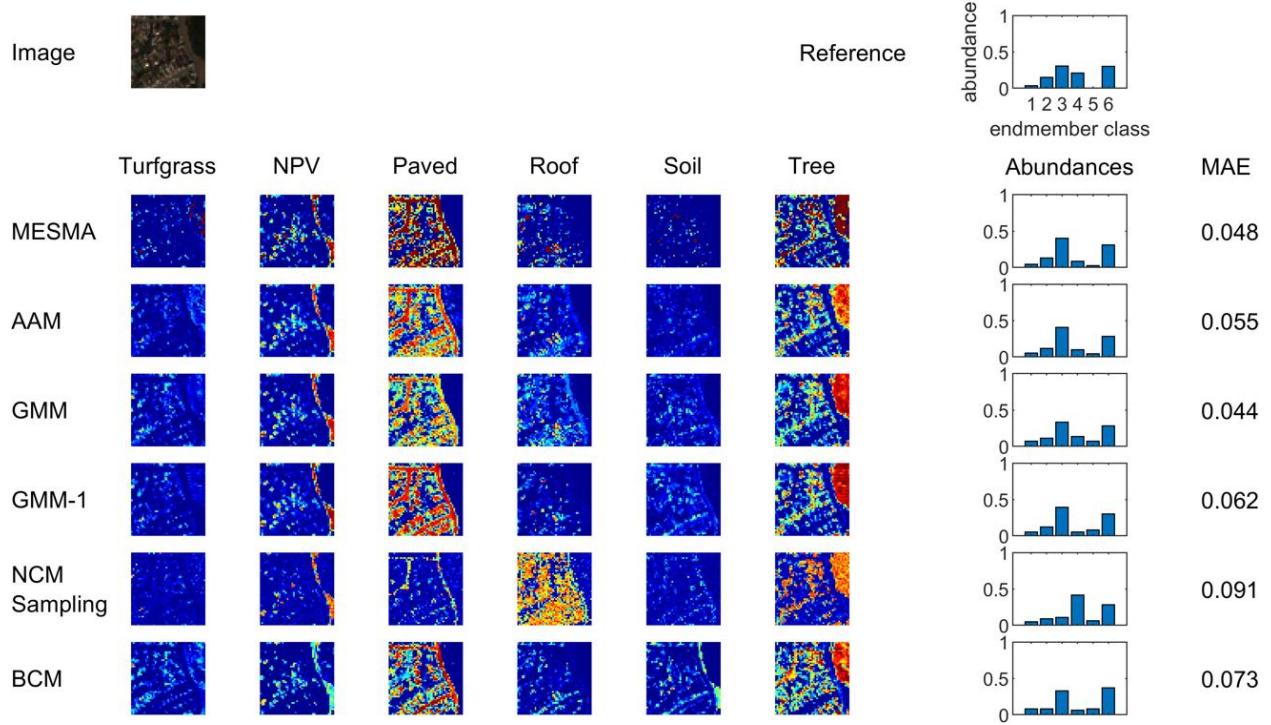
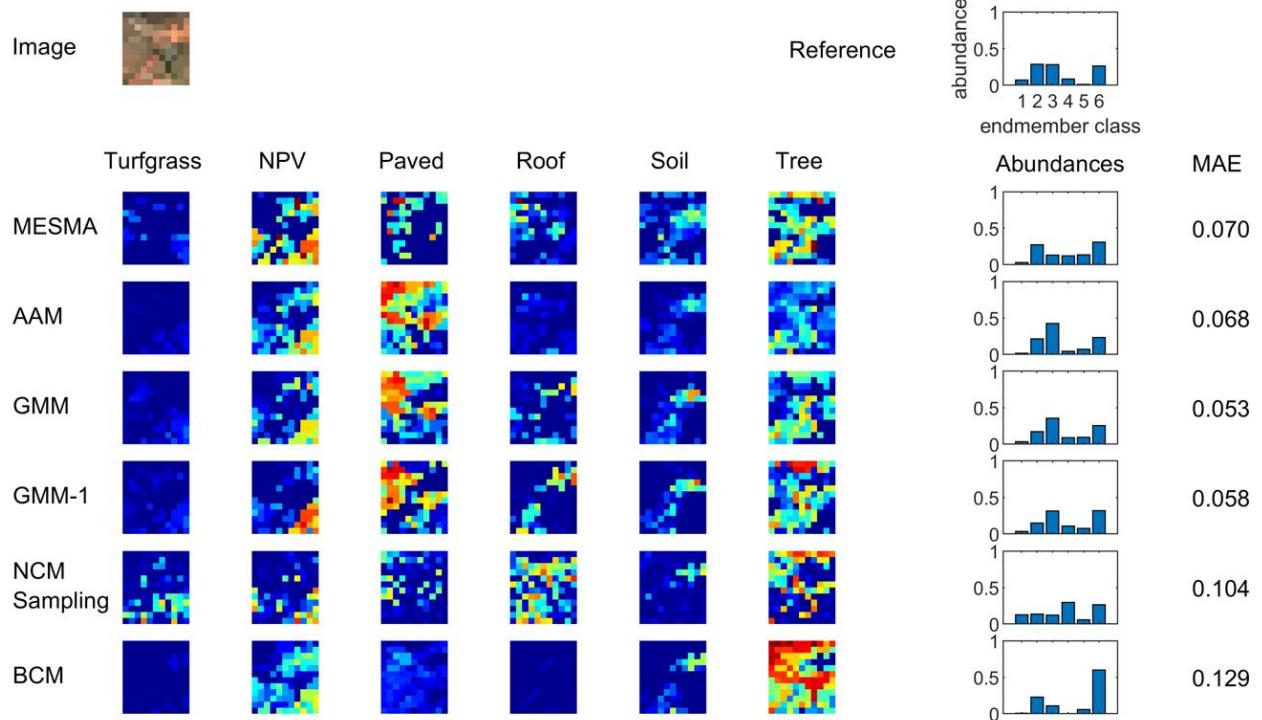


Figure S4: Abundance maps for polygon No. 4.

16 m



4 m

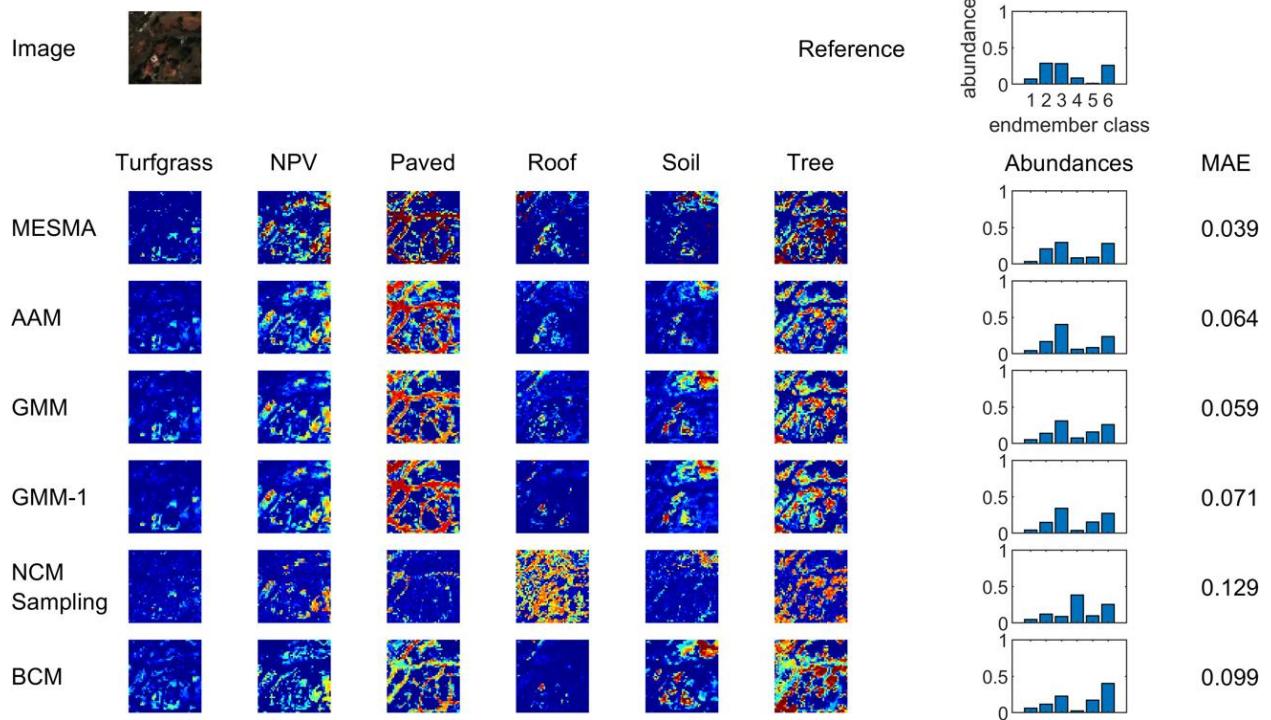
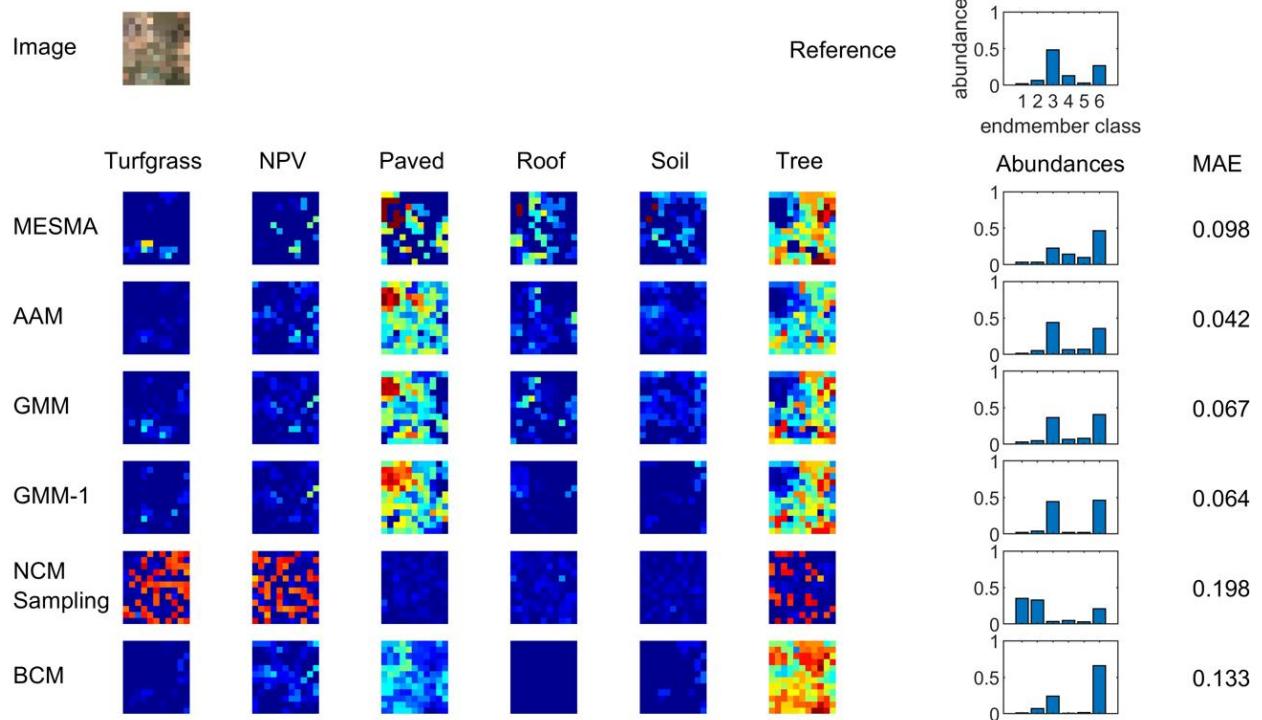


Figure S5: Abundance maps for polygon No. 5.

16 m



4 m

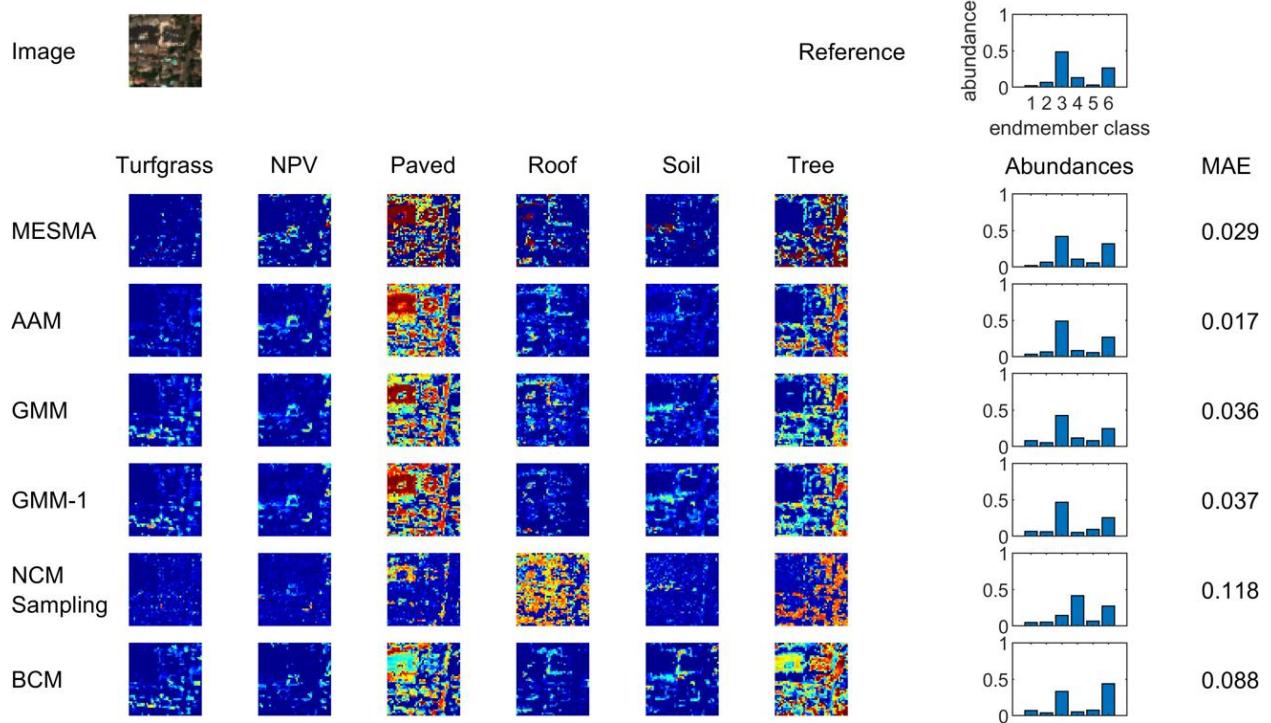
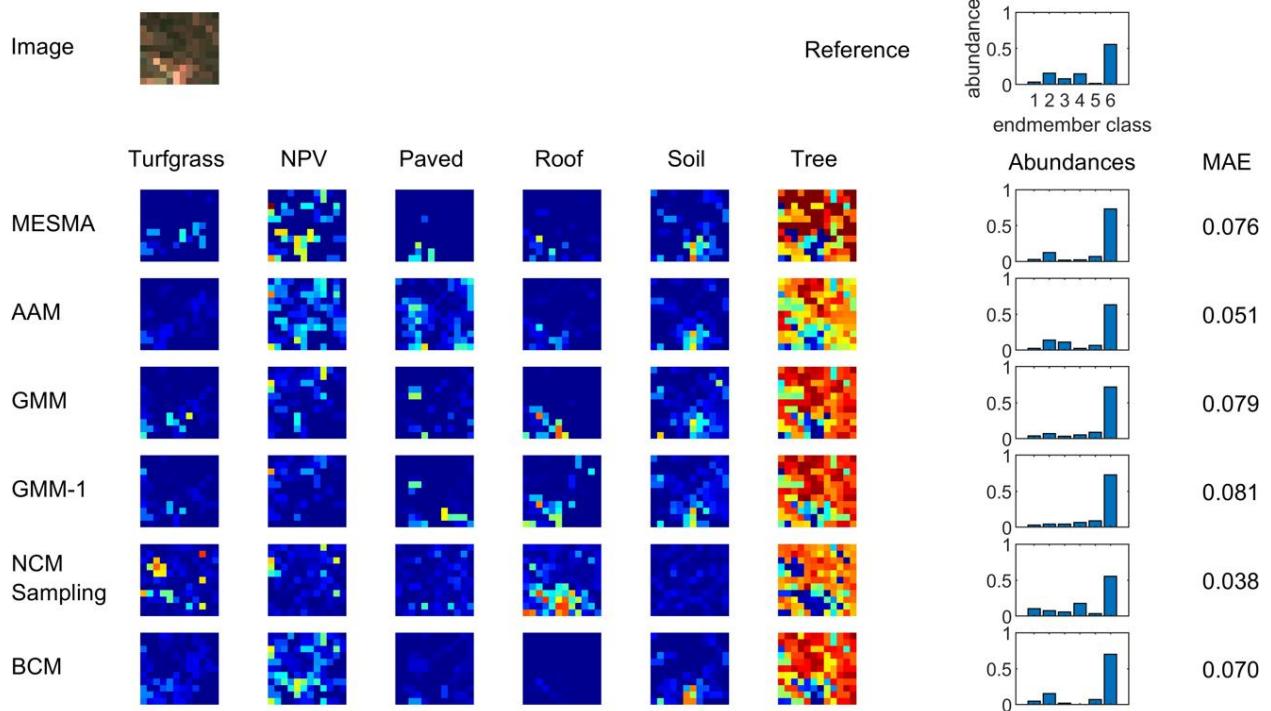


Figure S6: Abundance maps for polygon No. 6.

16 m



4 m

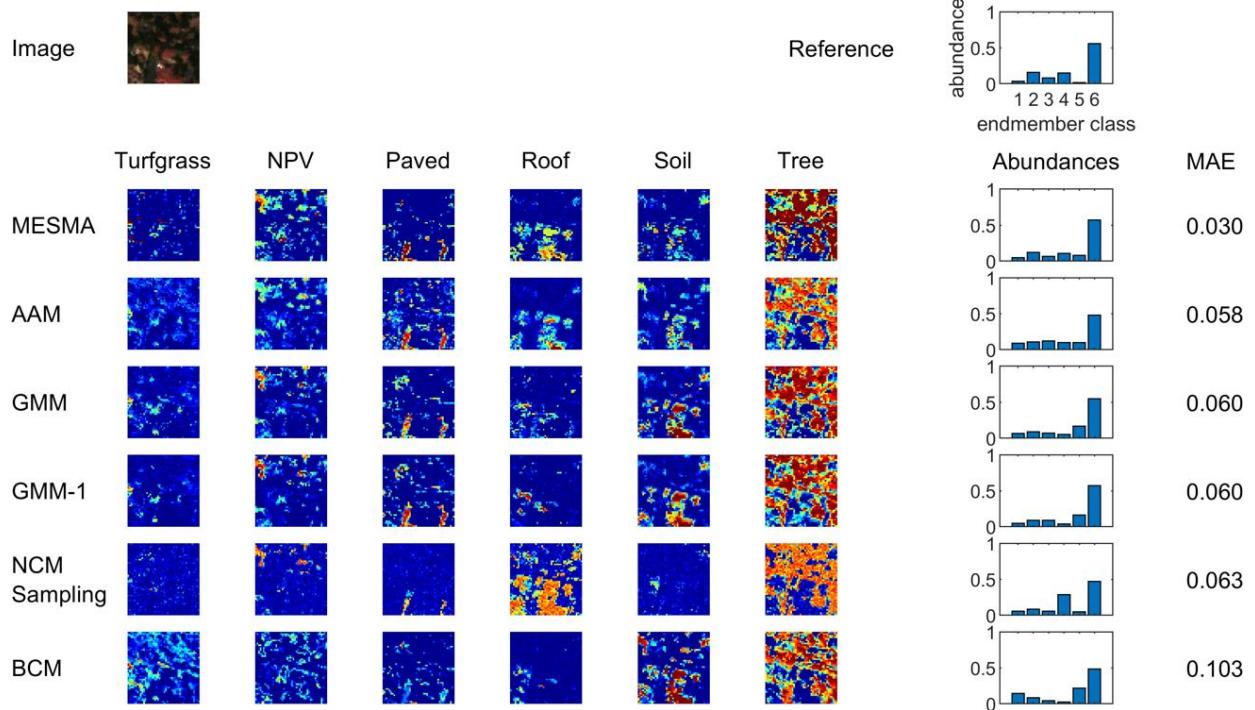
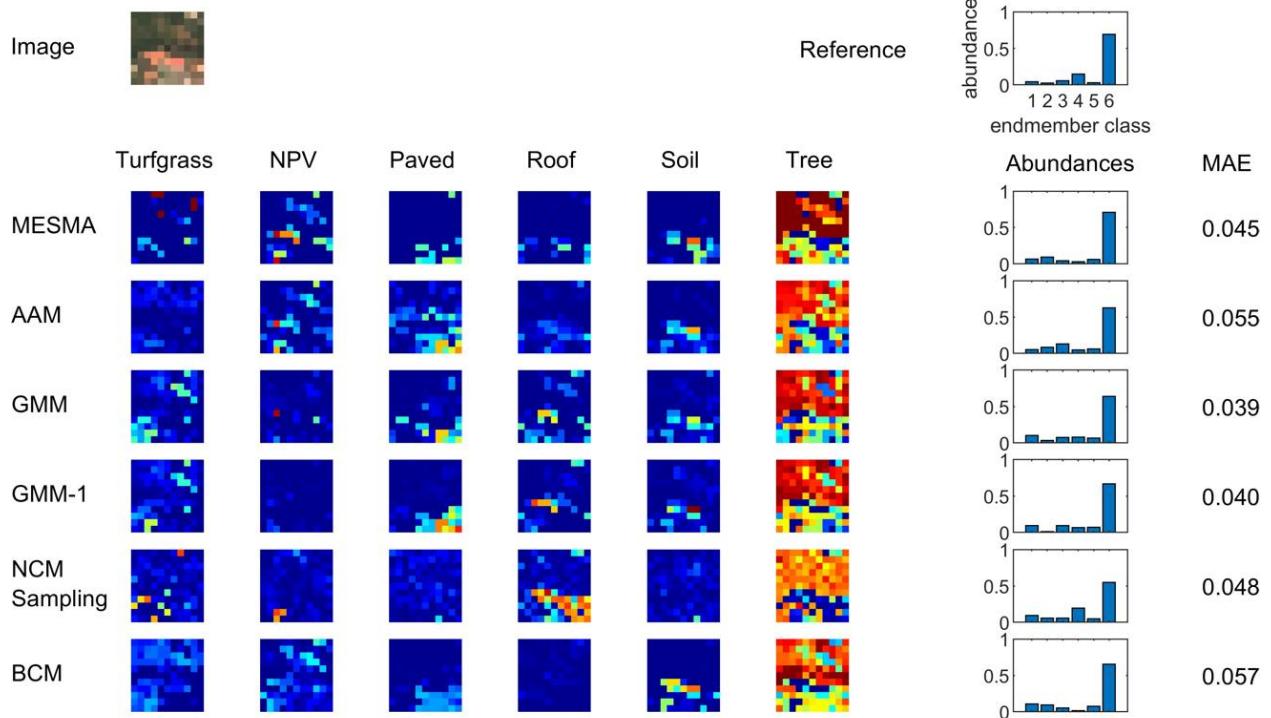


Figure S7: Abundance maps for polygon No. 7.

16 m



4 m

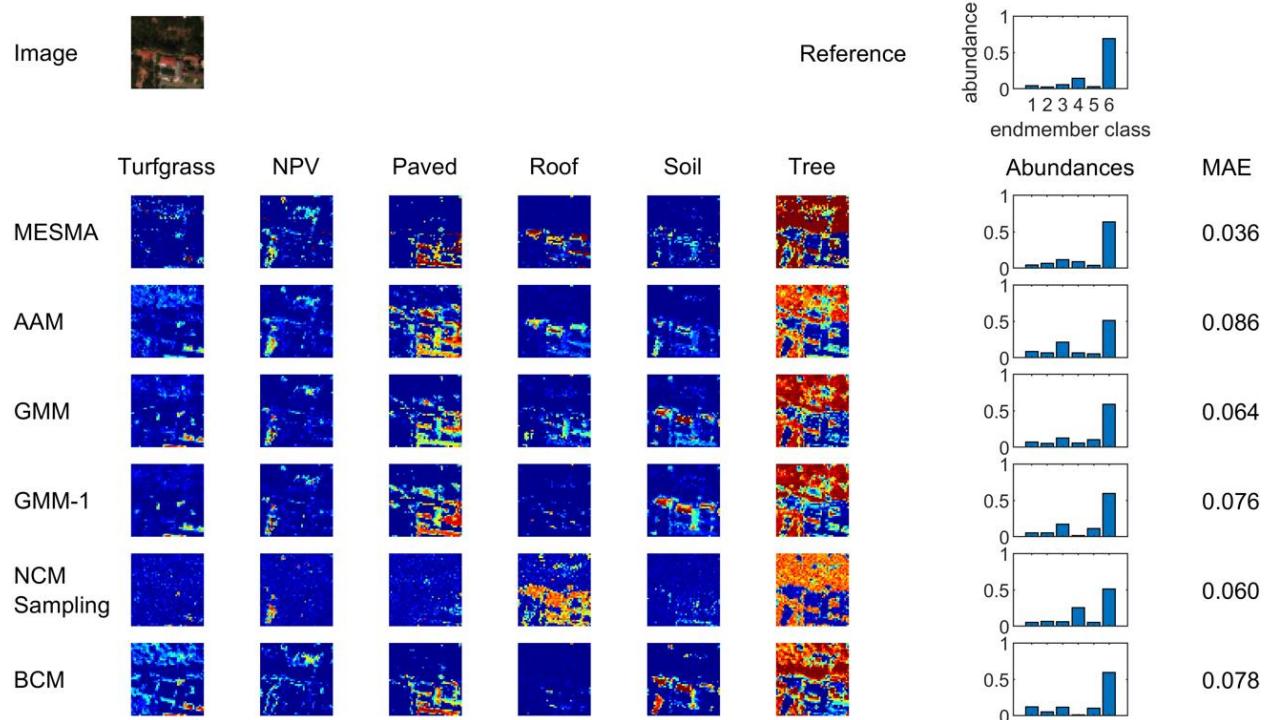
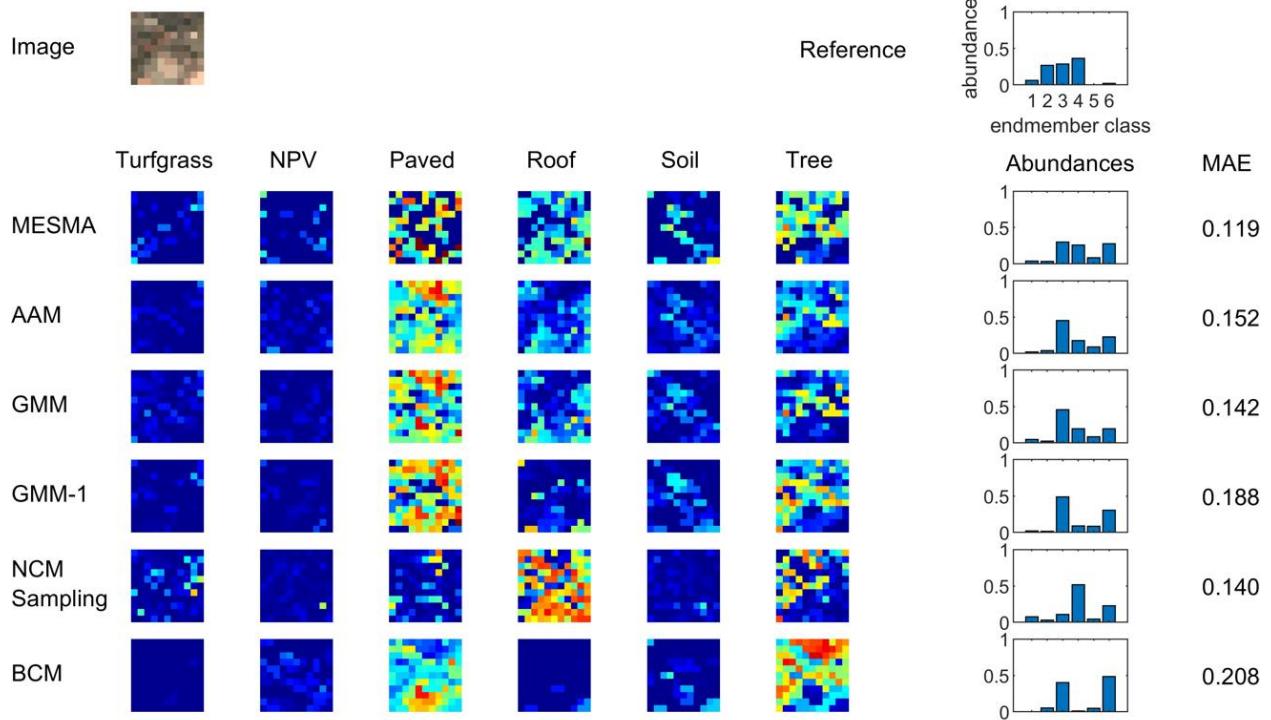


Figure S8: Abundance maps for polygon No. 8.

16 m



4 m

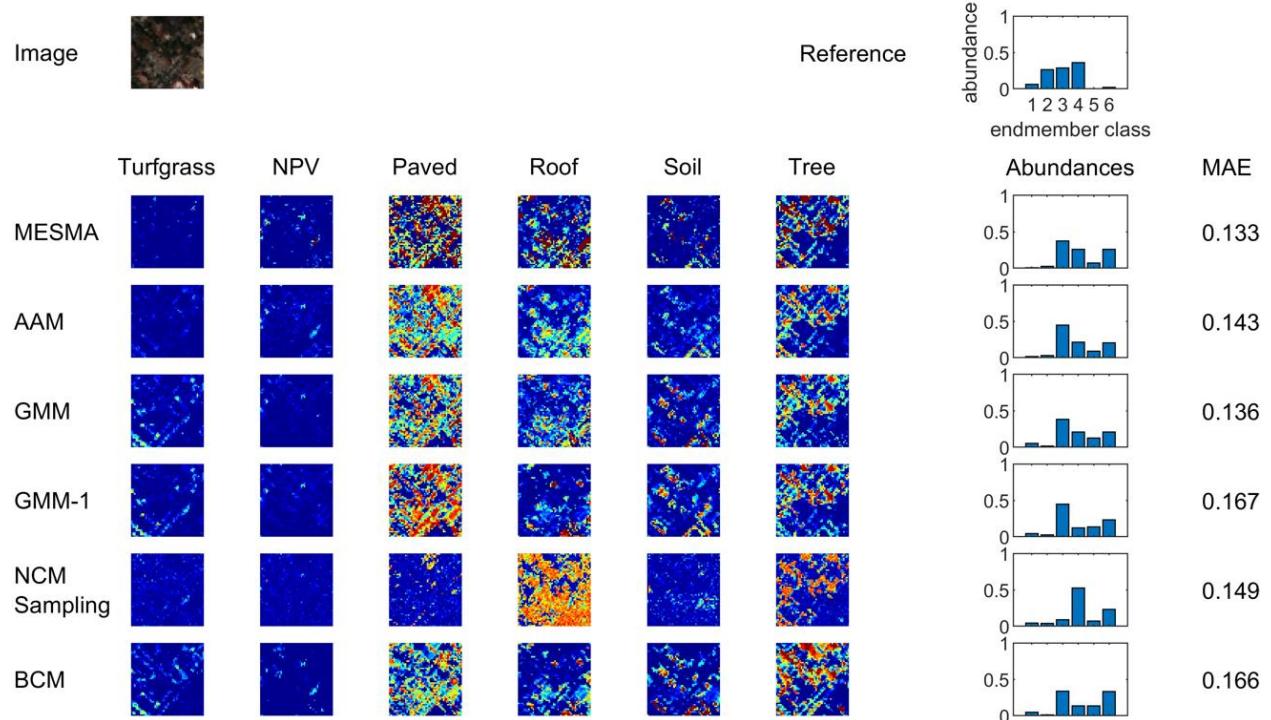
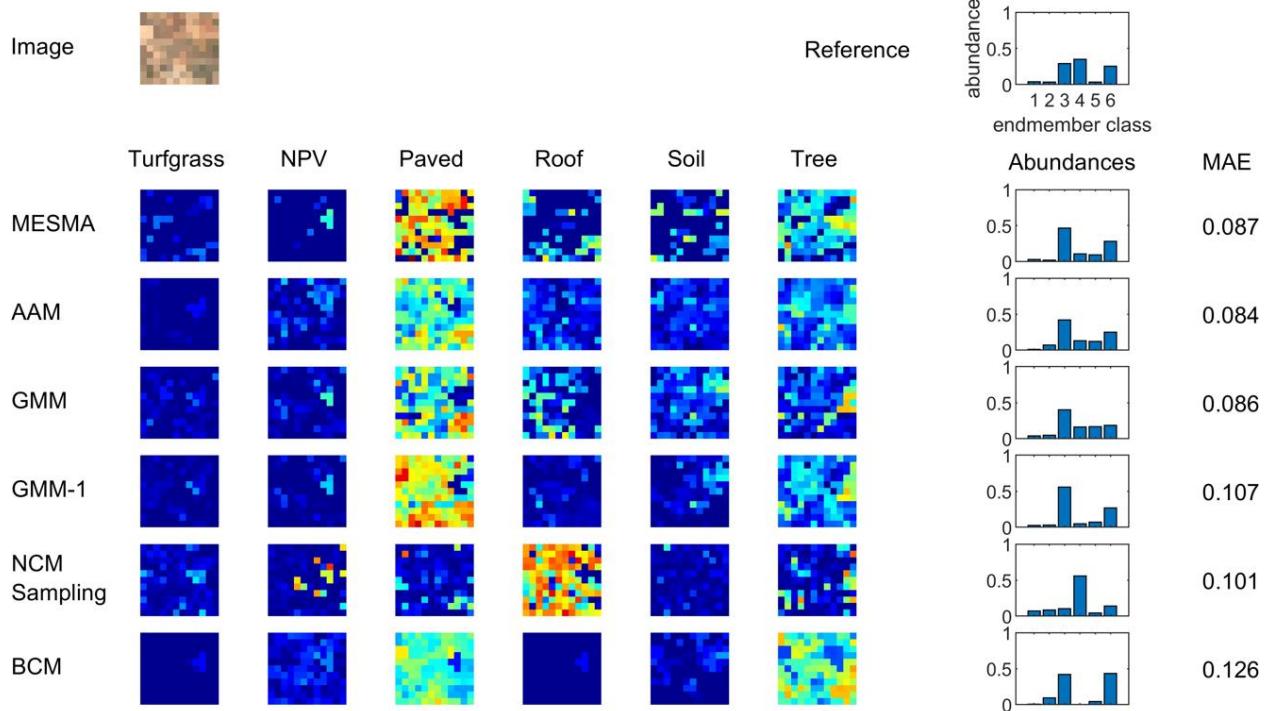


Figure S9: Abundance maps for polygon No. 9.

16 m



4 m

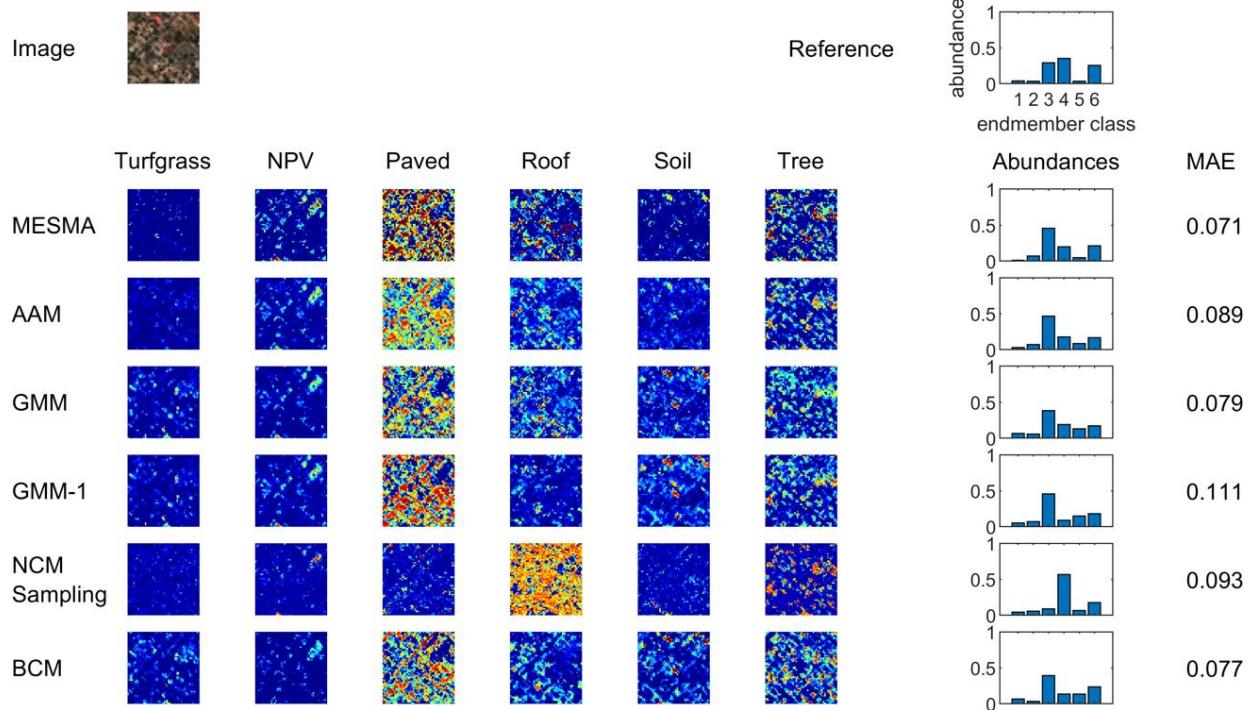
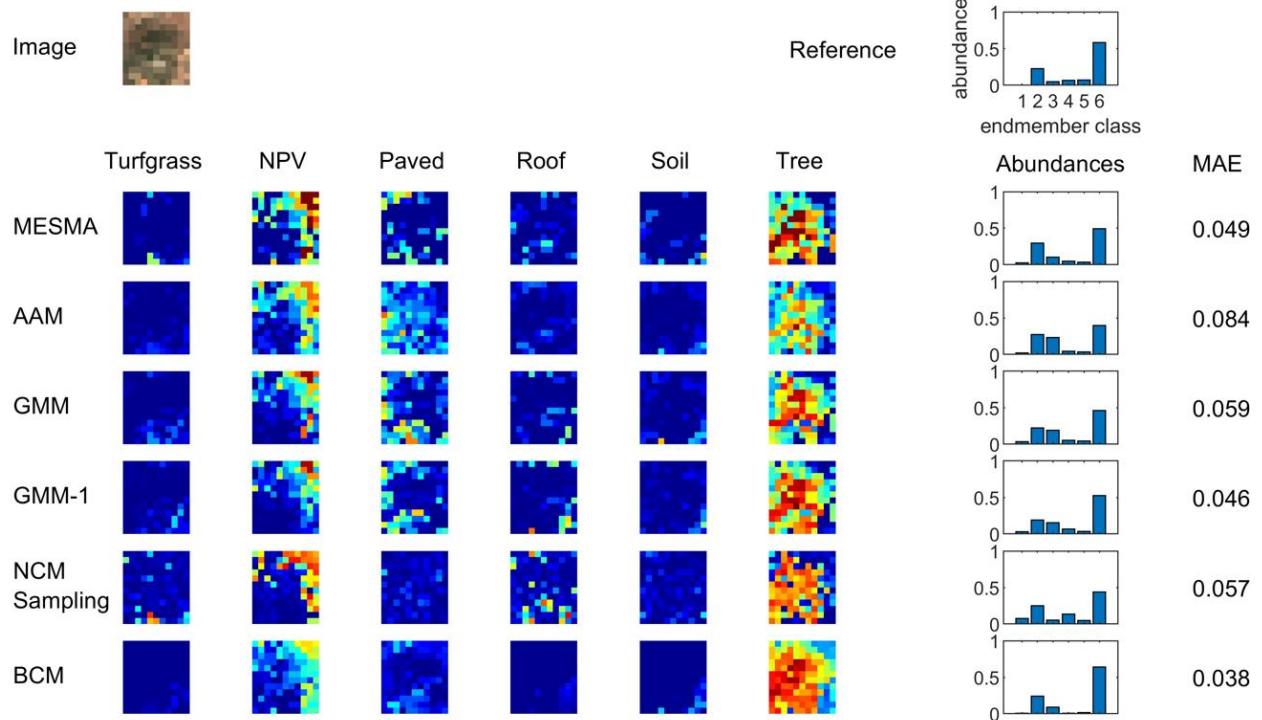


Figure S10: Abundance maps for polygon No. 10.

16 m



4 m

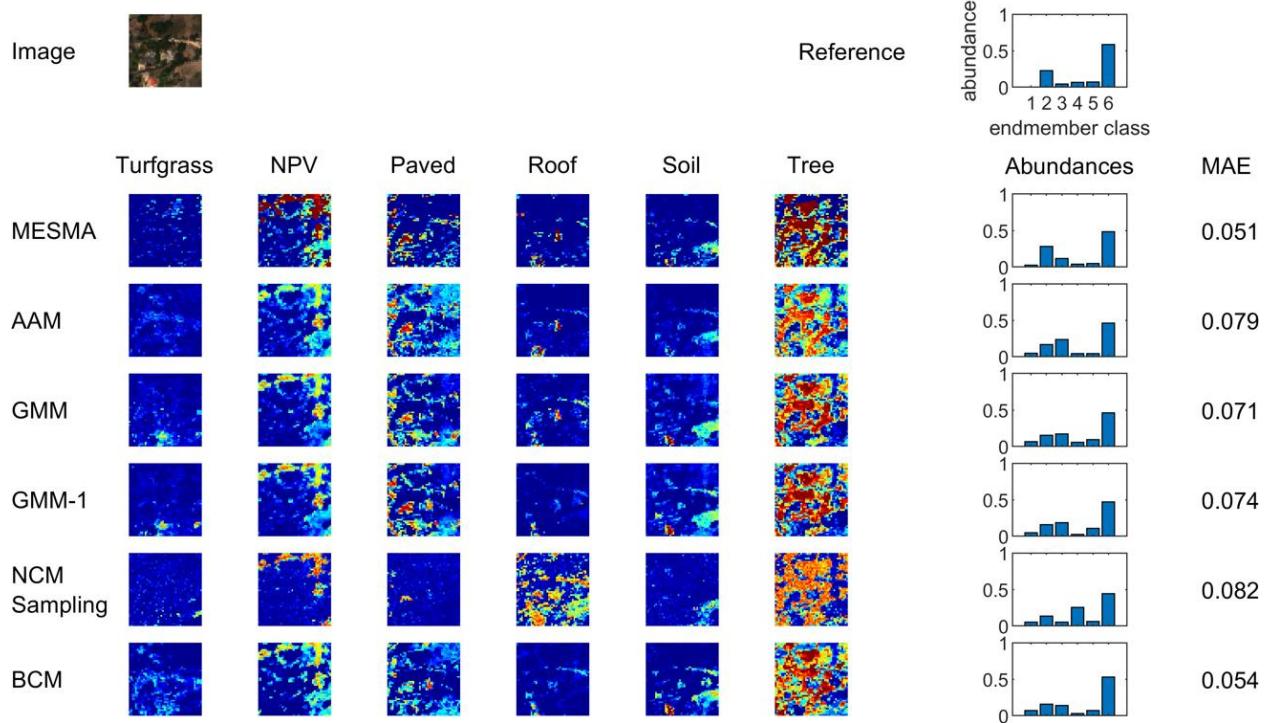
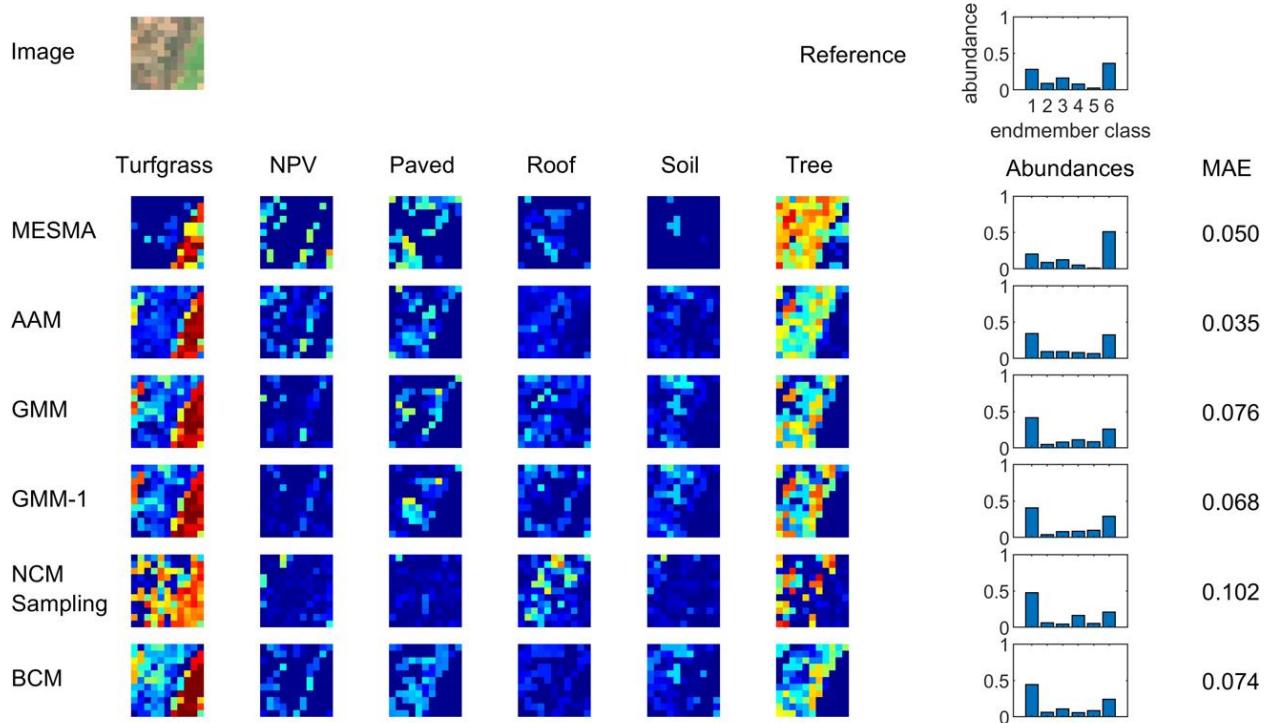


Figure S11: Abundance maps for polygon No. 11.

16 m



4 m

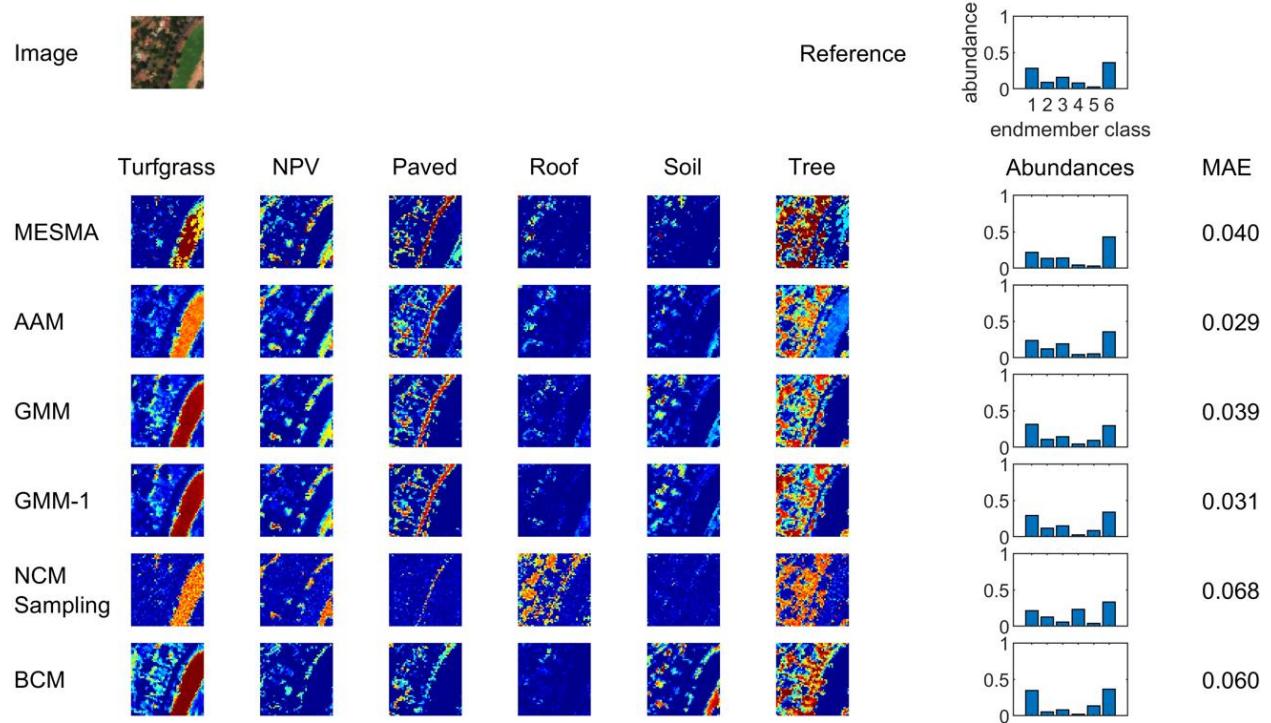
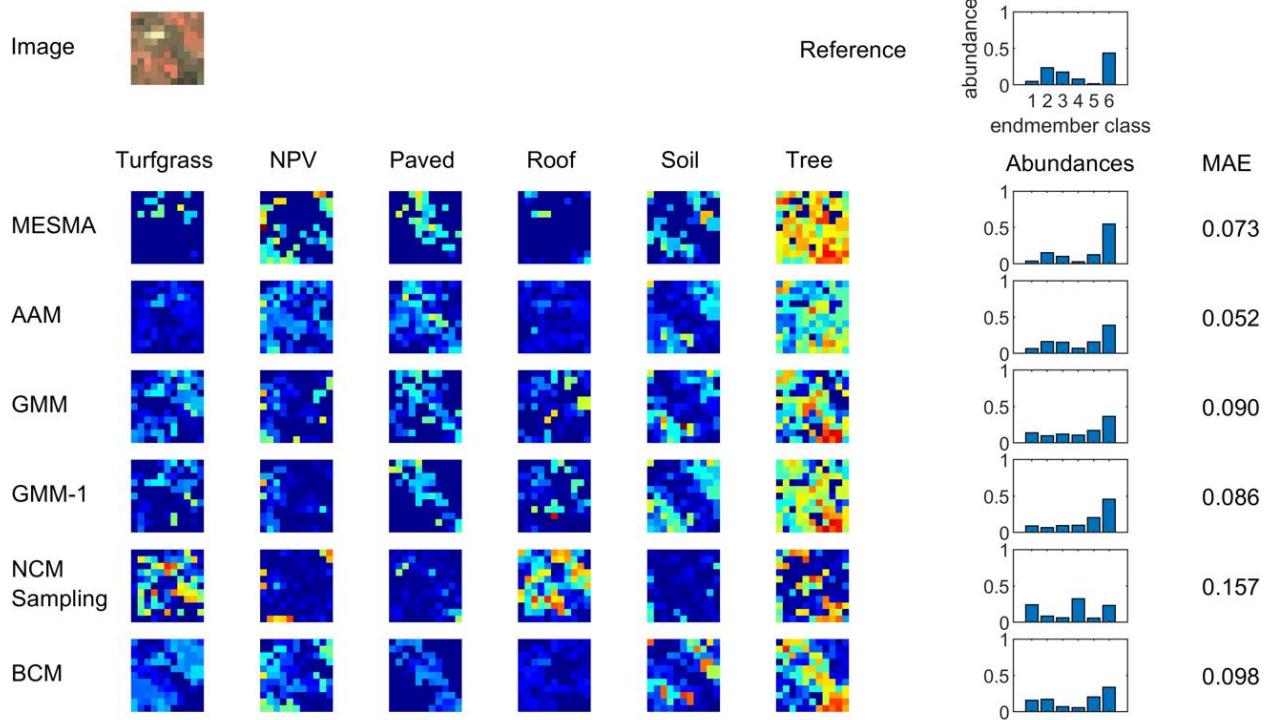


Figure S12: Abundance maps for polygon No. 12.

16 m



4 m

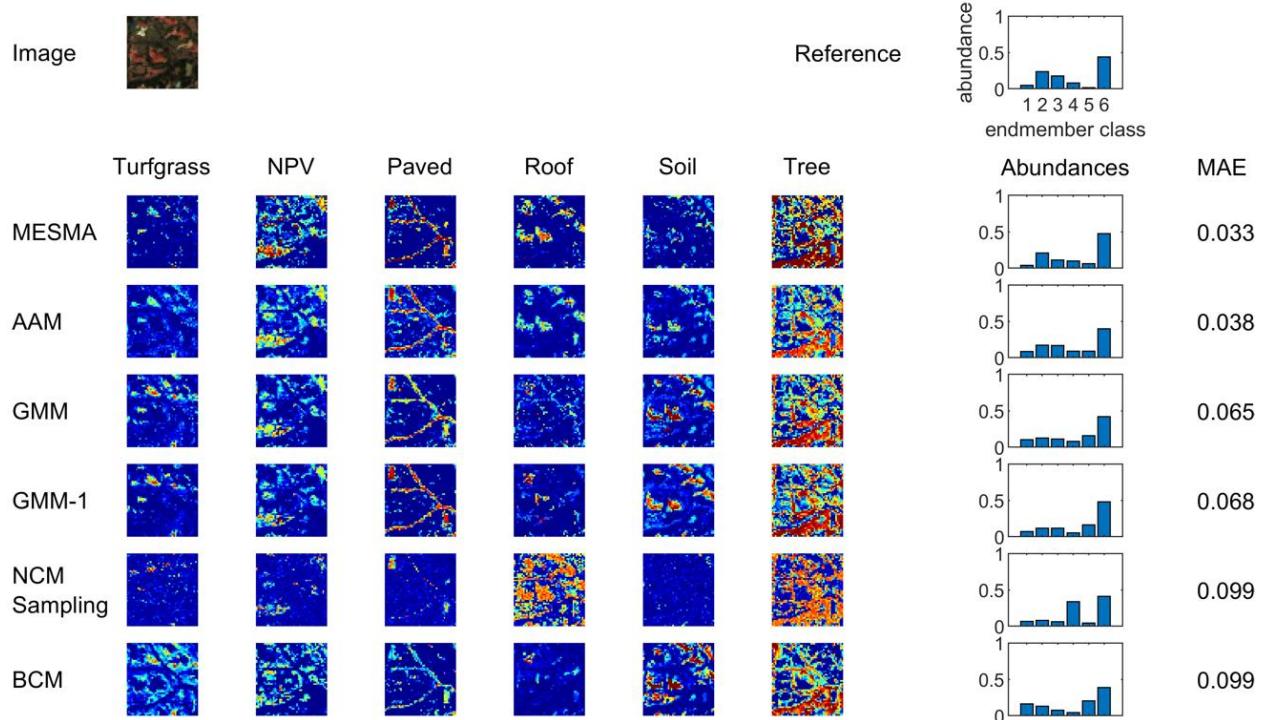
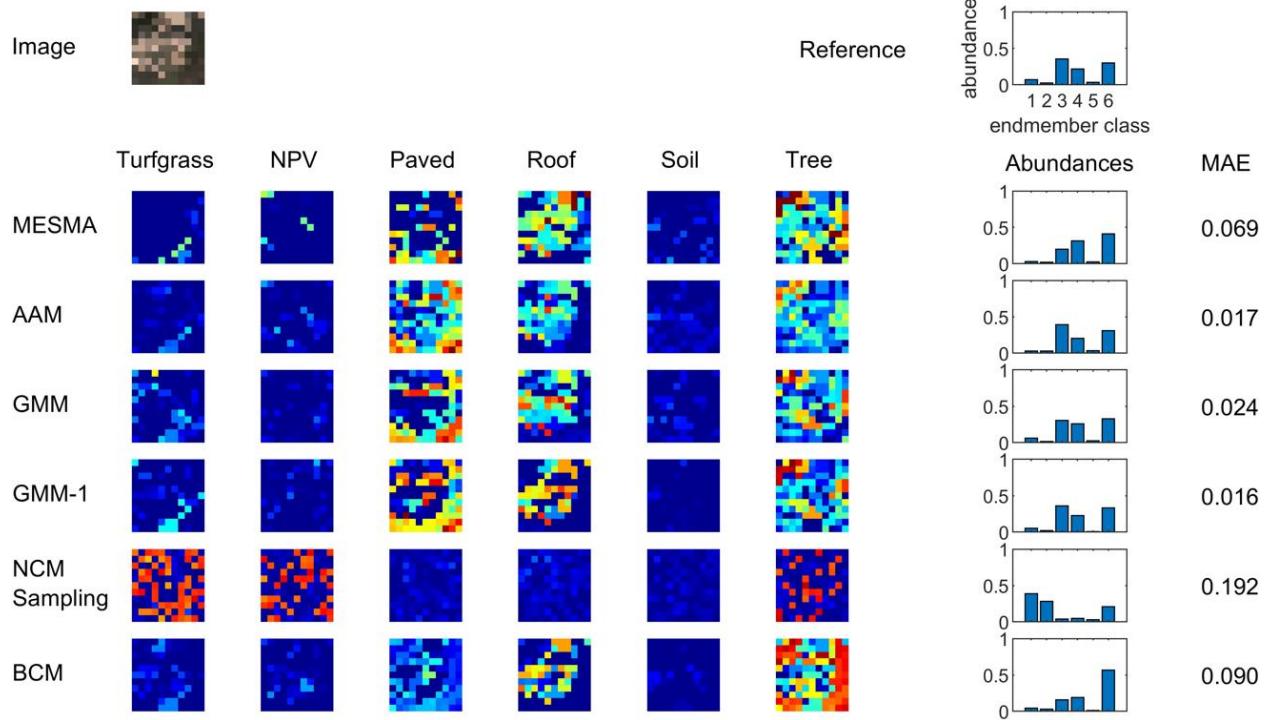


Figure S13: Abundance maps for polygon No. 13.

16 m



4 m

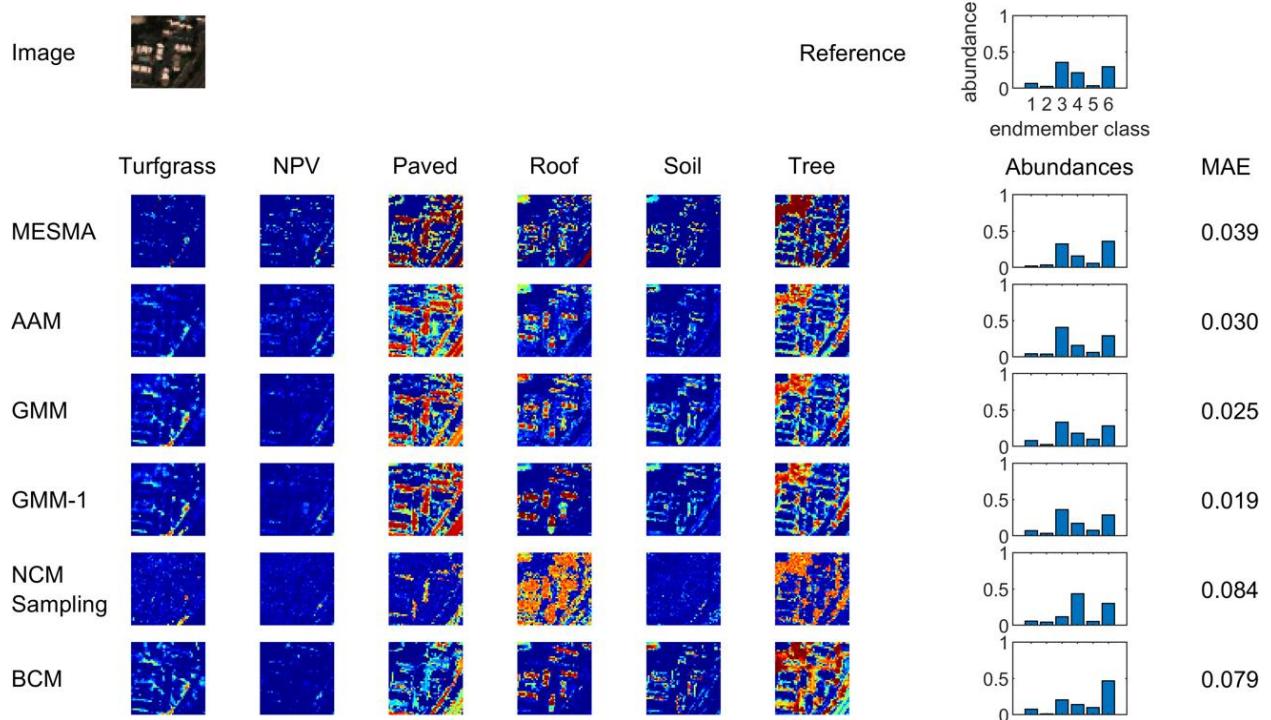
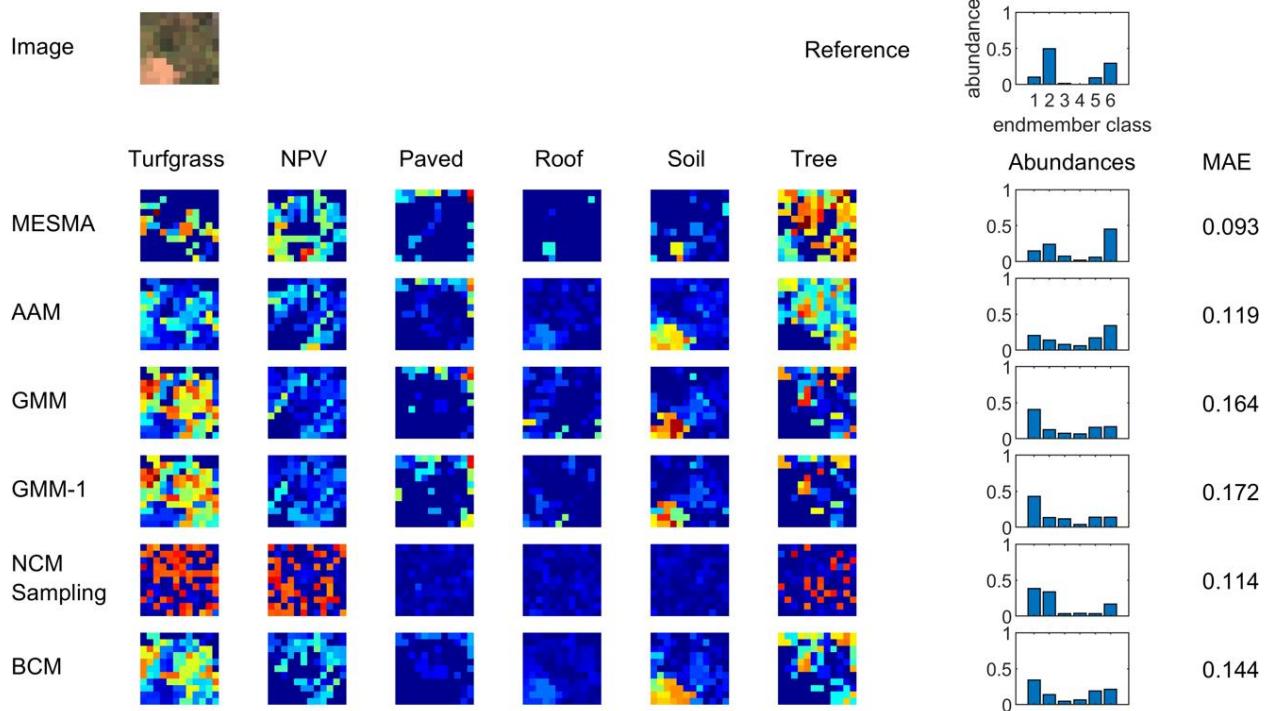


Figure S14: Abundance maps for polygon No. 14.

16 m



4 m

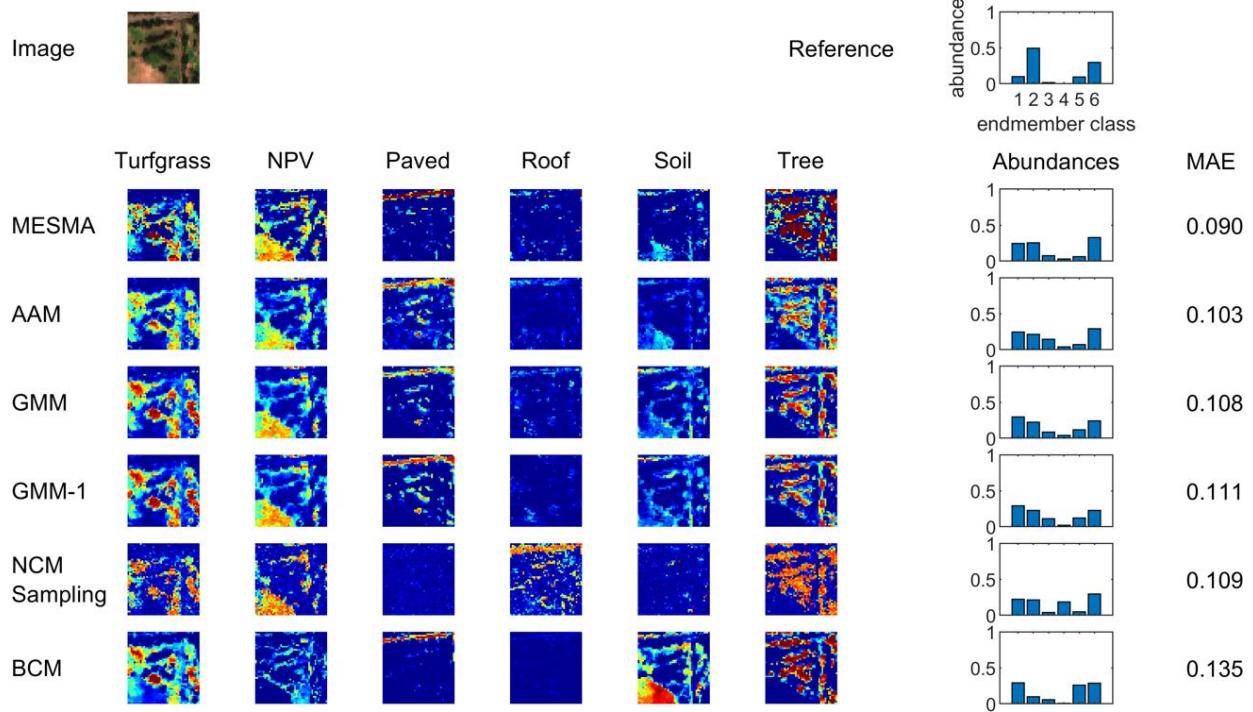
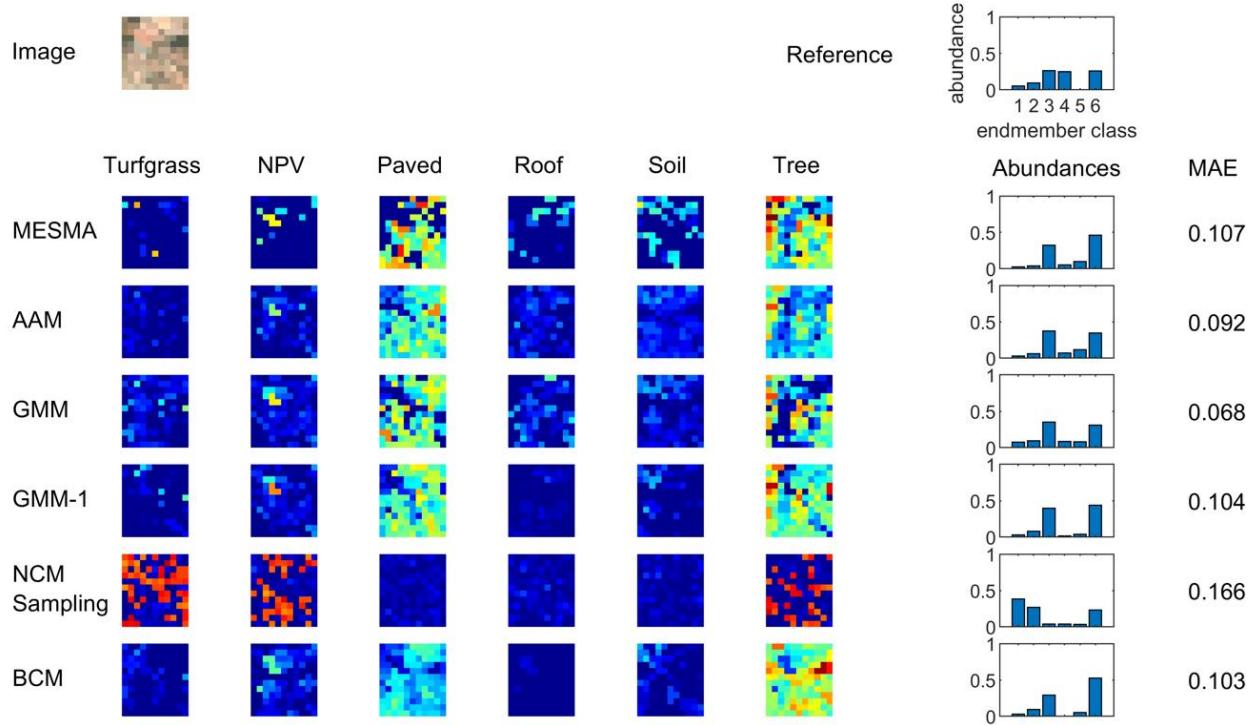


Figure S15: Abundance maps for polygon No. 15.

16 m



4 m

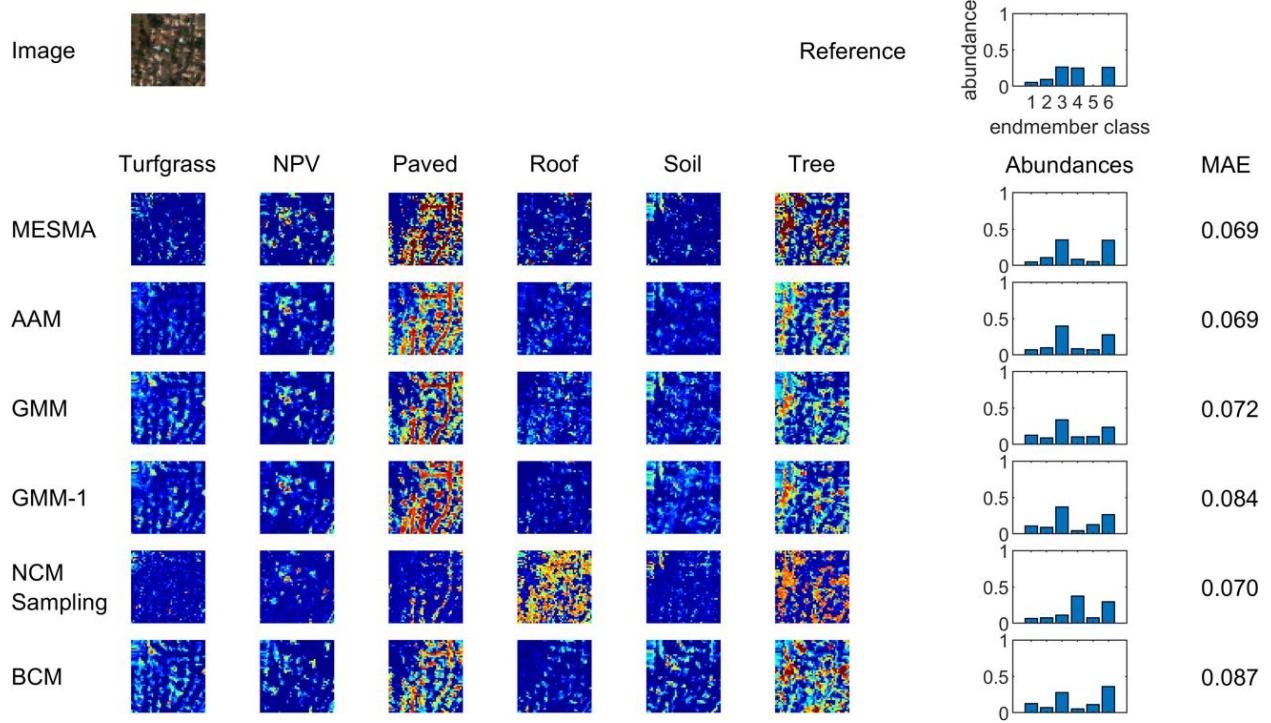
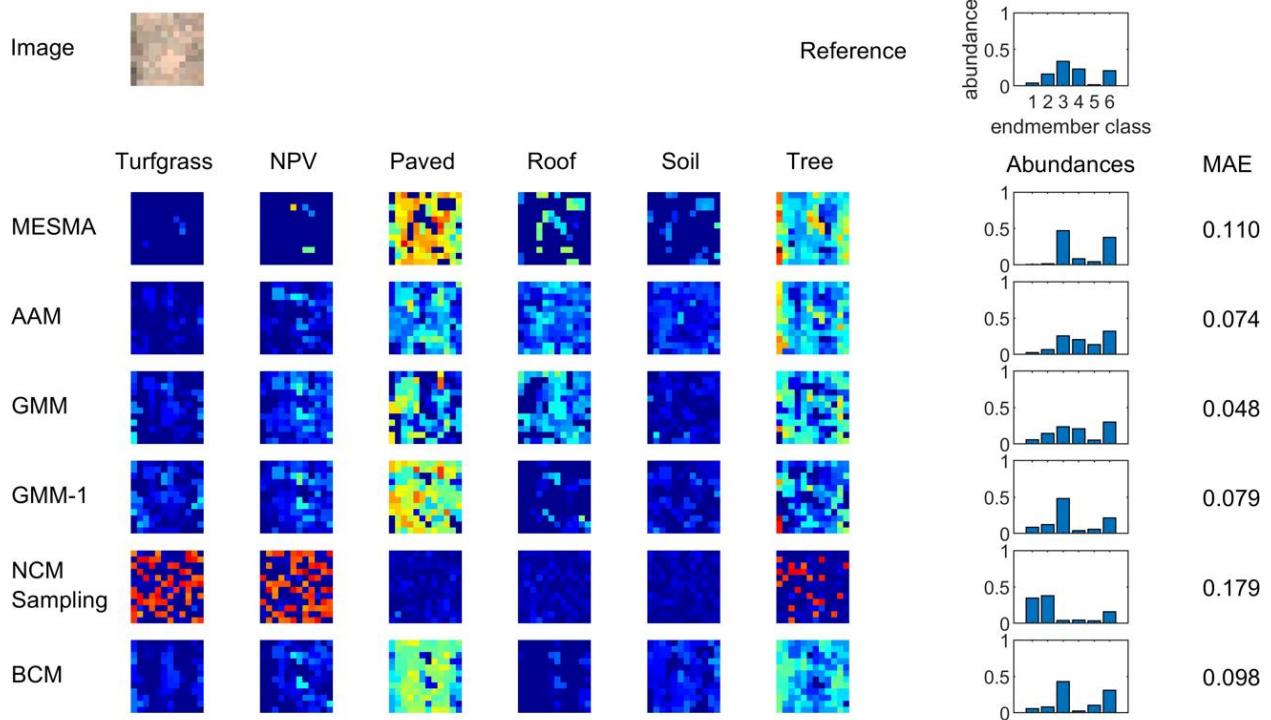


Figure S16: Abundance maps for polygon No. 16.

16 m



4 m

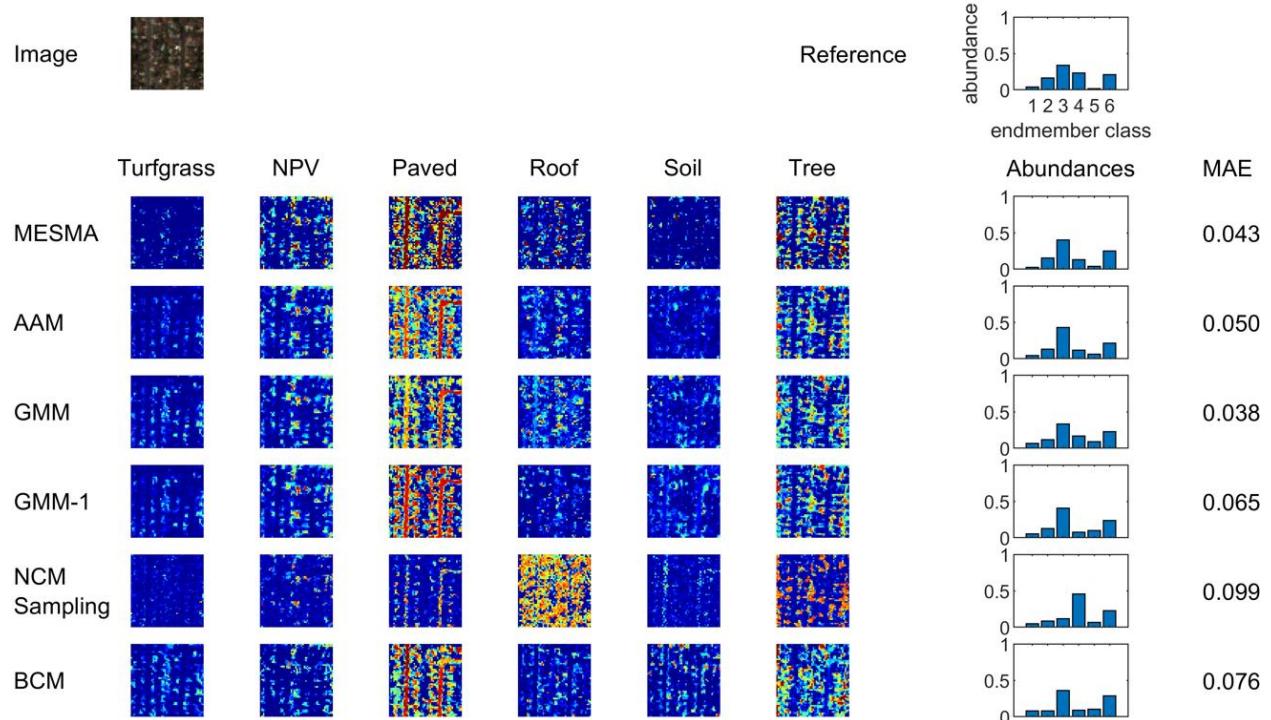
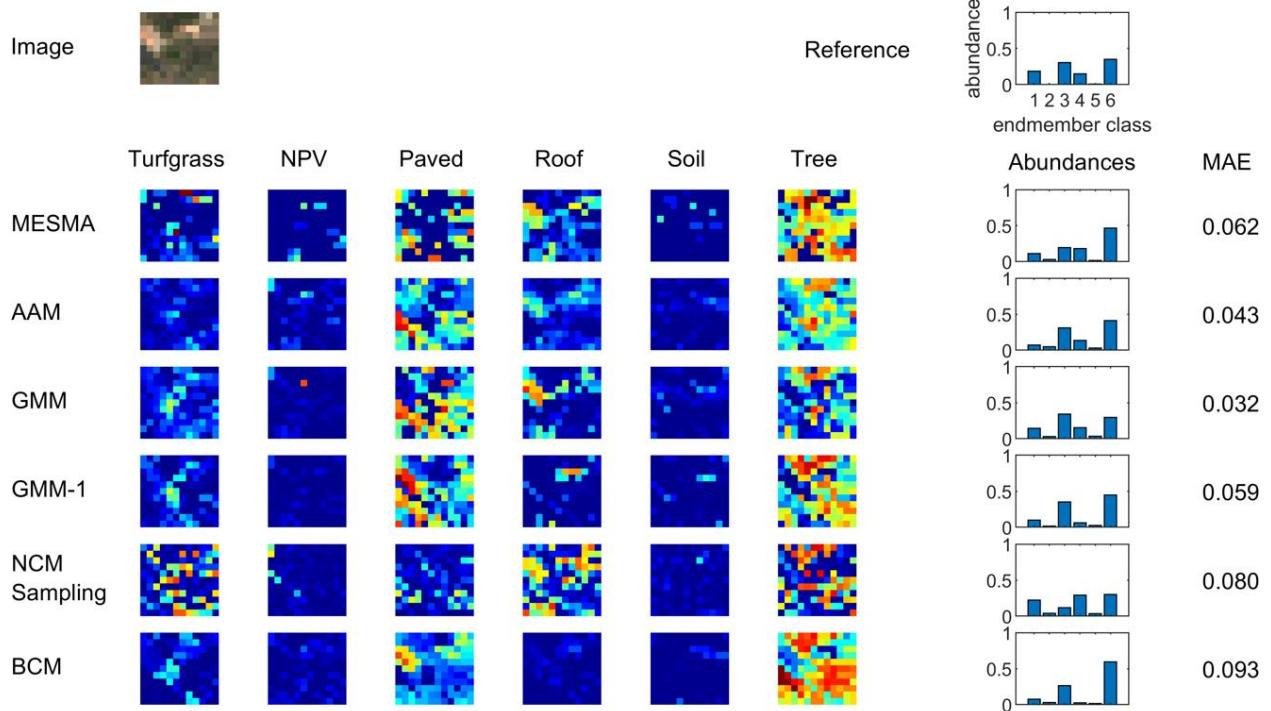


Figure S17: Abundance maps for polygon No. 17.

16 m



4 m

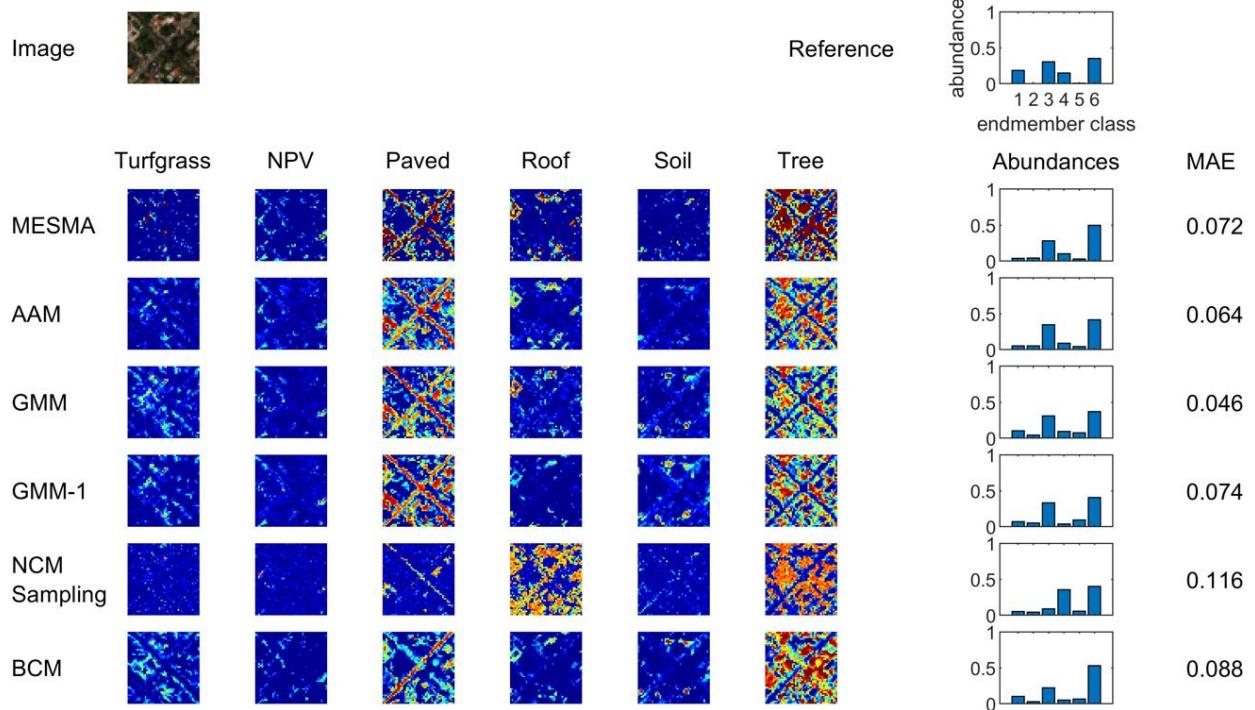
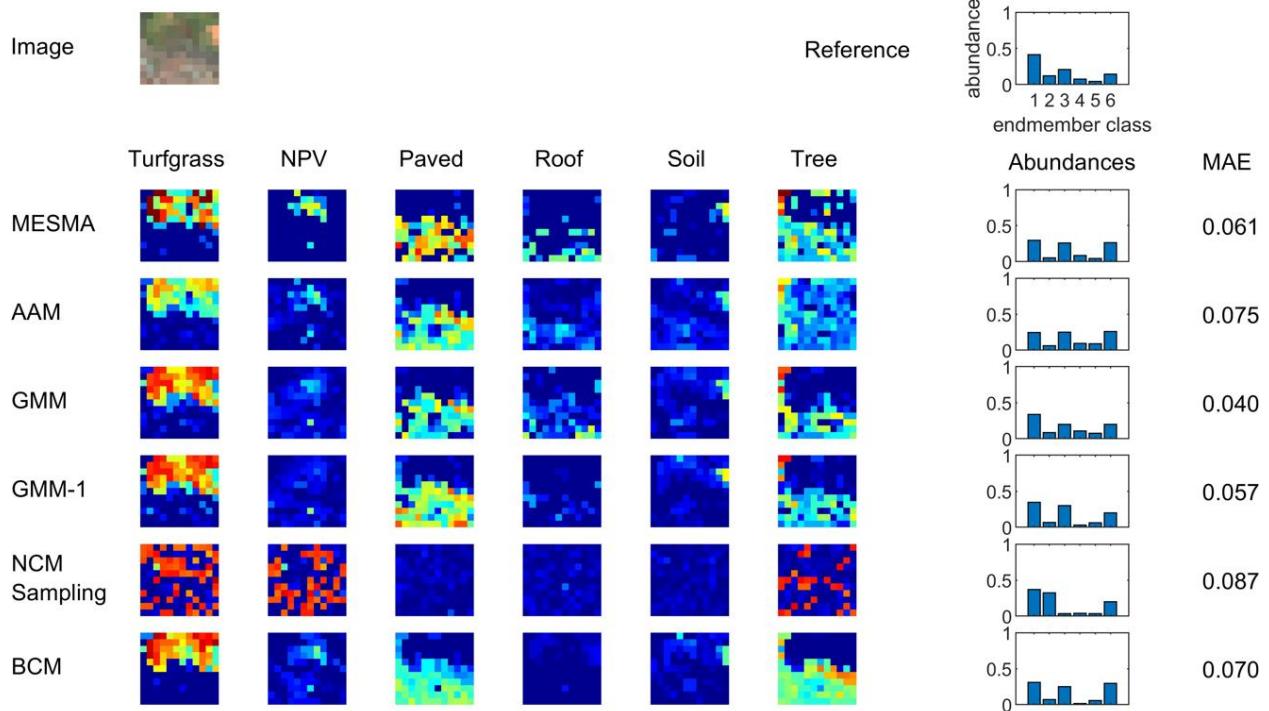


Figure S18: Abundance maps for polygon No. 18.

16 m



4 m

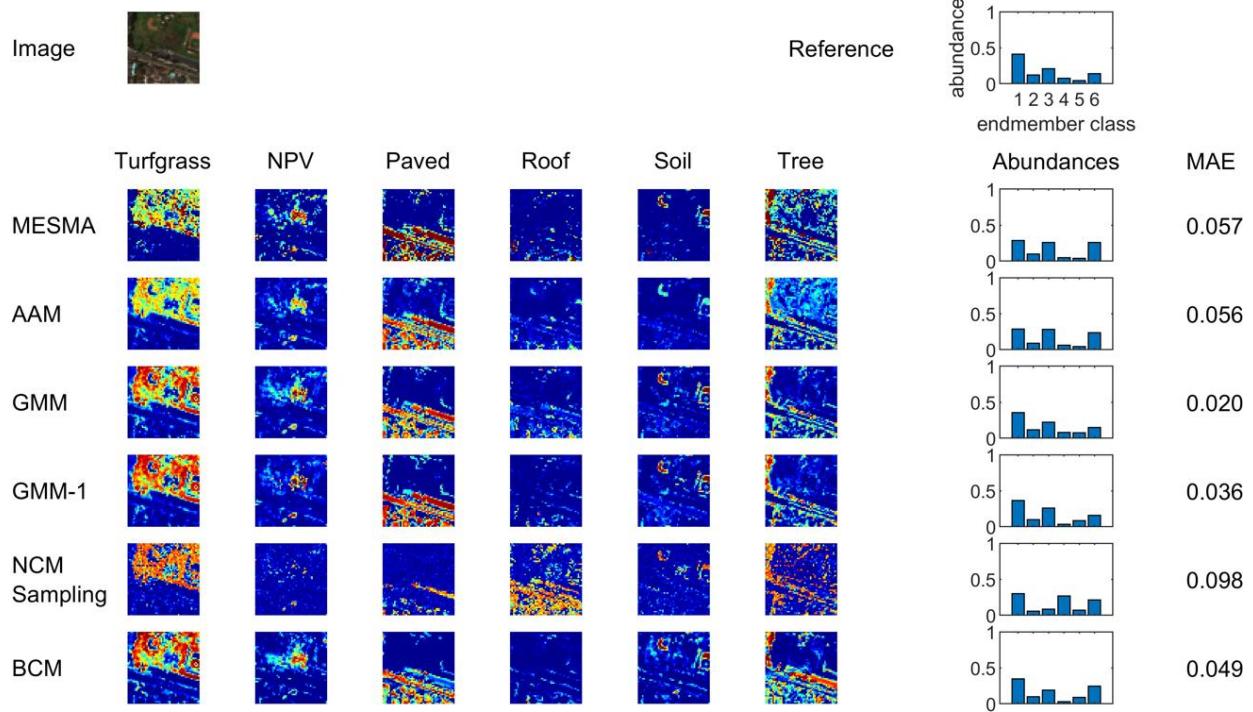
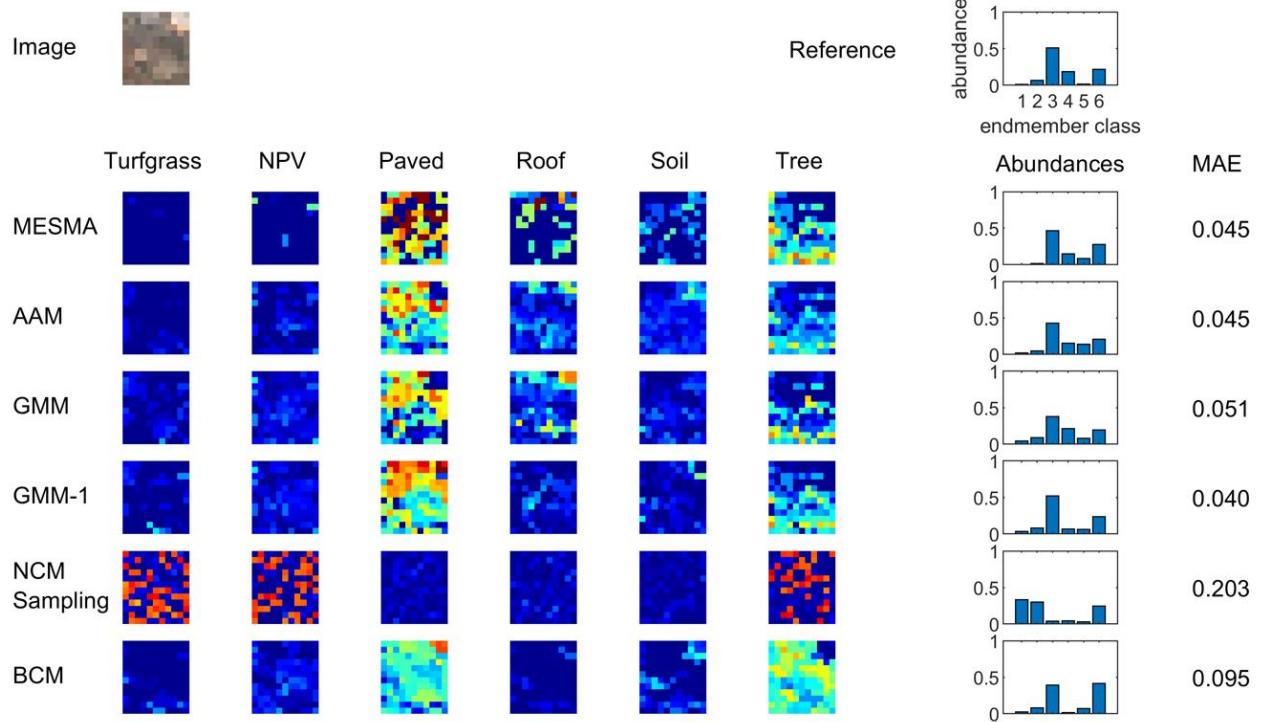


Figure S19: Abundance maps for polygon No. 19.

16 m



4 m

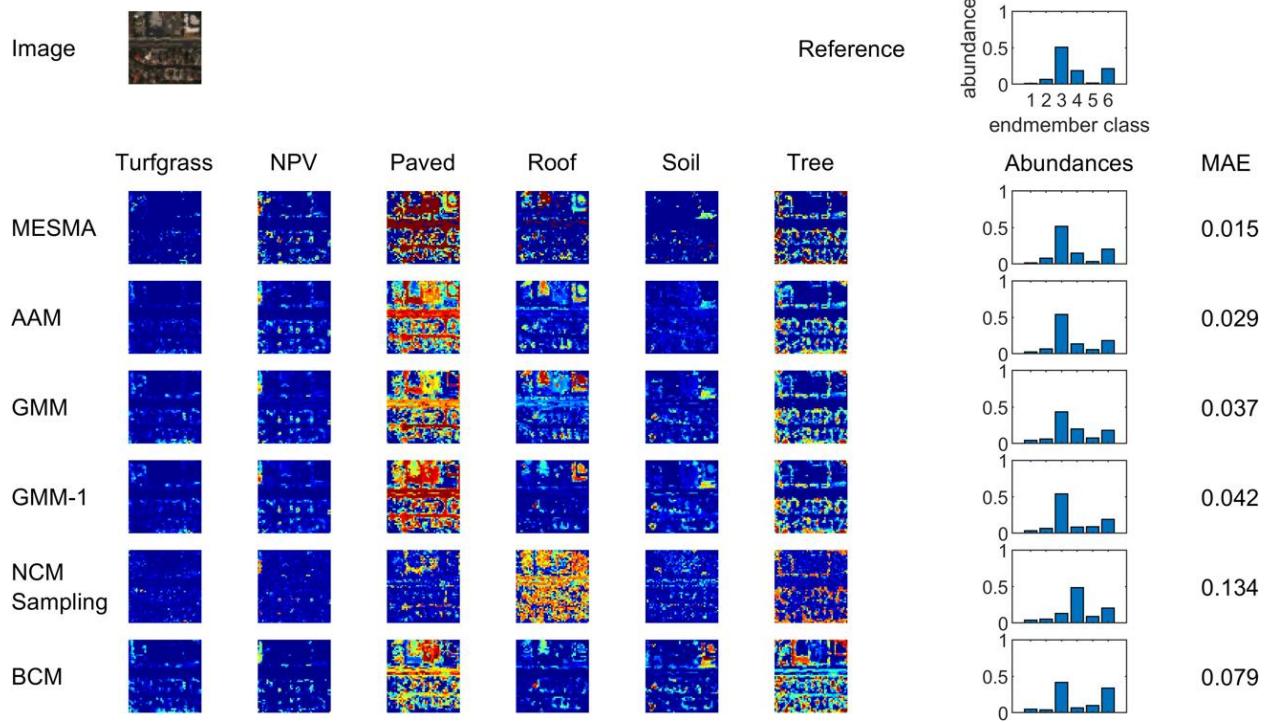
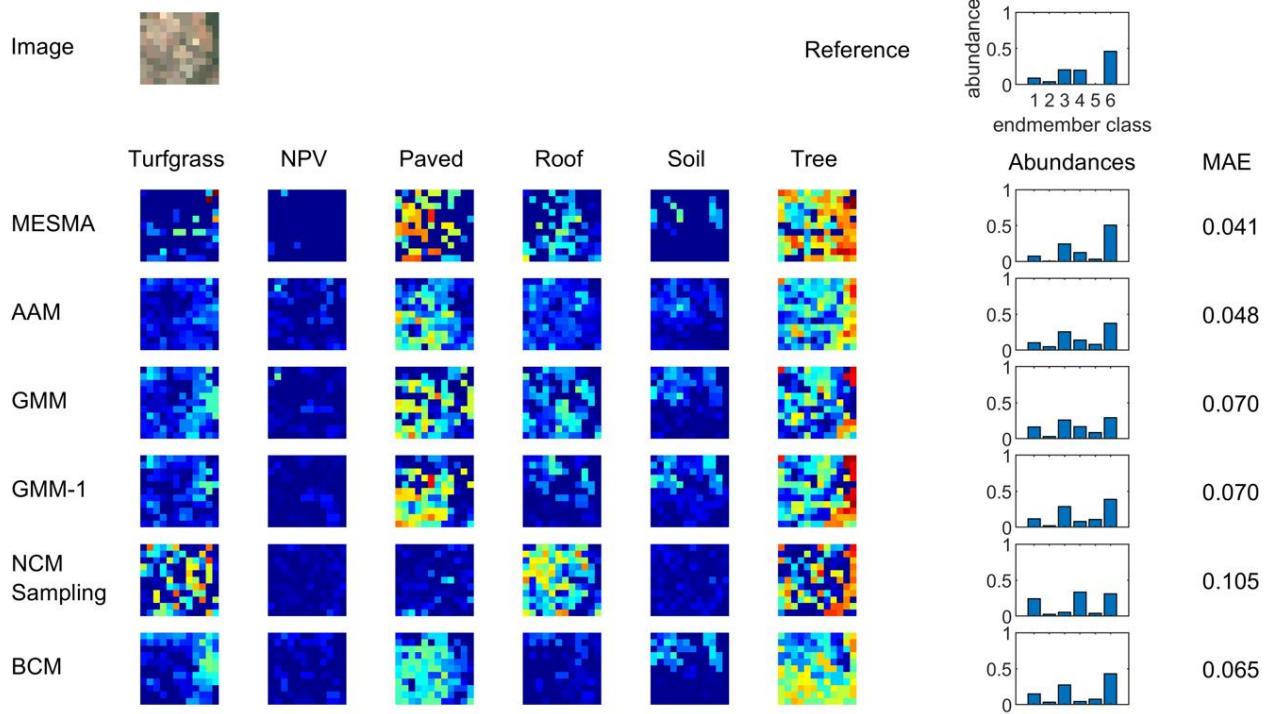


Figure S20: Abundance maps for polygon No. 20.

16 m



4 m

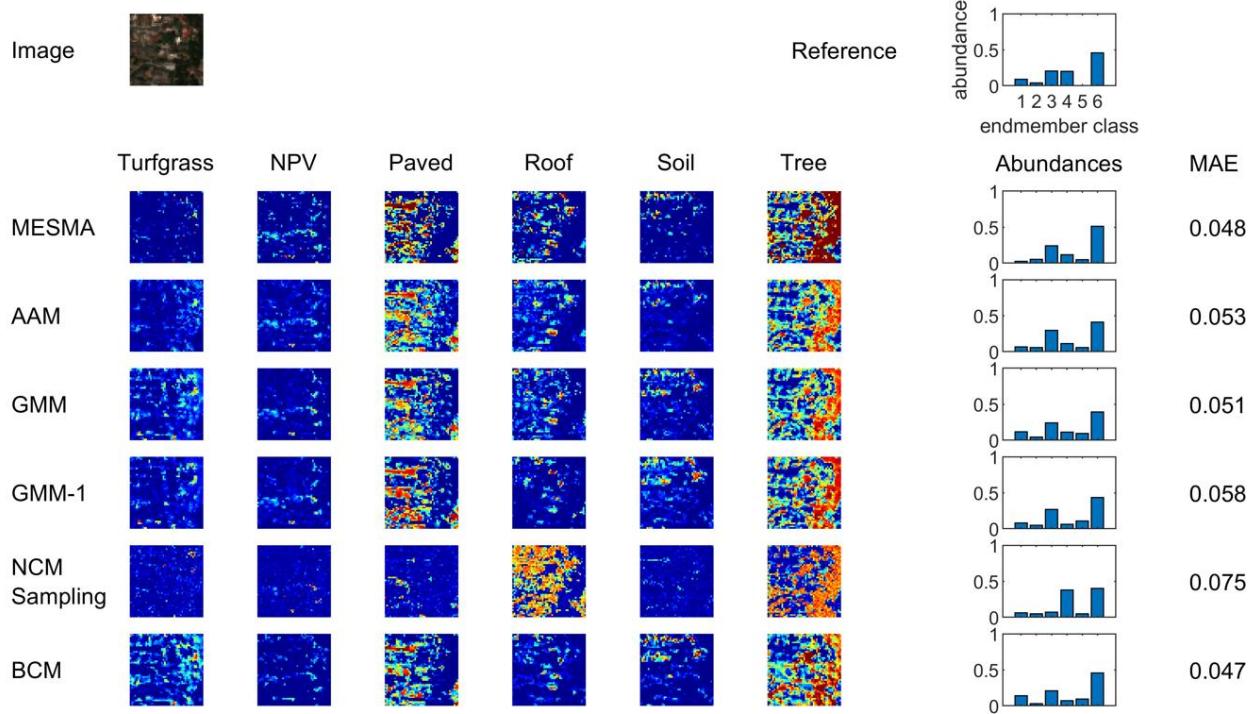
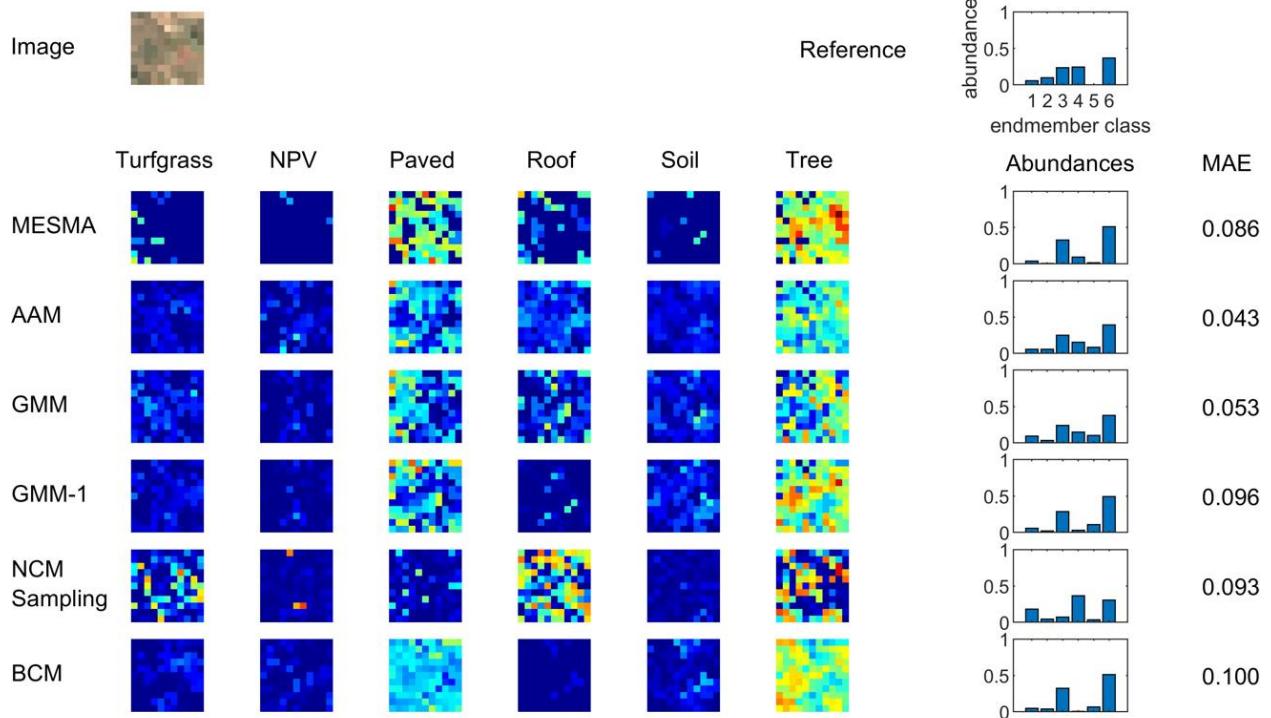


Figure S21: Abundance maps for polygon No. 21.

16 m



4 m

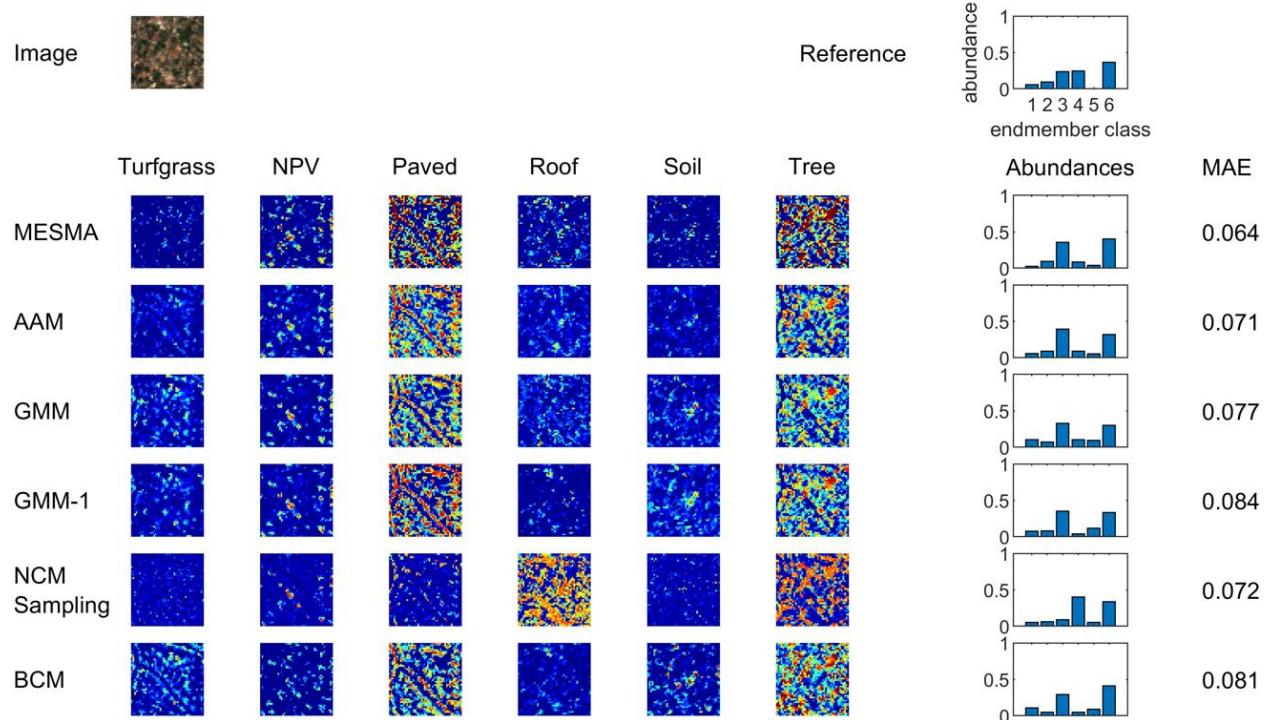
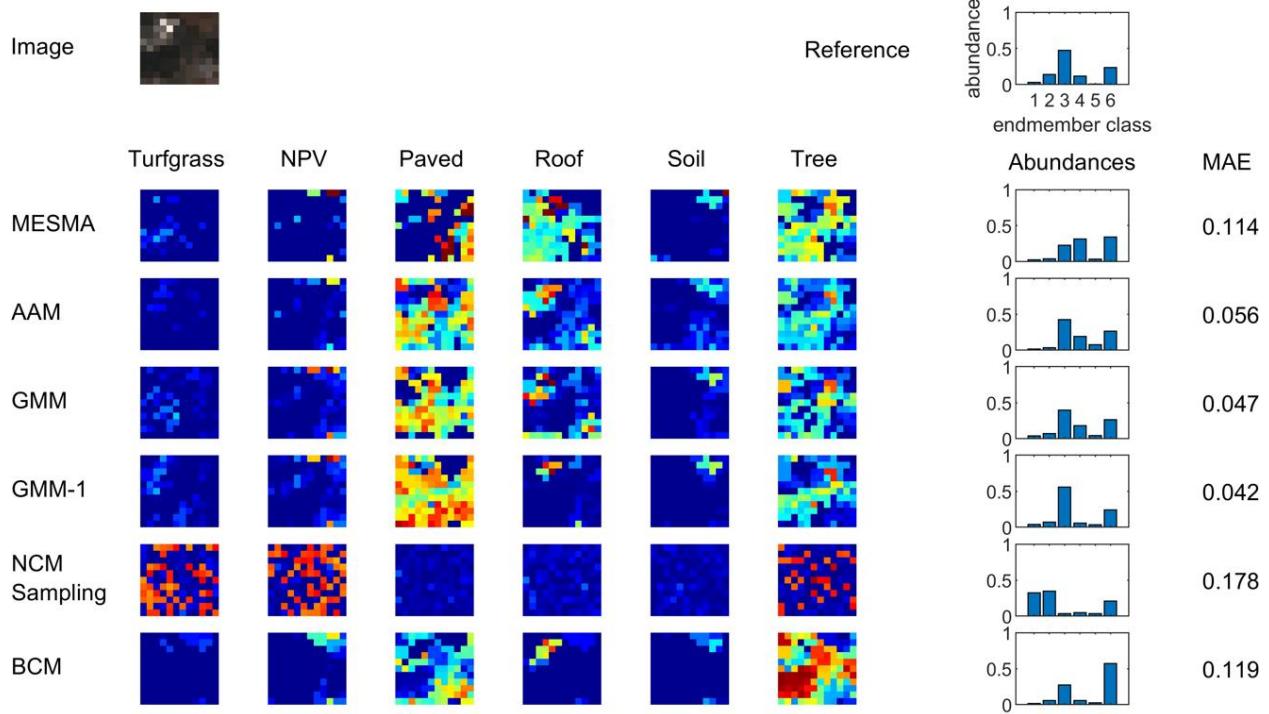


Figure S22: Abundance maps for polygon No. 22.

16 m



4 m

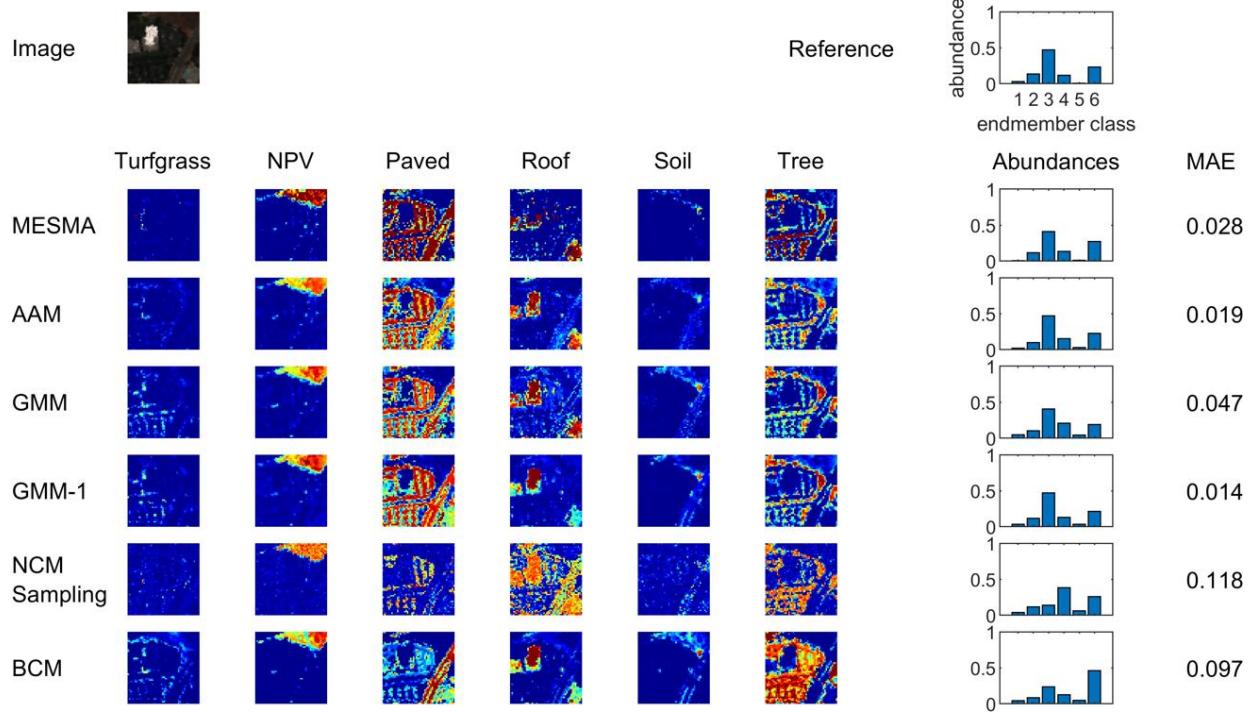
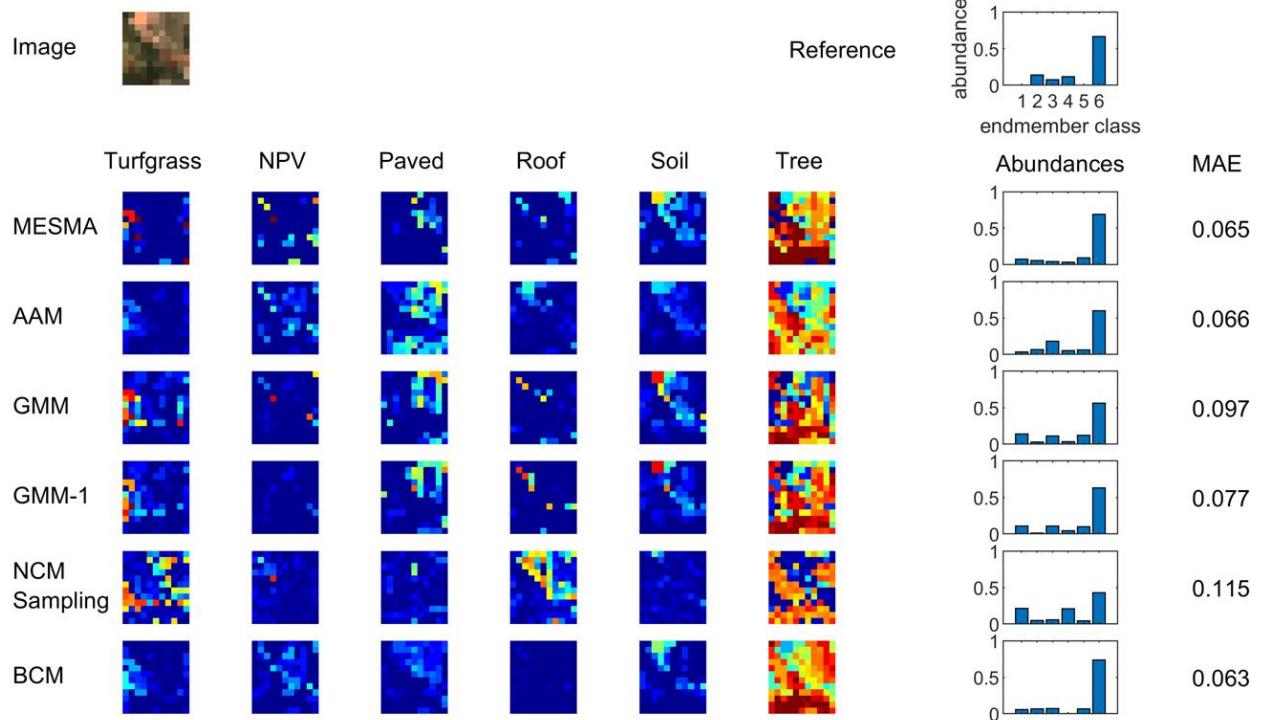


Figure S23: Abundance maps for polygon No. 23.

16 m



4 m

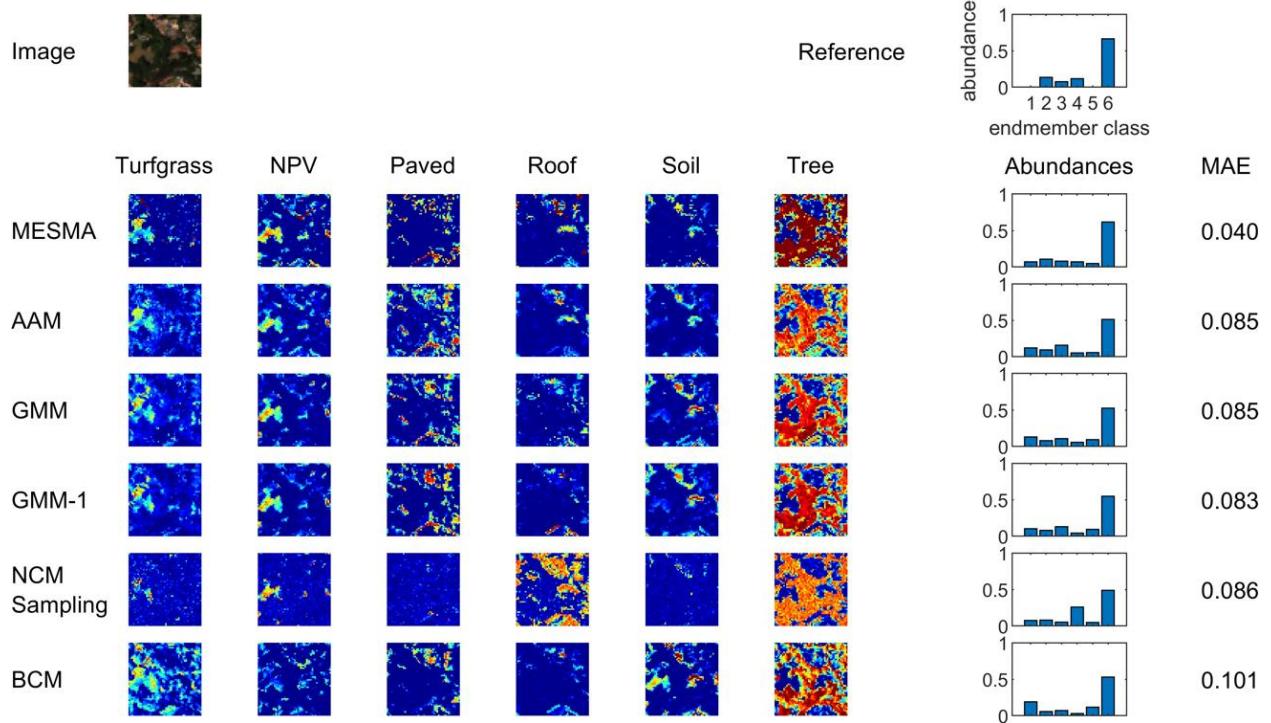
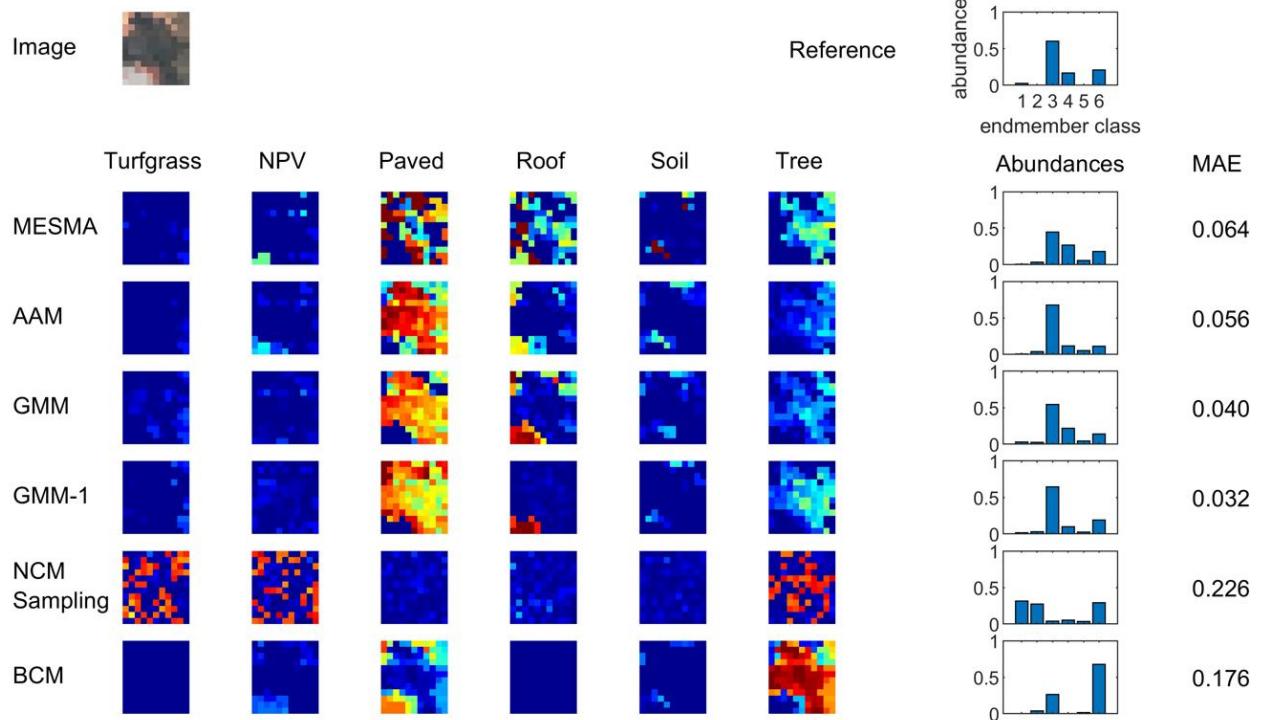


Figure S24: Abundance maps for polygon No. 24.

16 m



4 m

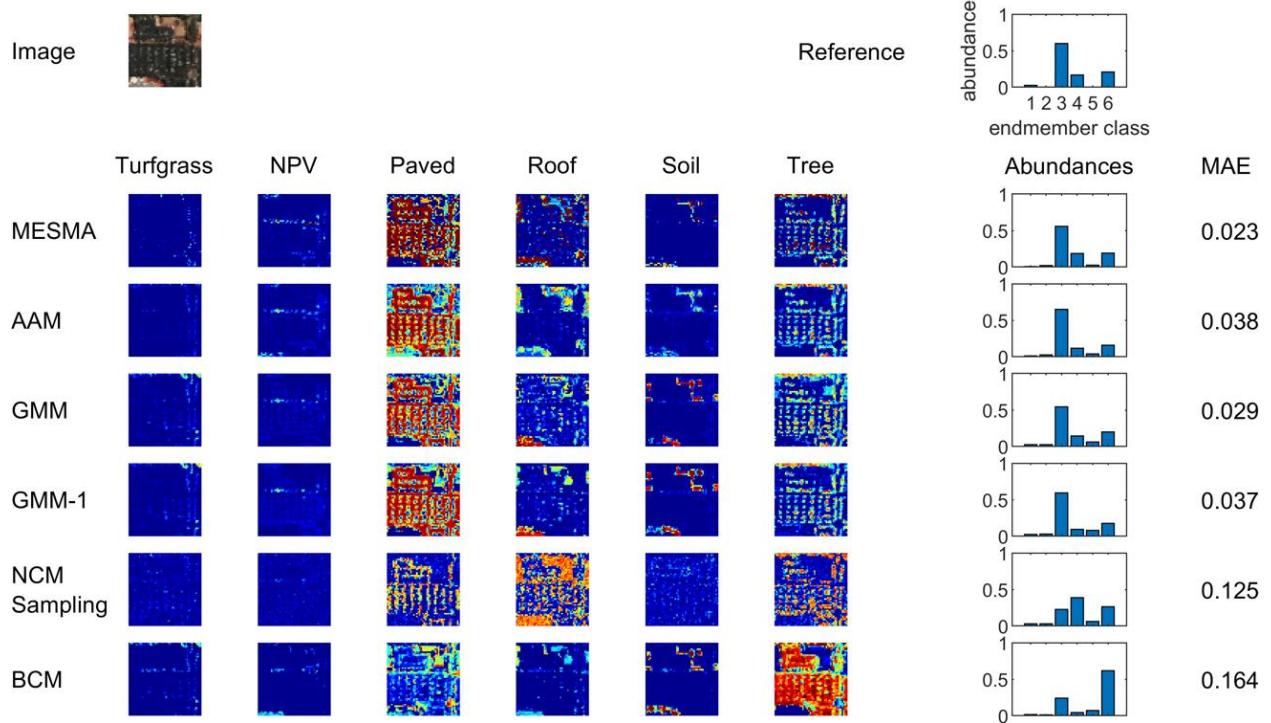
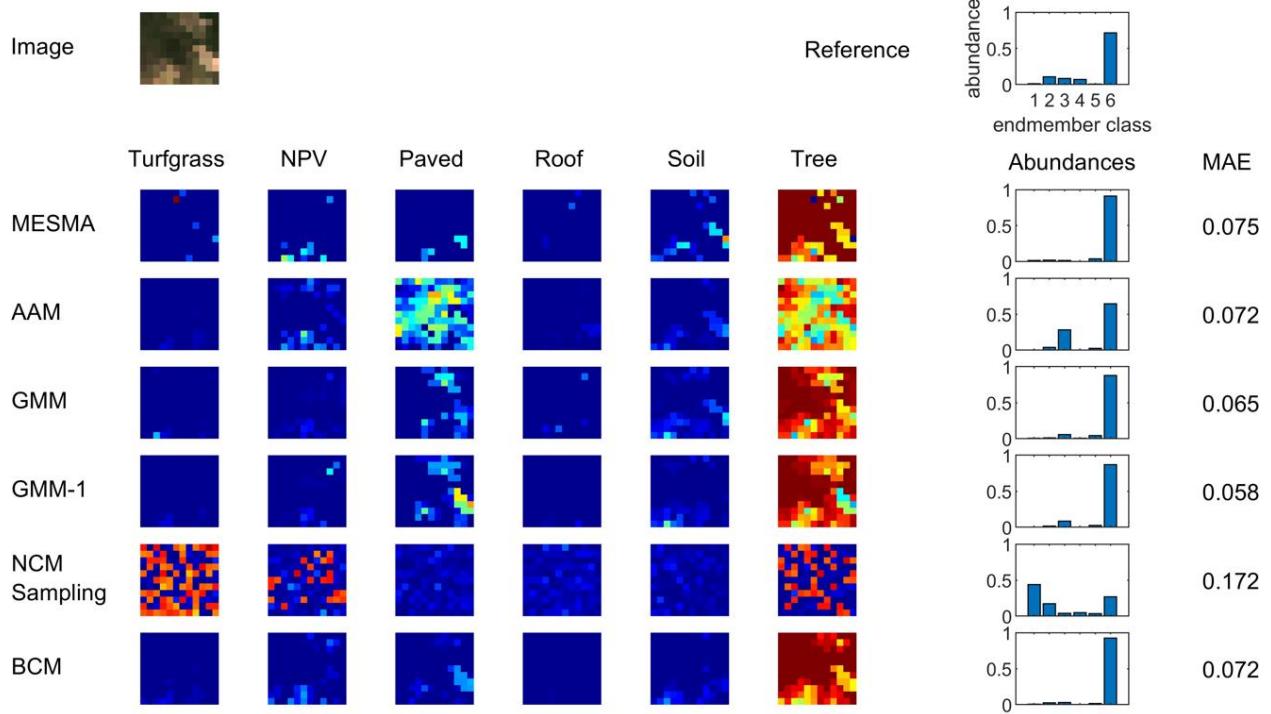


Figure S25: Abundance maps for polygon No. 25.

16 m



4 m

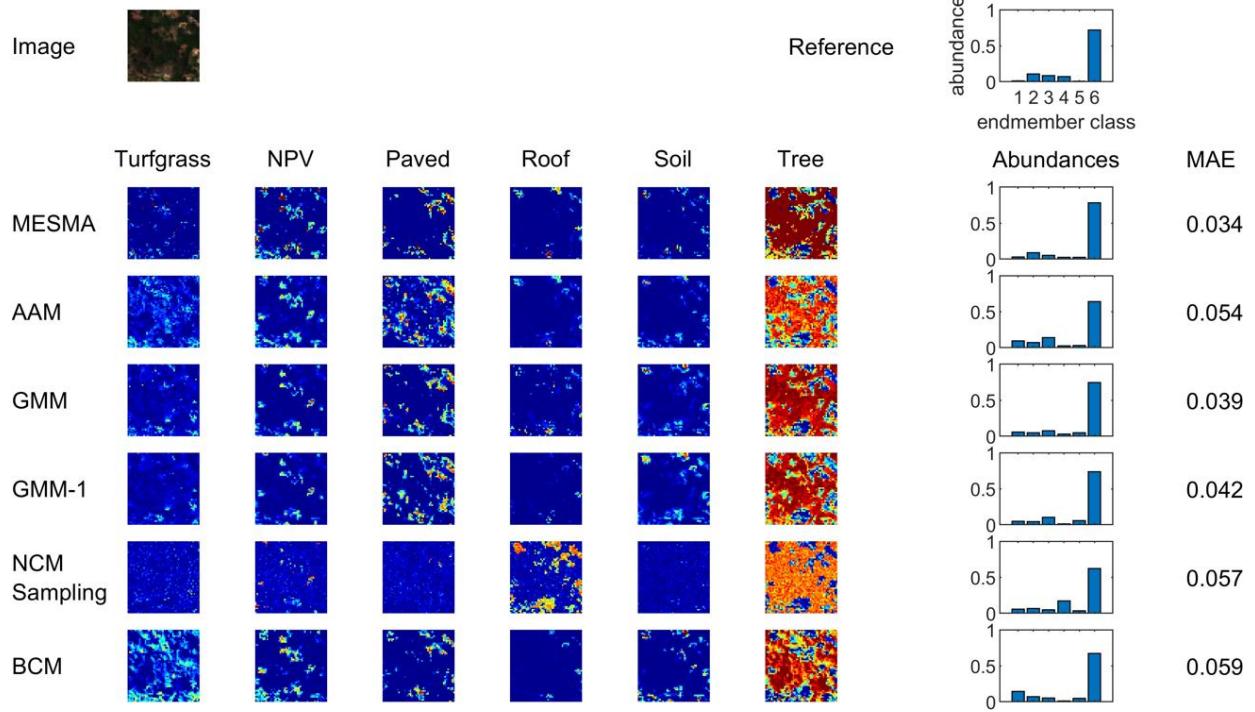
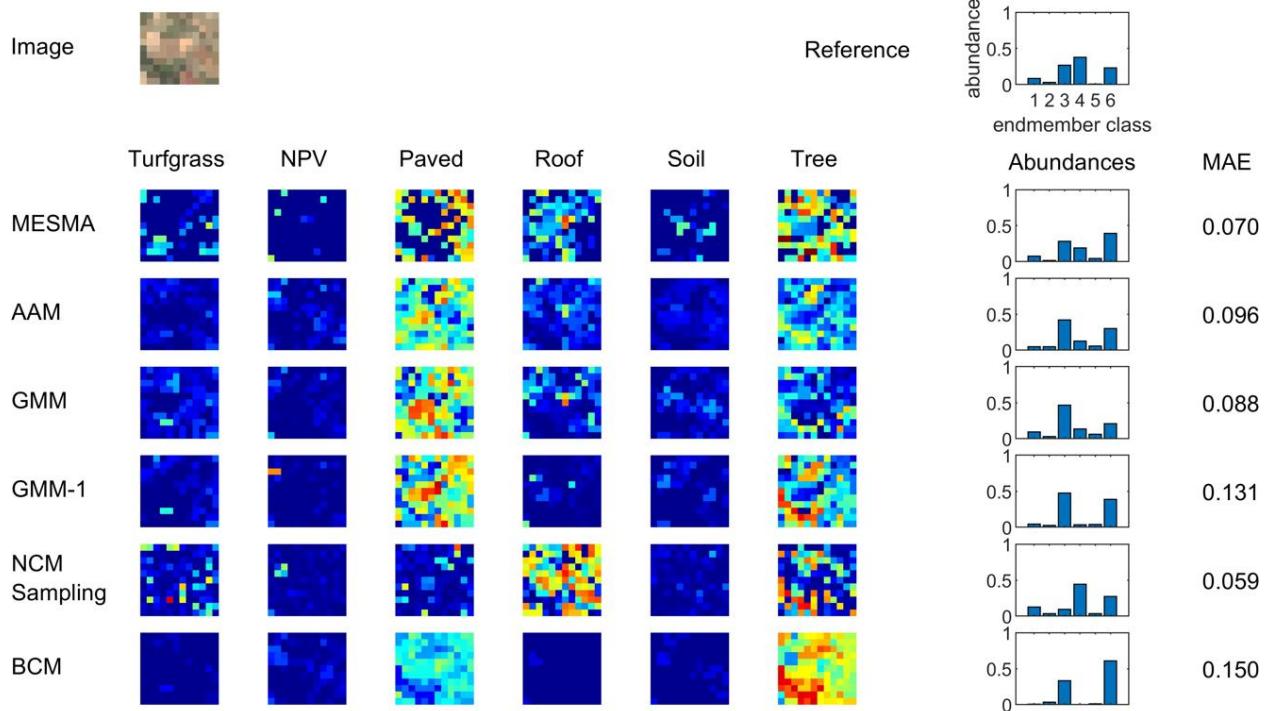


Figure S26: Abundance maps for polygon No. 26.

16 m



4 m

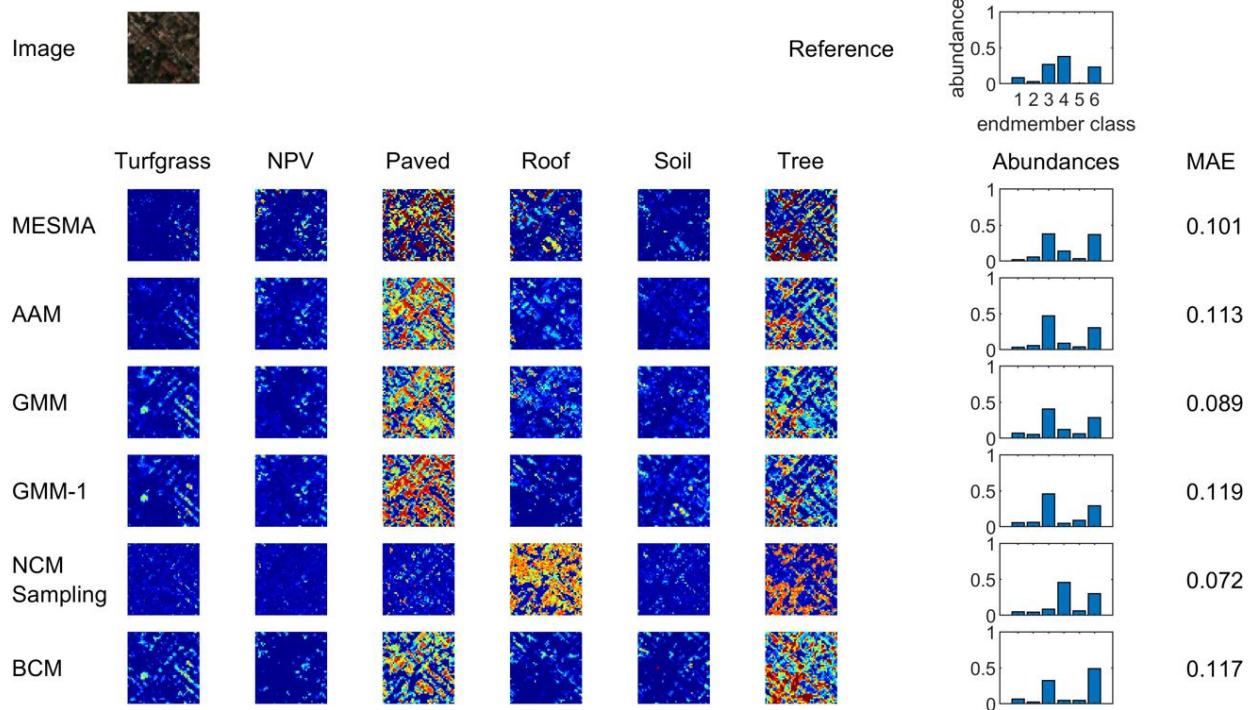
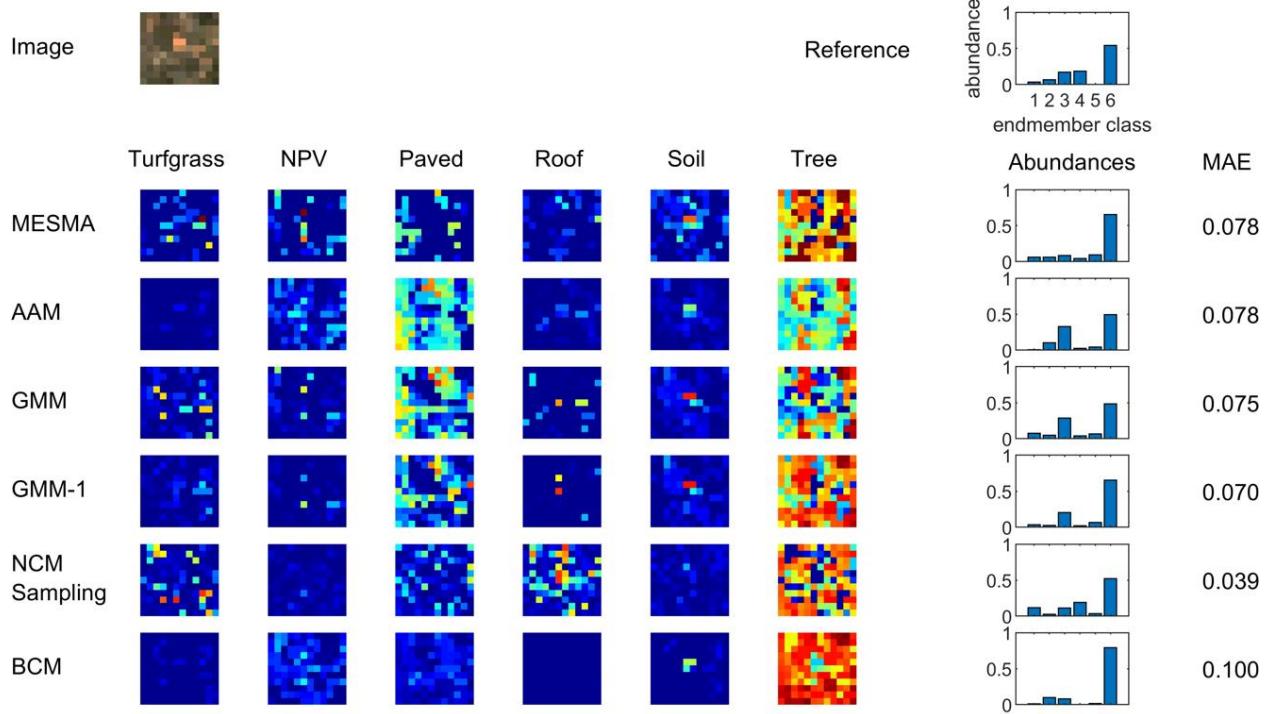


Figure S27: Abundance maps for polygon No. 27.

16 m



4 m

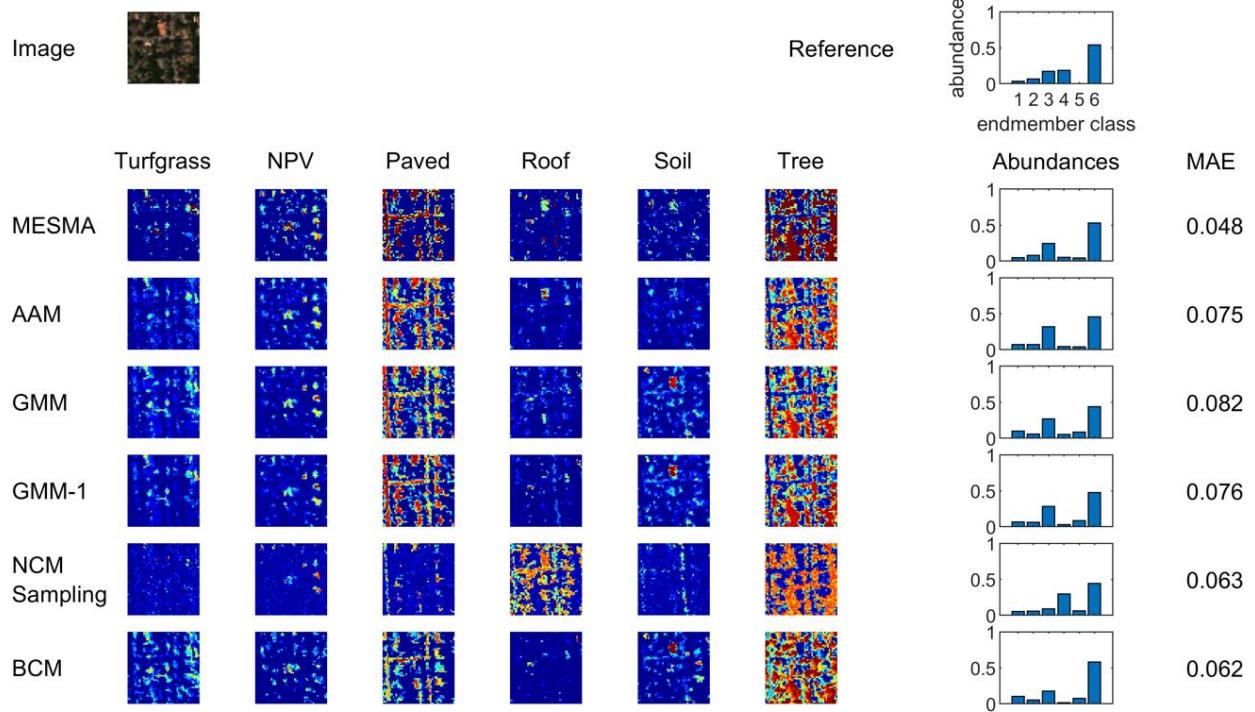
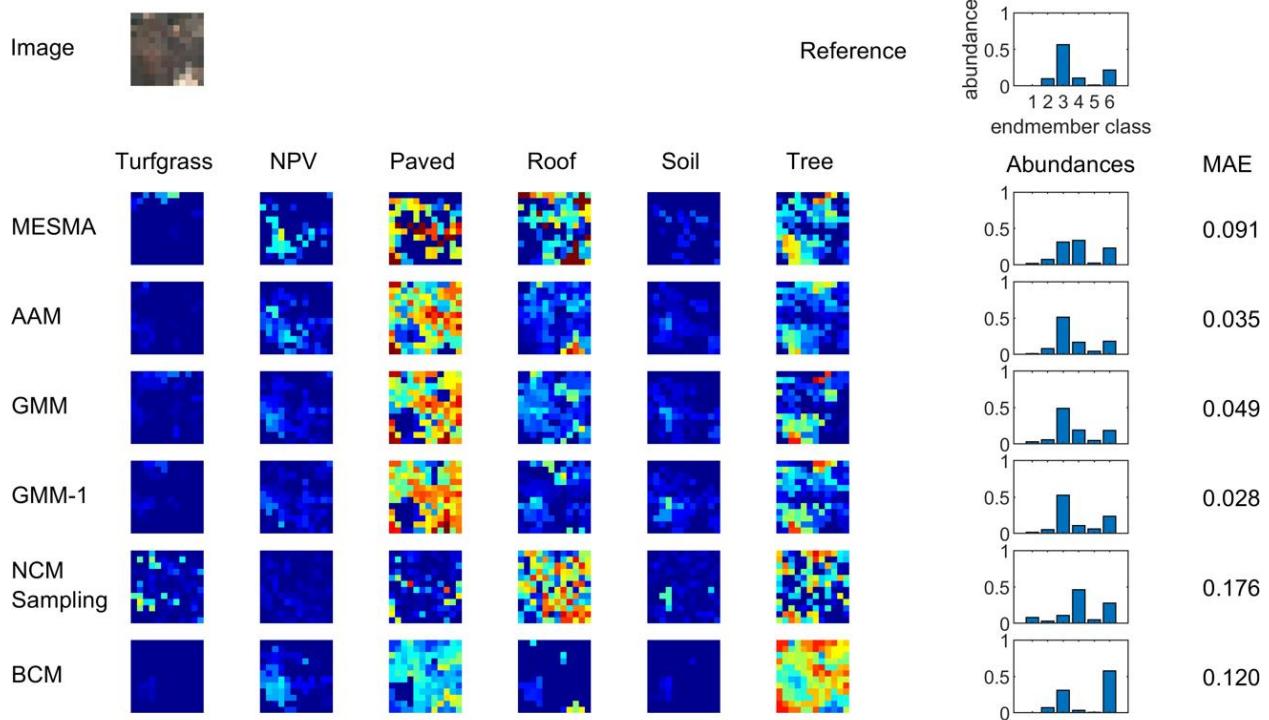


Figure S28: Abundance maps for polygon No. 28.

16 m



4 m

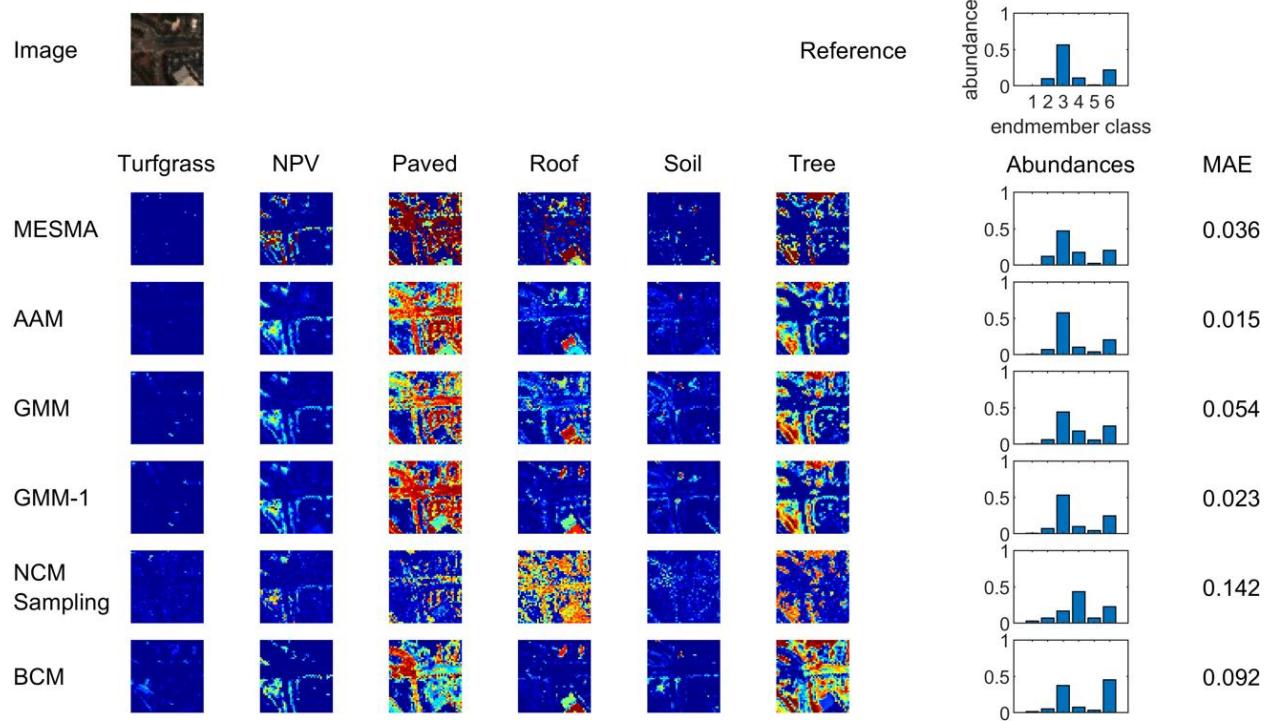
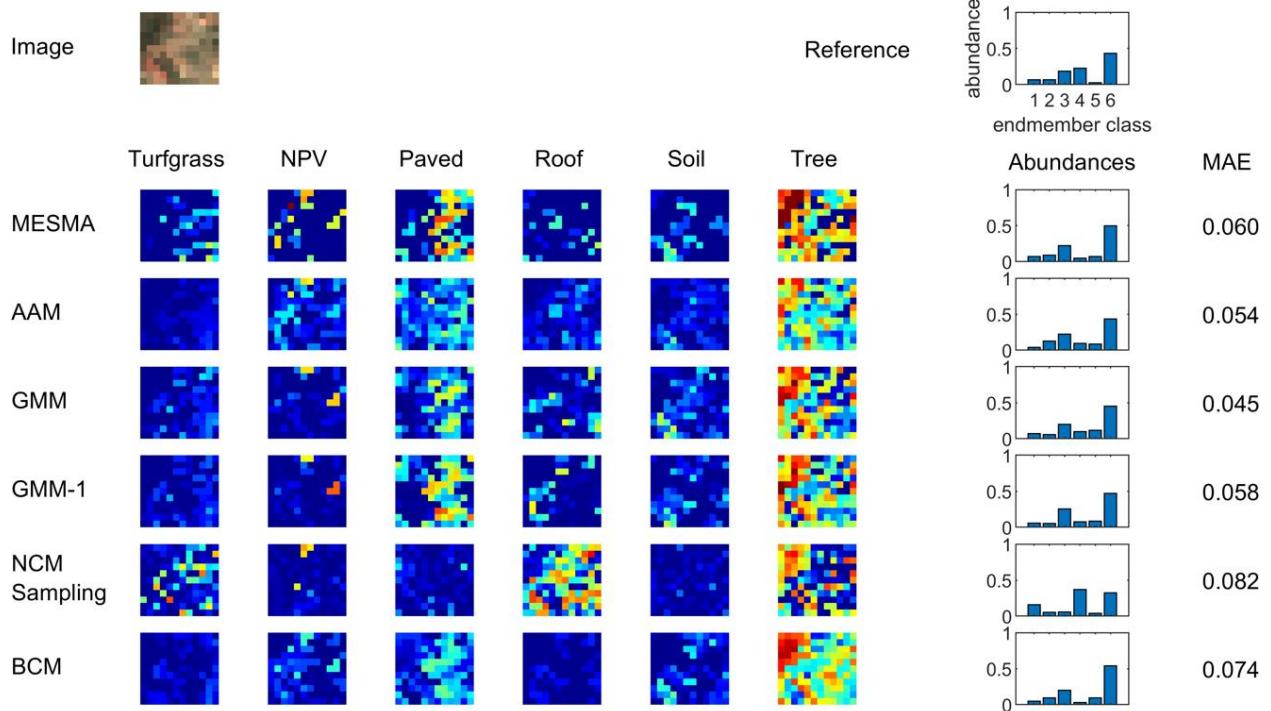


Figure S29: Abundance maps for polygon No. 29.

16 m



4 m

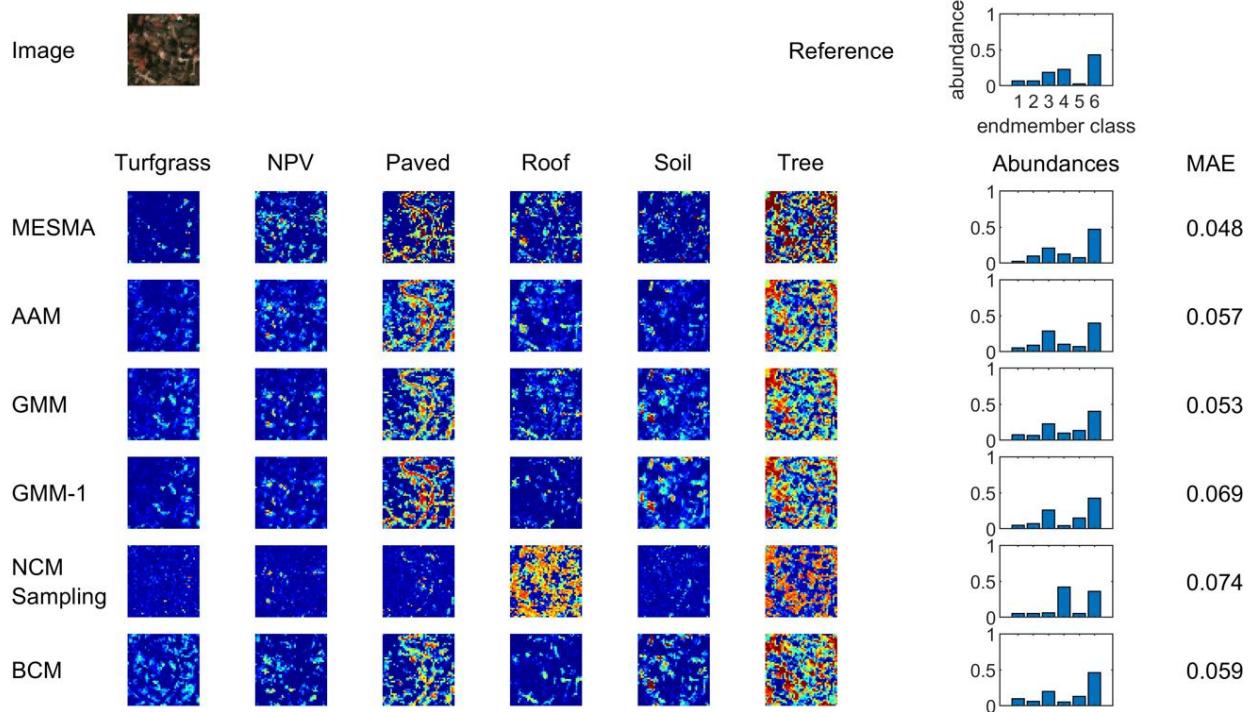
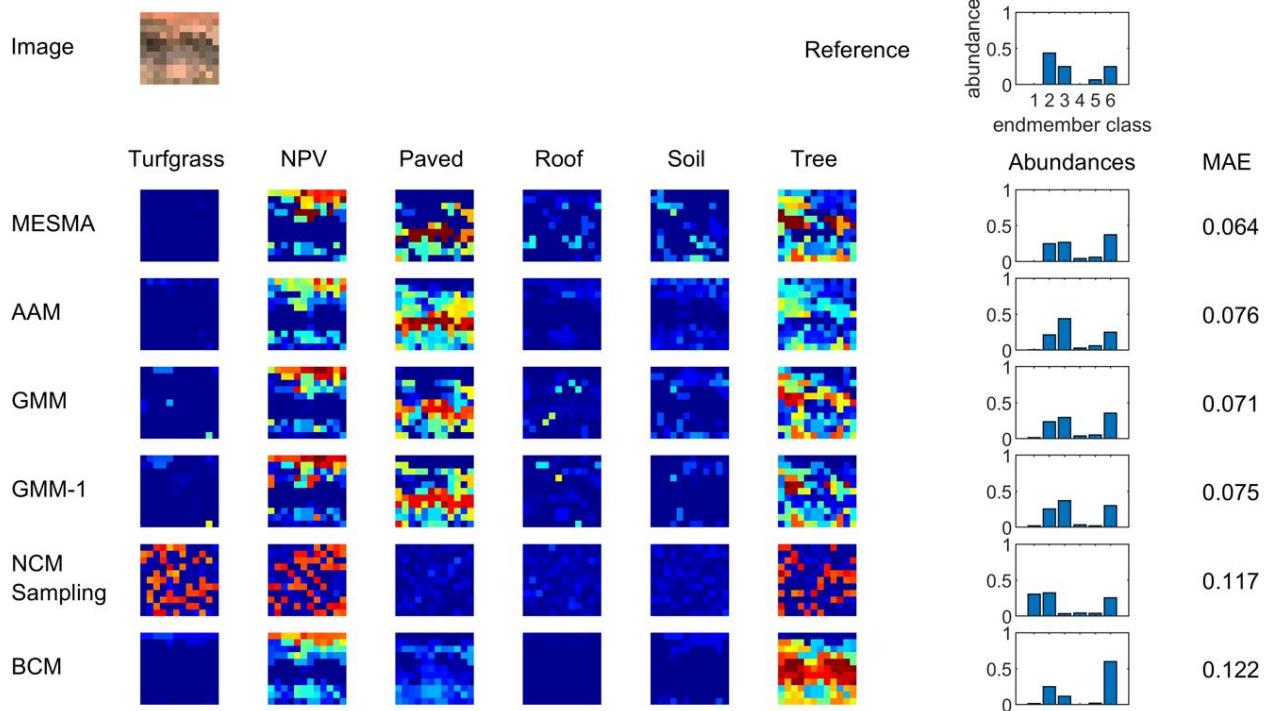


Figure S30: Abundance maps for polygon No. 30.

16 m



4 m

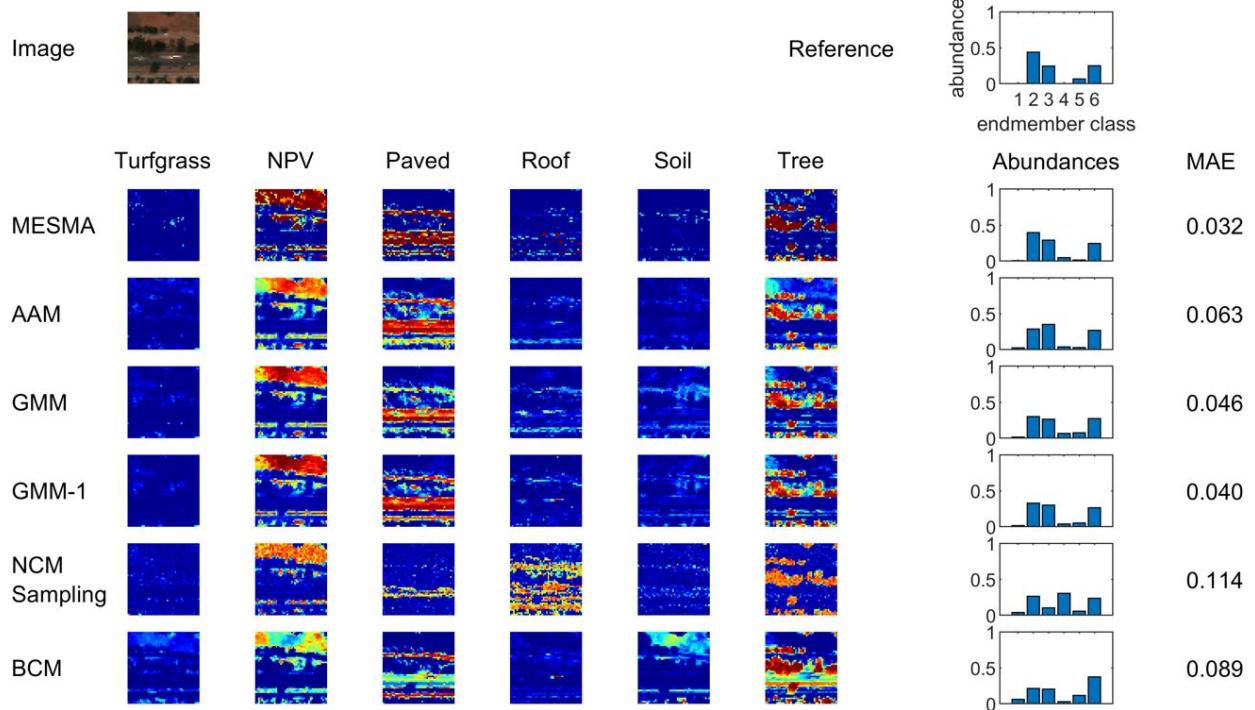
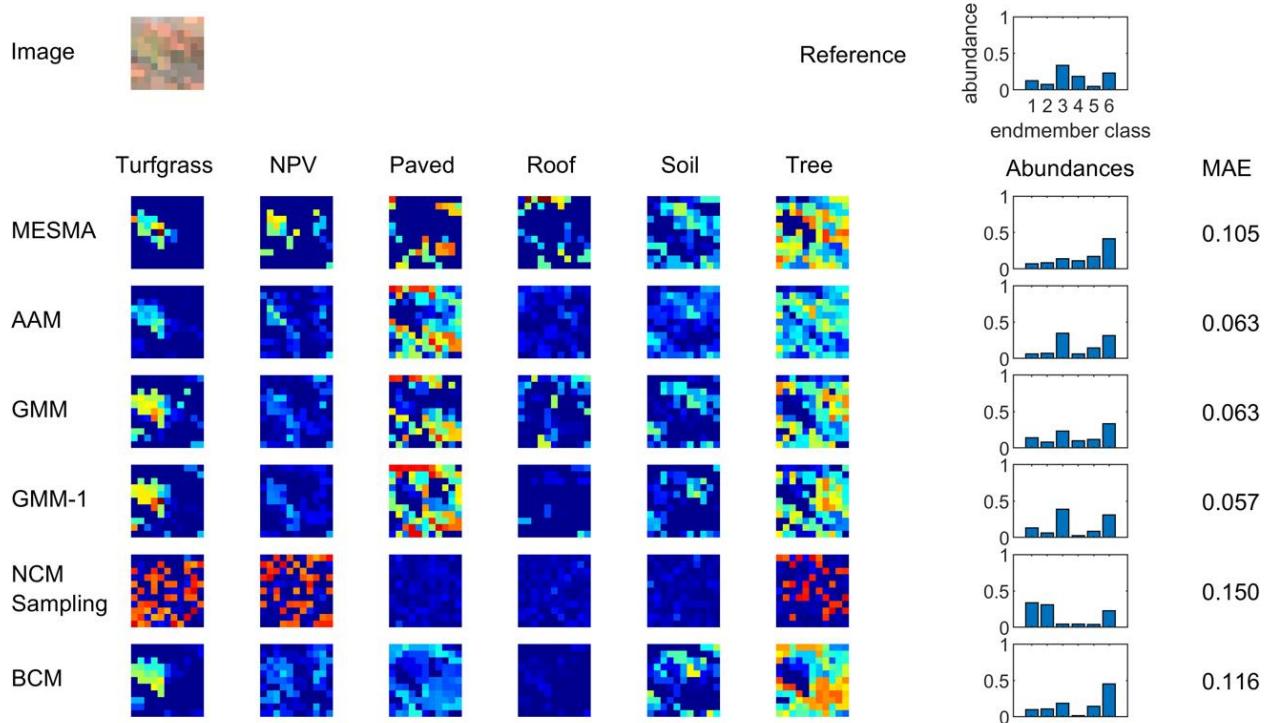


Figure S31: Abundance maps for polygon No. 31.

16 m



4 m

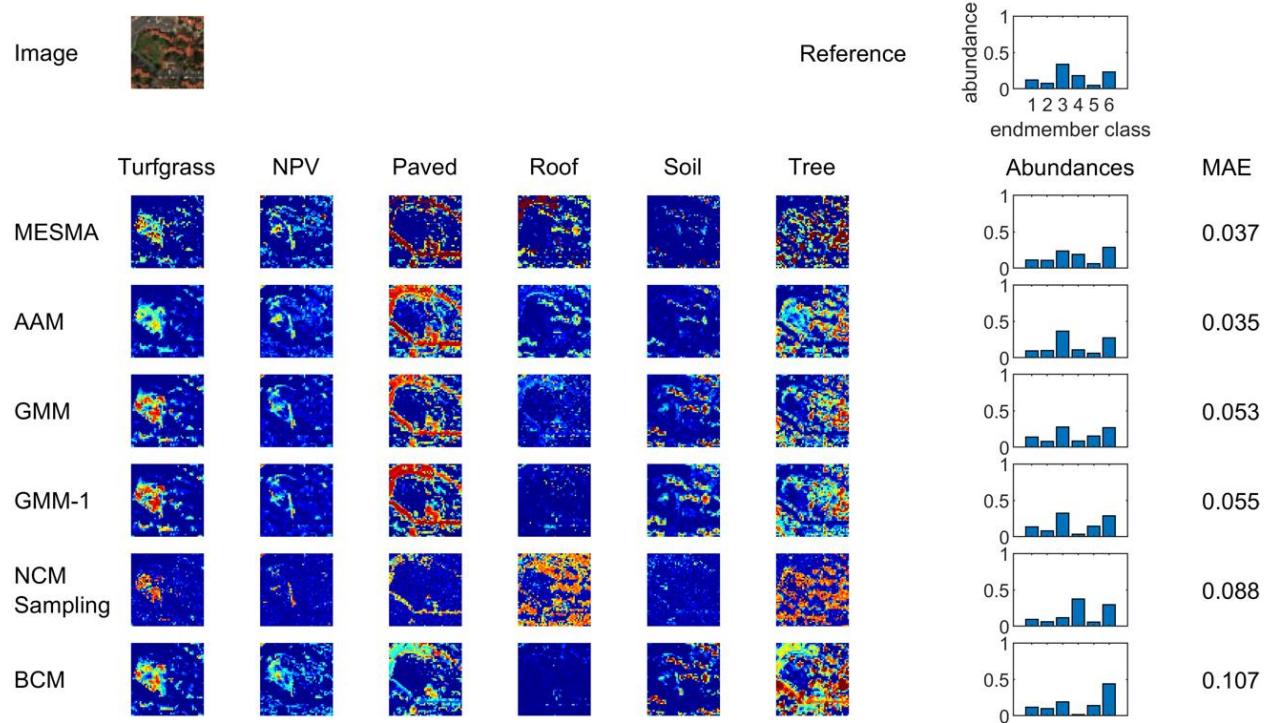
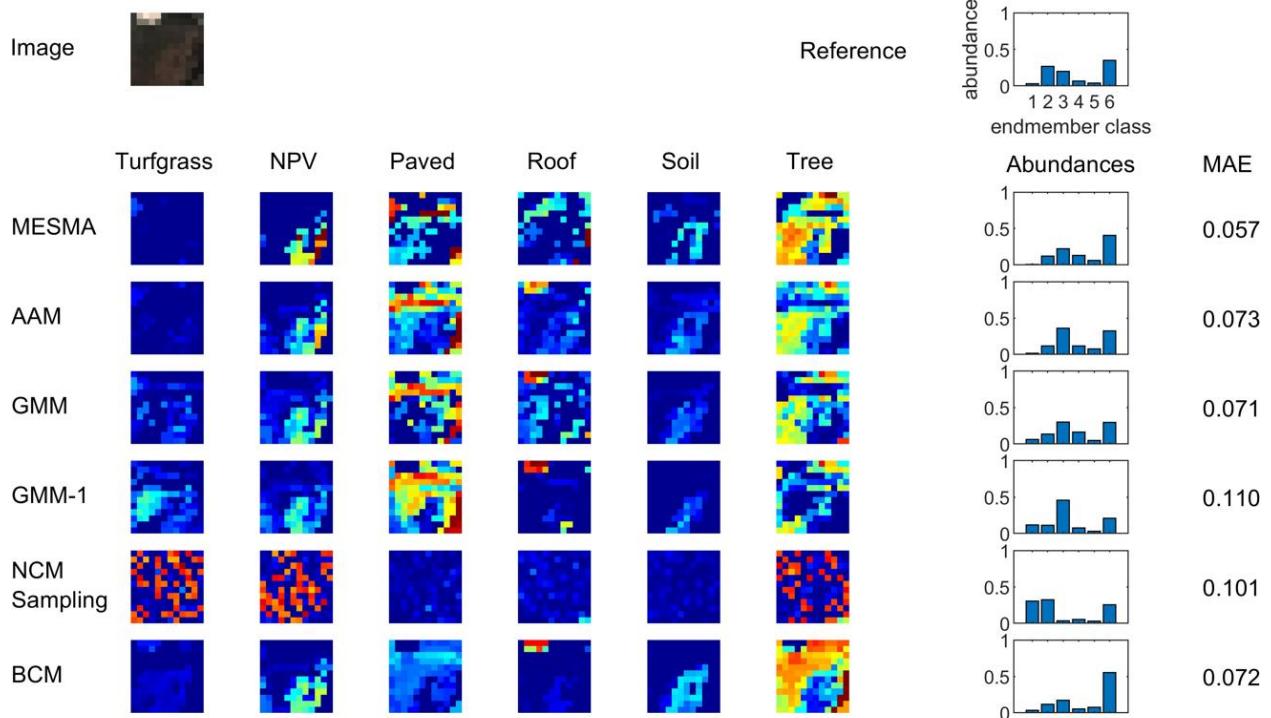


Figure S32: Abundance maps for polygon No. 32.

16 m



4 m

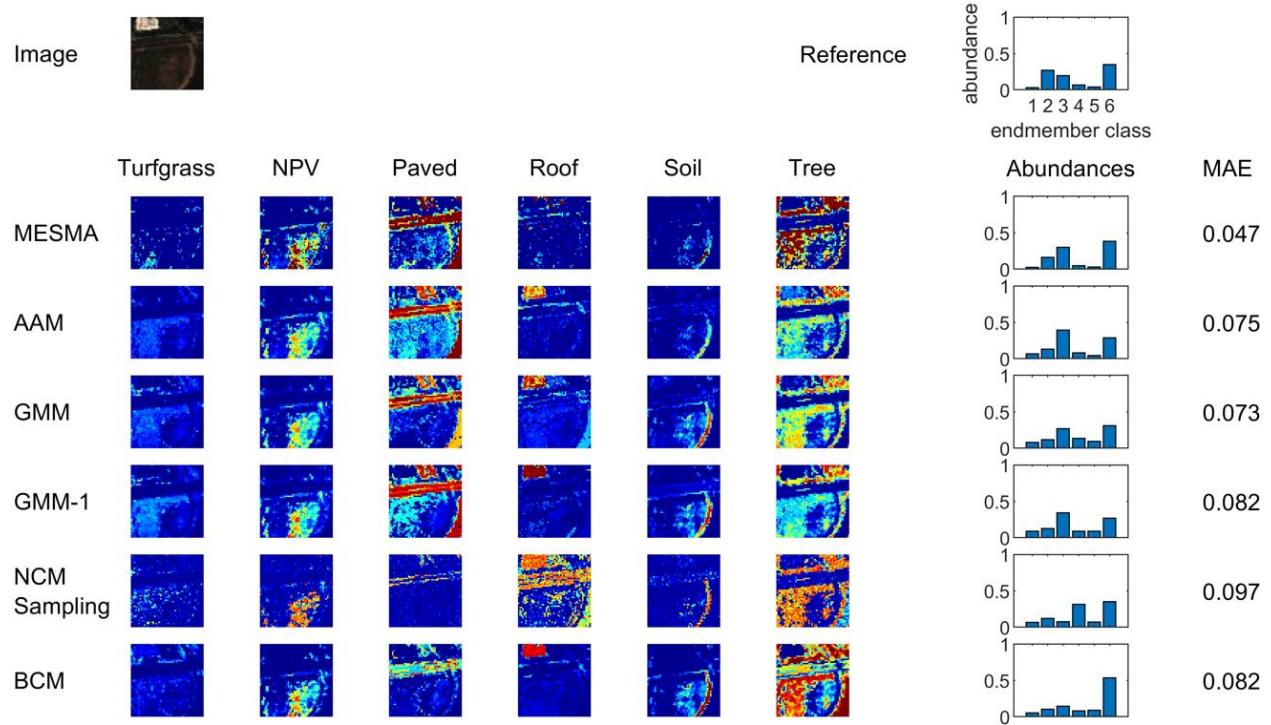
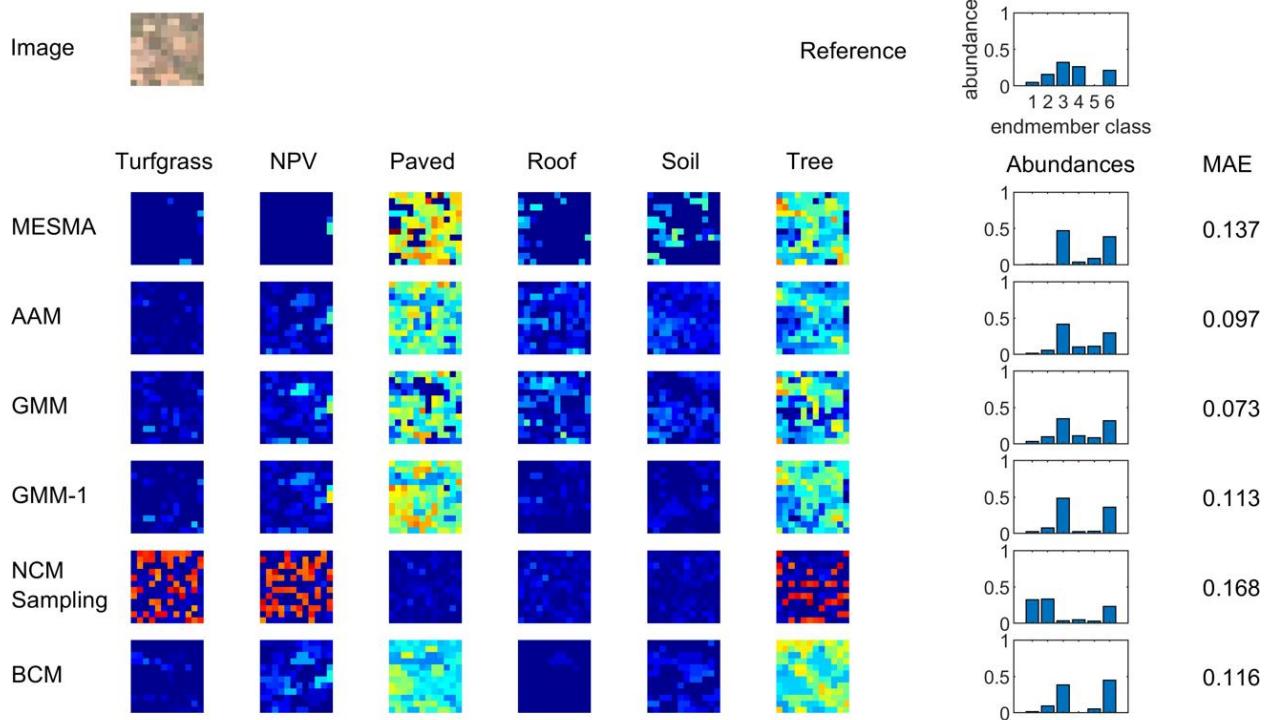


Figure S33: Abundance maps for polygon No. 33.

16 m



4 m

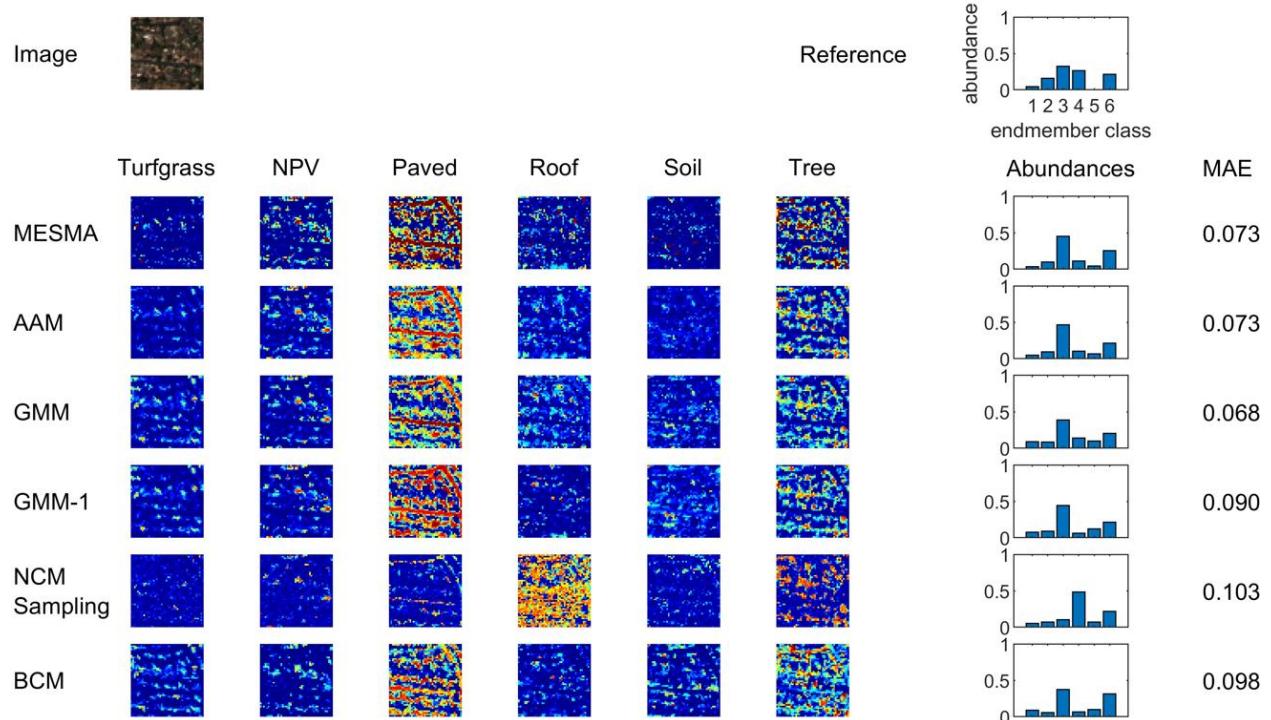
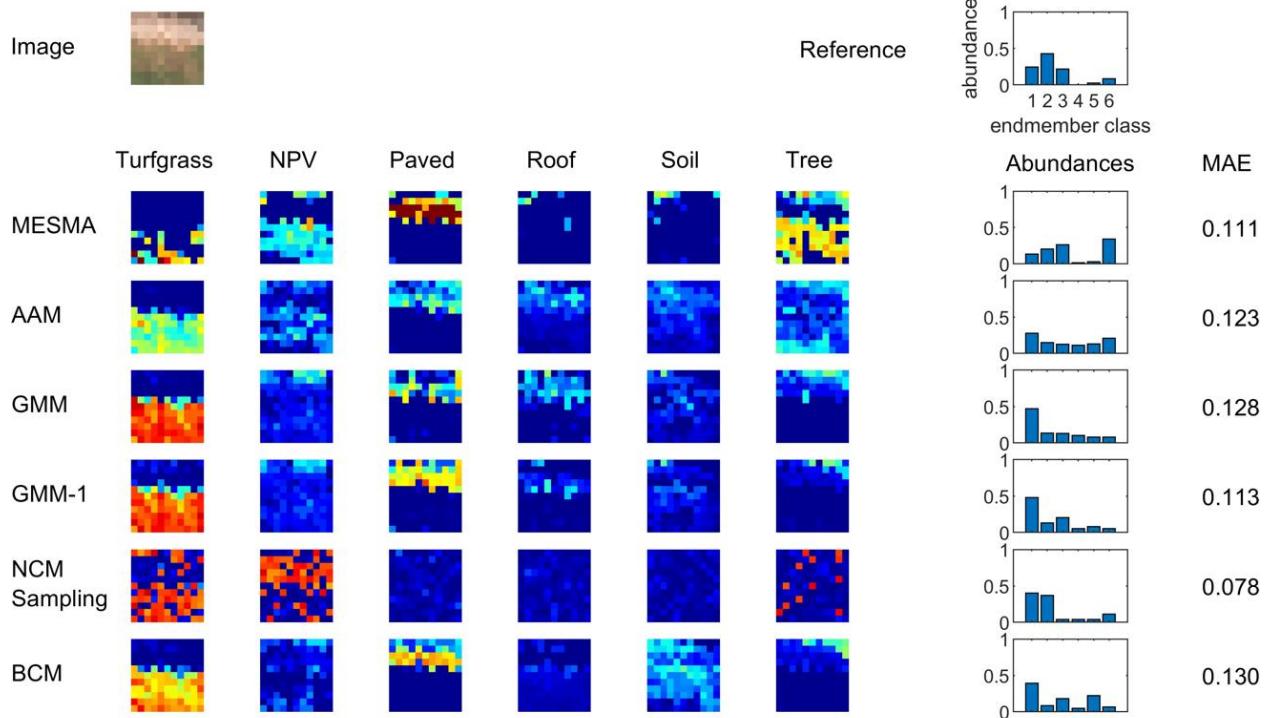


Figure S34: Abundance maps for polygon No. 34.

16 m



4 m

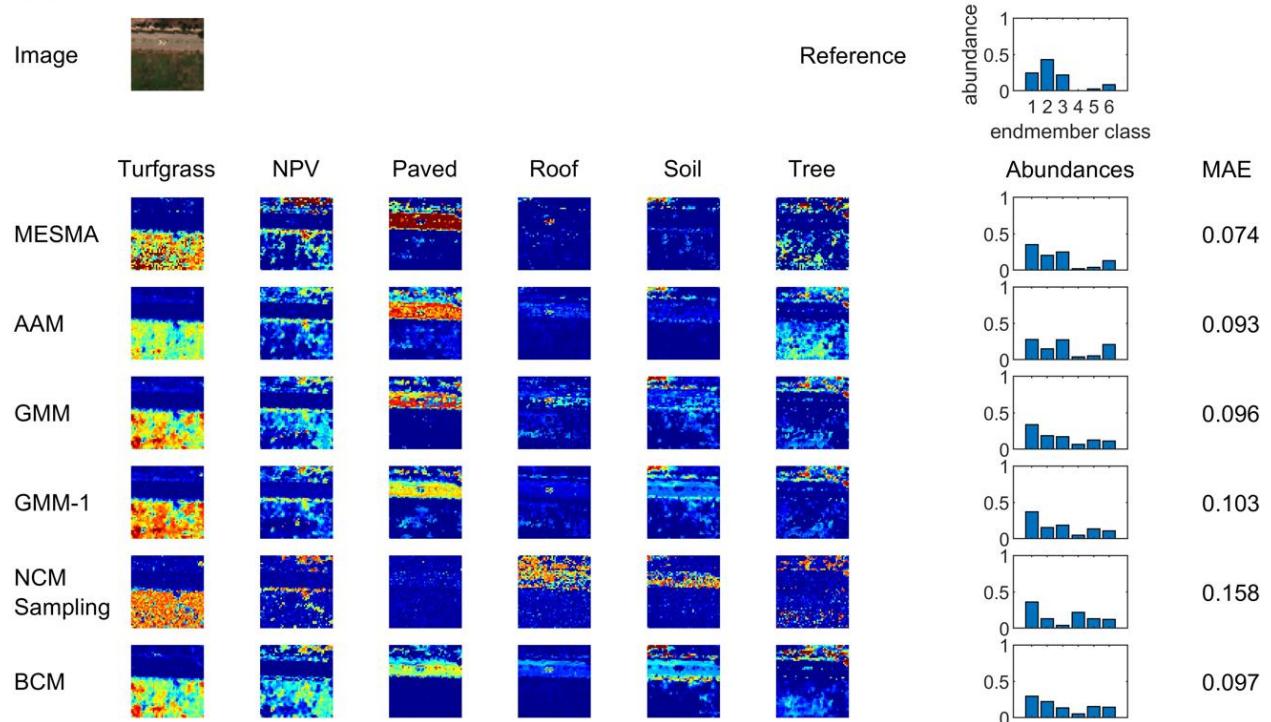
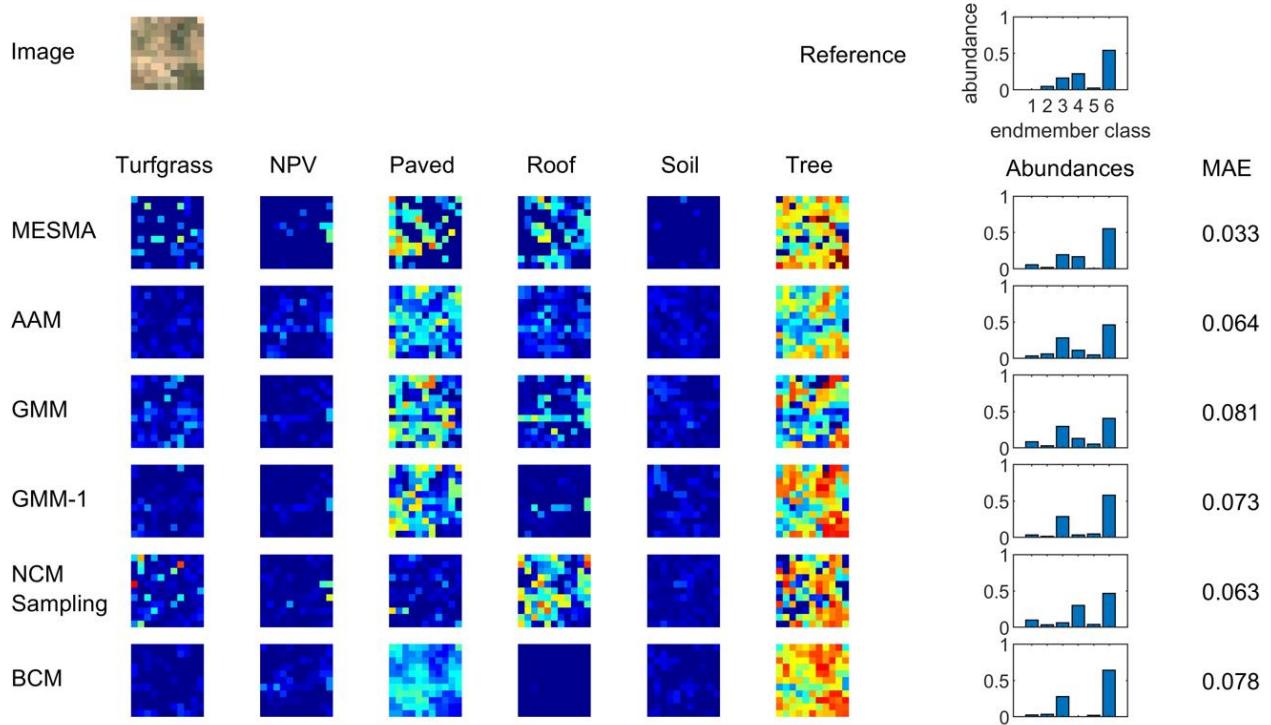


Figure S35: Abundance maps for polygon No. 35.

16 m



4 m

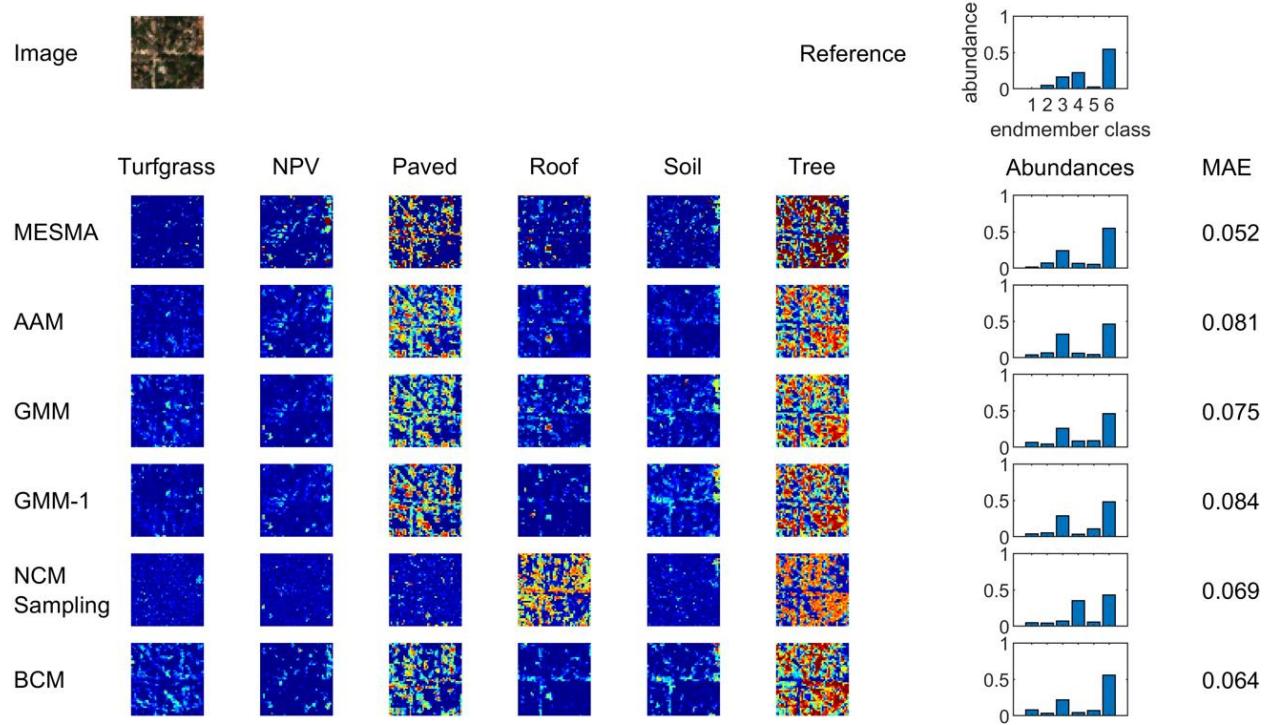
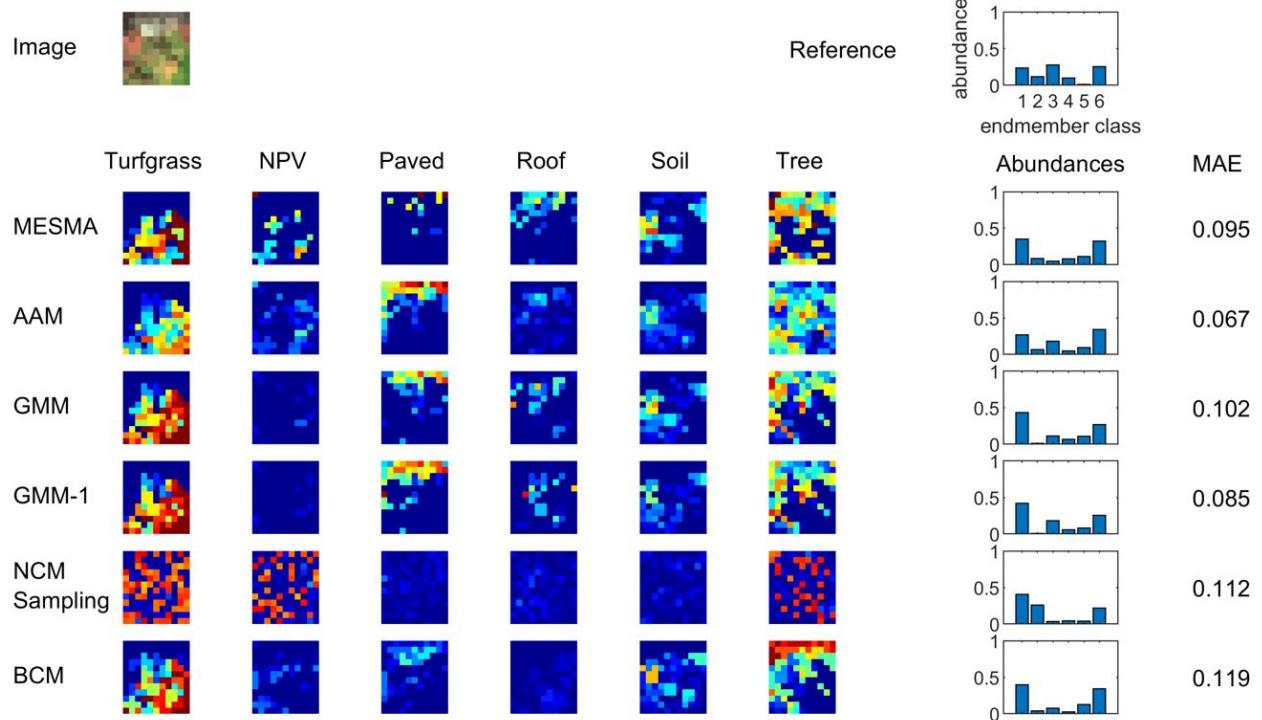


Figure S36: Abundance maps for polygon No. 36.

16 m



4 m

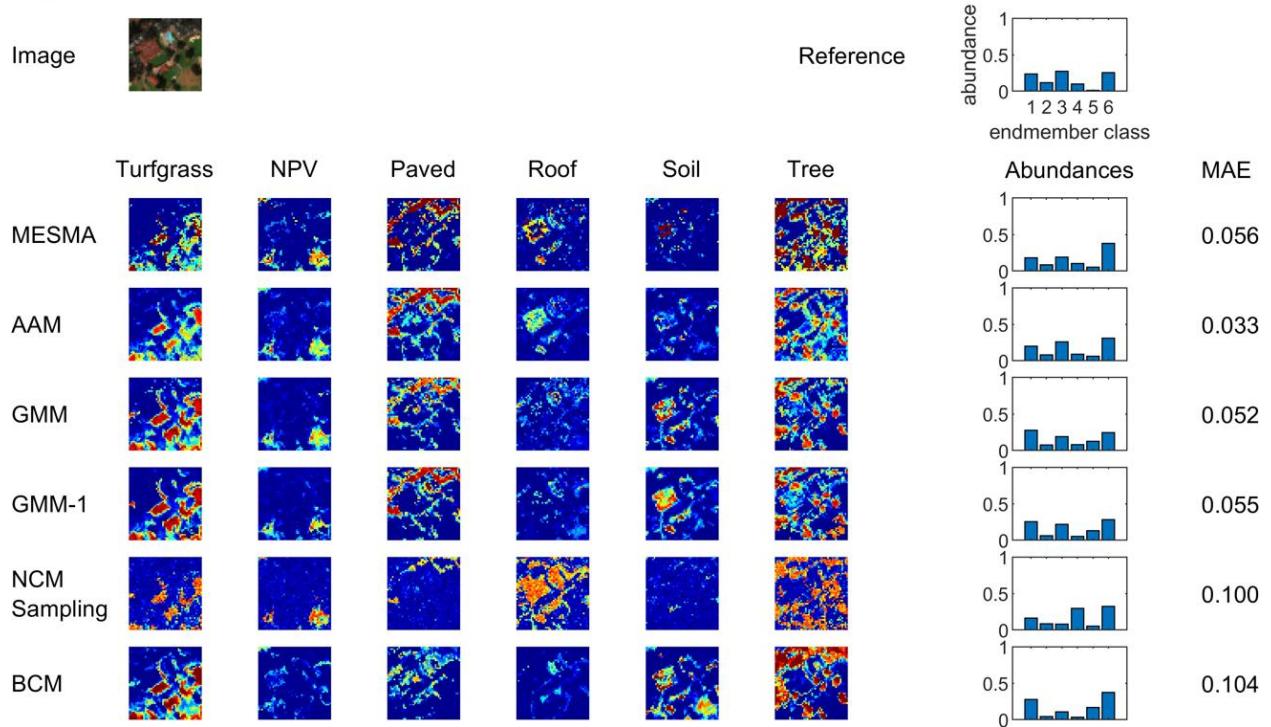
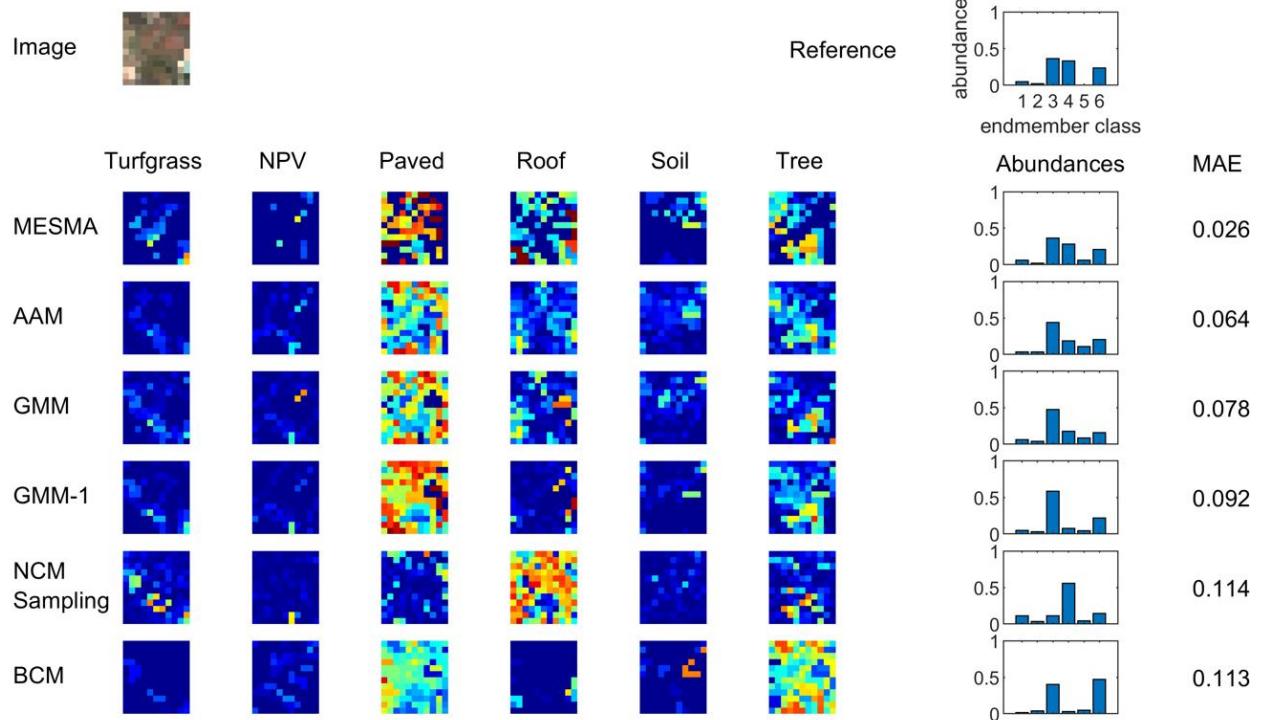


Figure S37: Abundance maps for polygon No. 37.

16 m



4 m

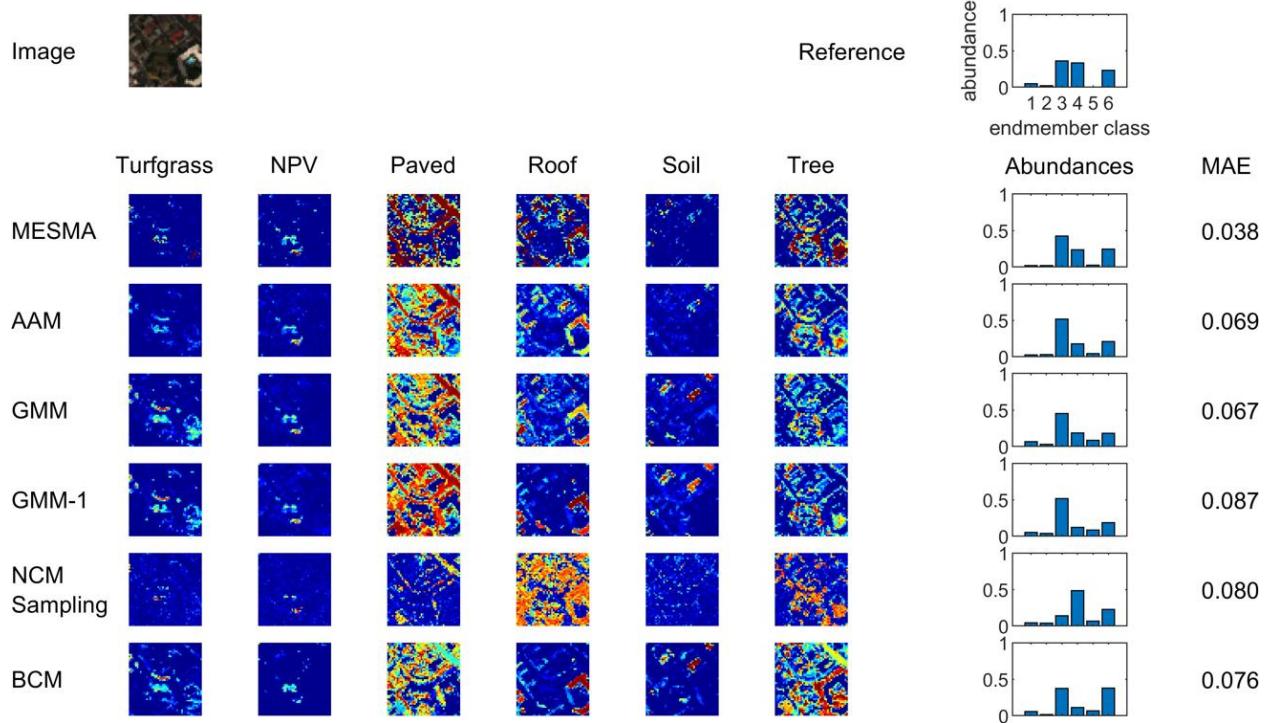
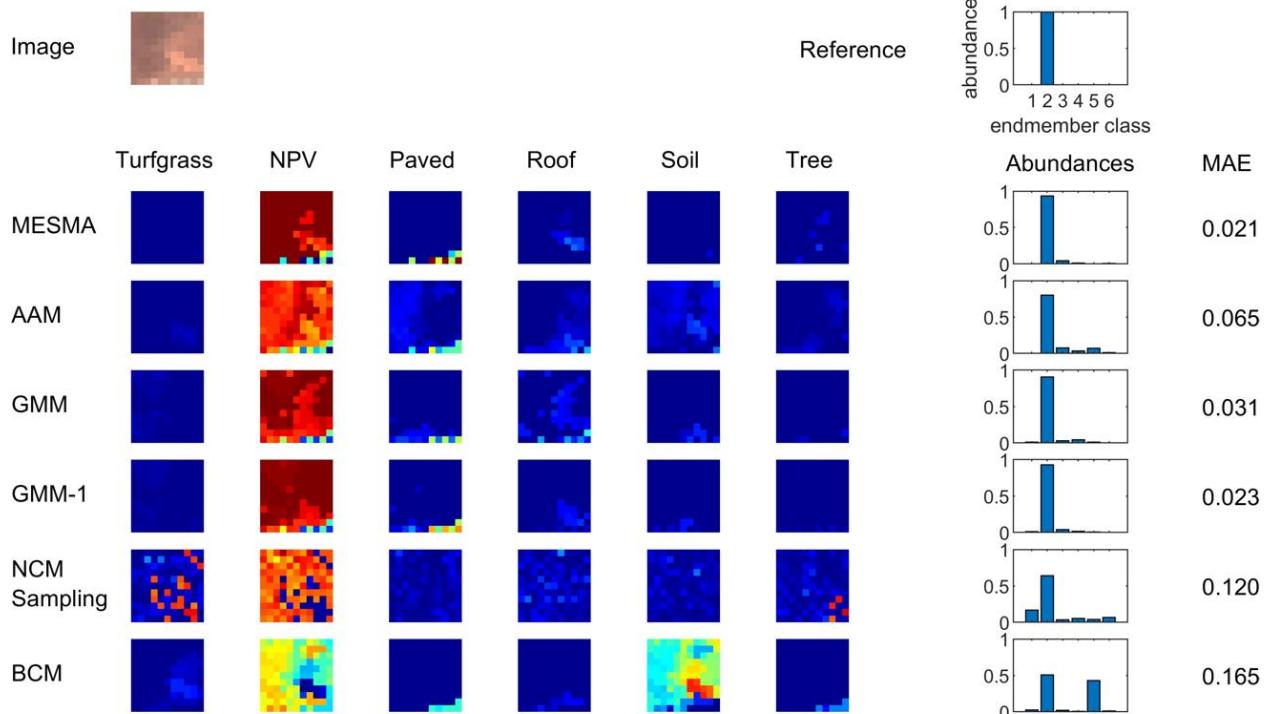


Figure S38: Abundance maps for polygon No. 38.

16 m



4 m

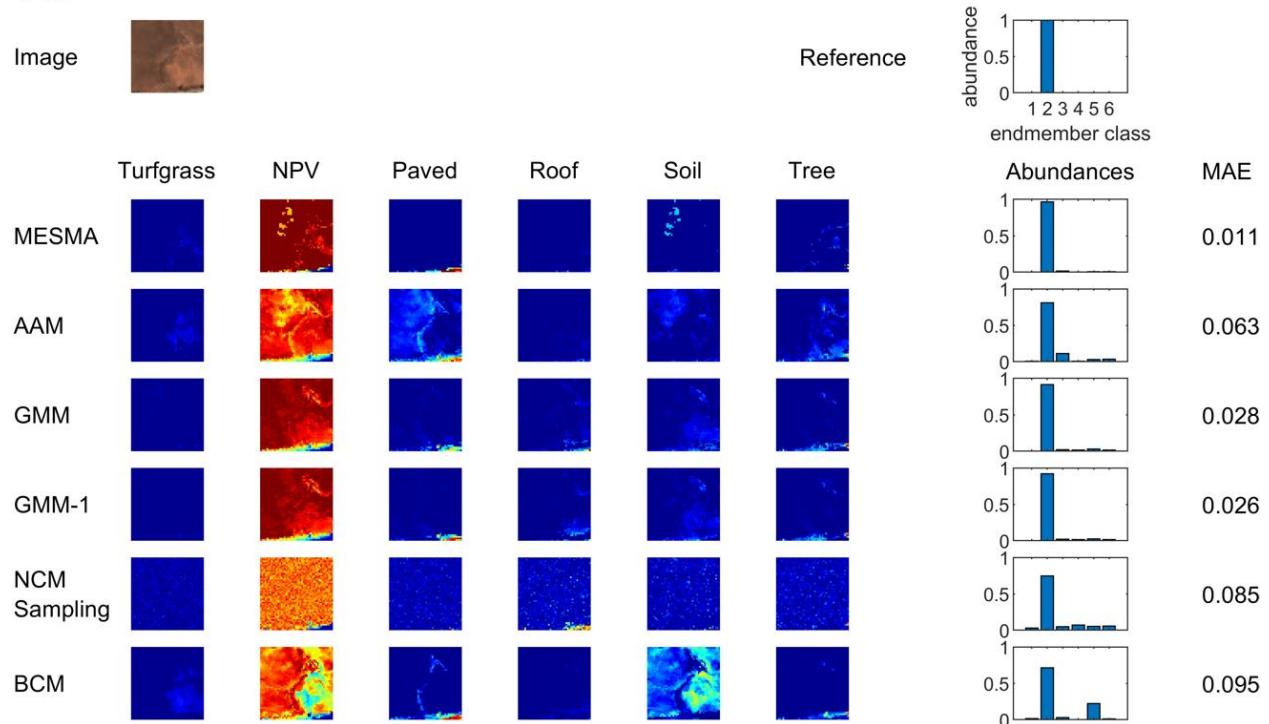
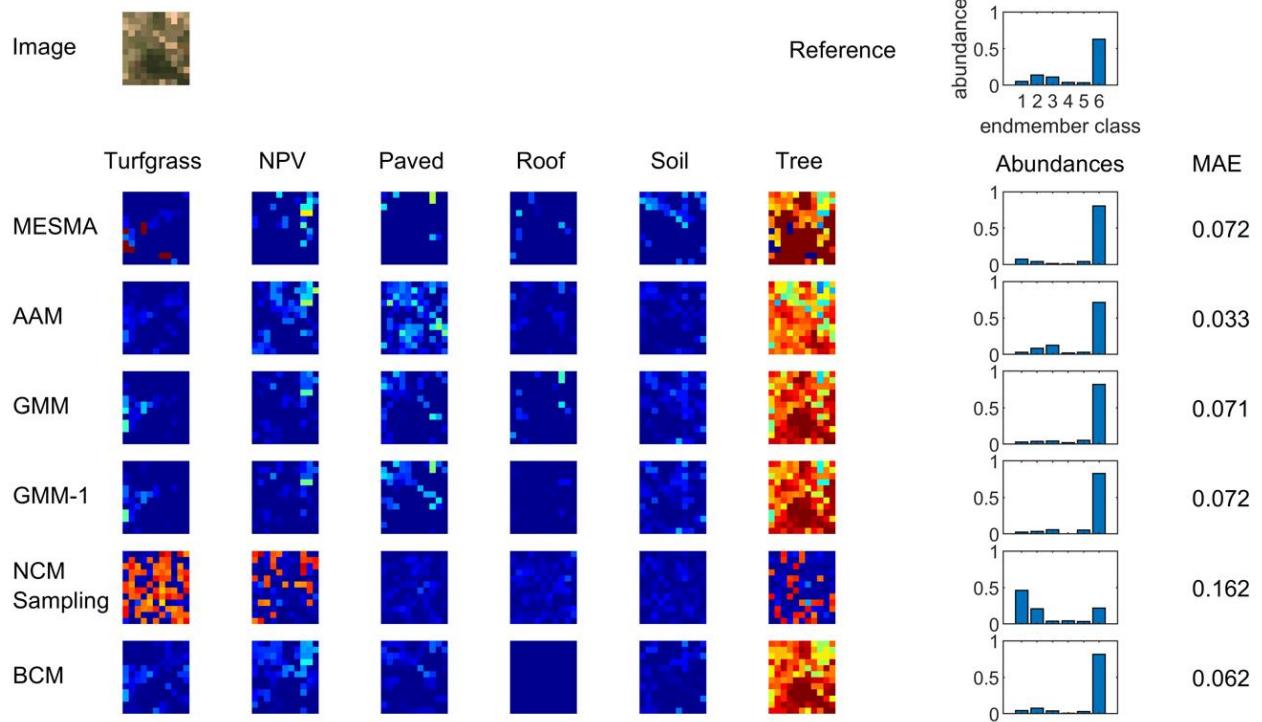


Figure S39: Abundance maps for polygon No. 39.

16 m



4 m

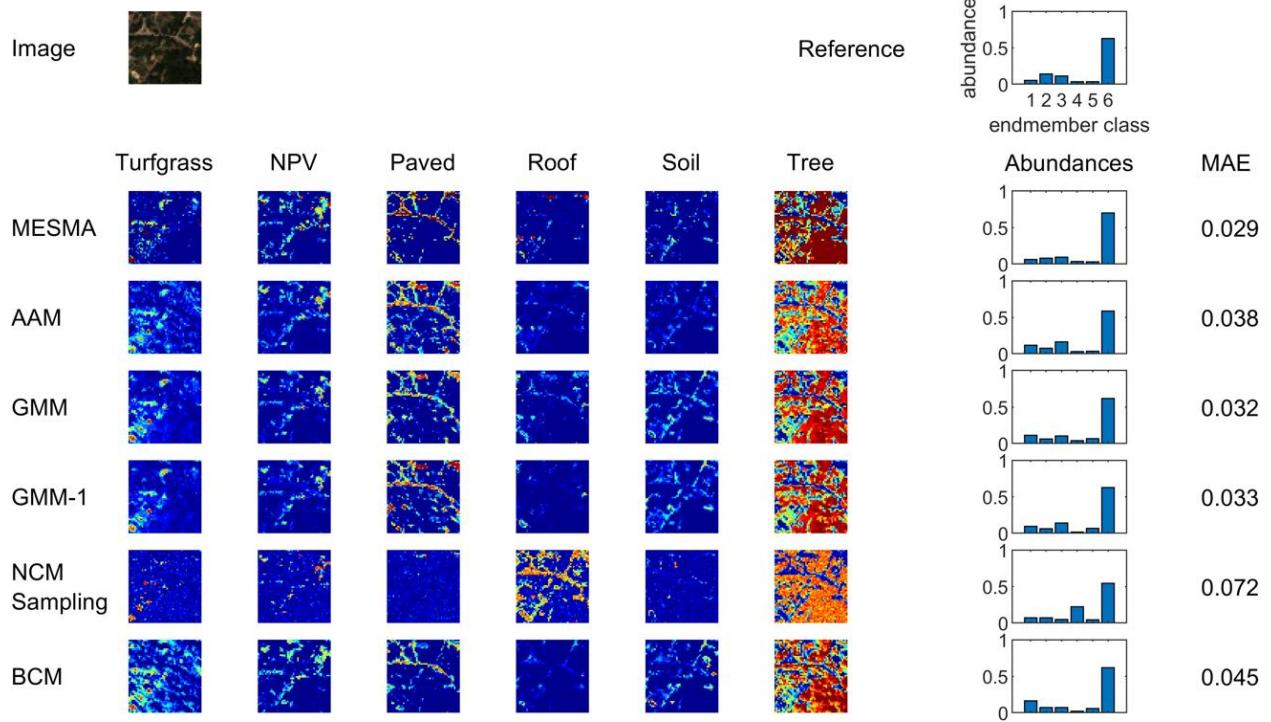
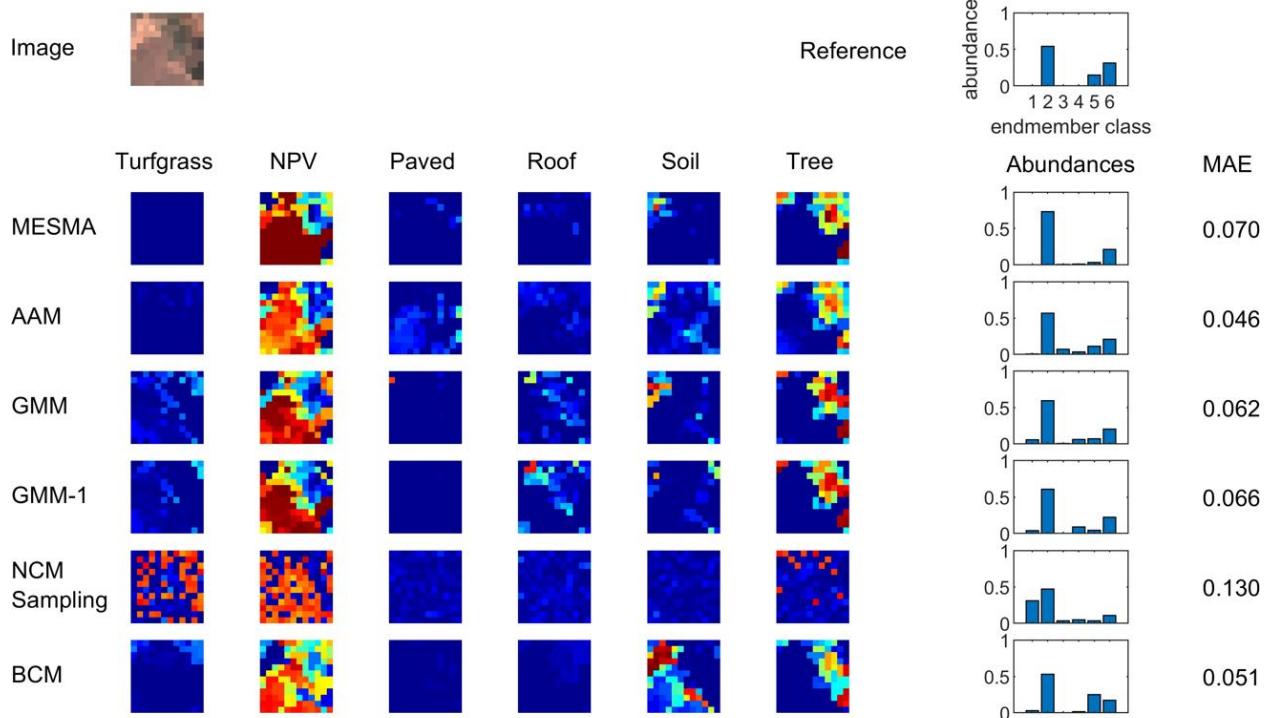


Figure S40: Abundance maps for polygon No. 40.

16 m



4 m

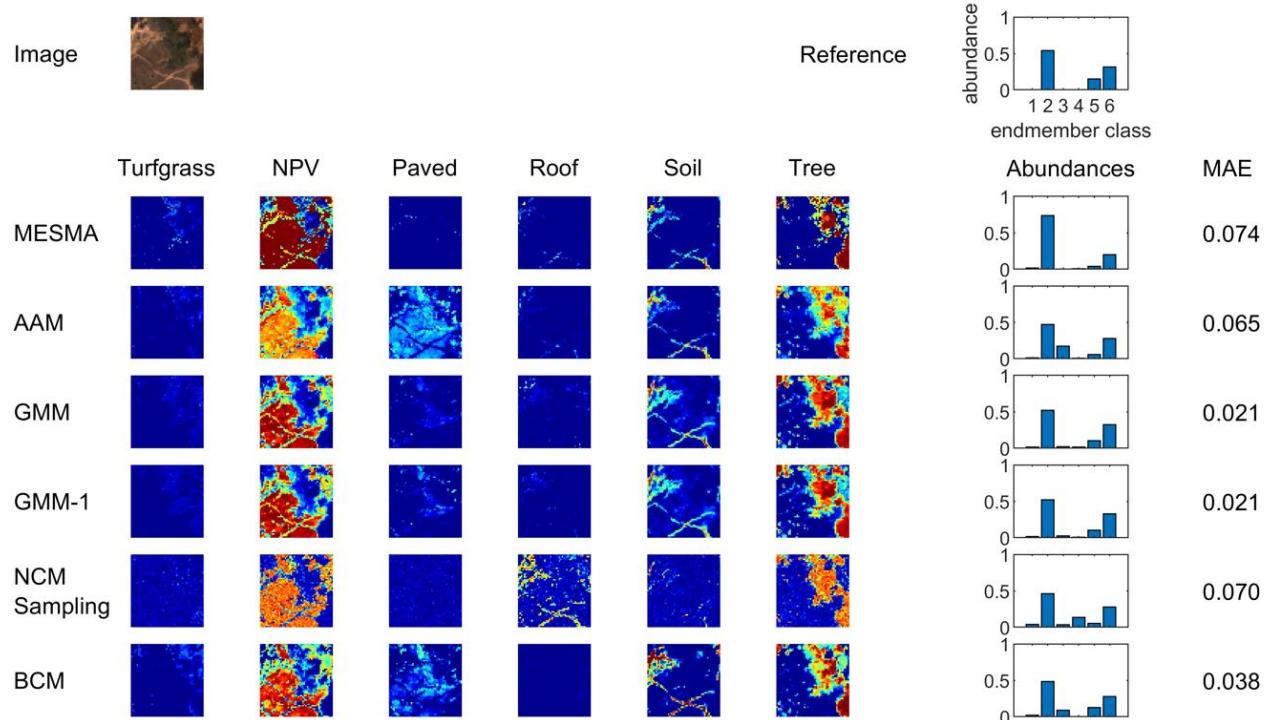
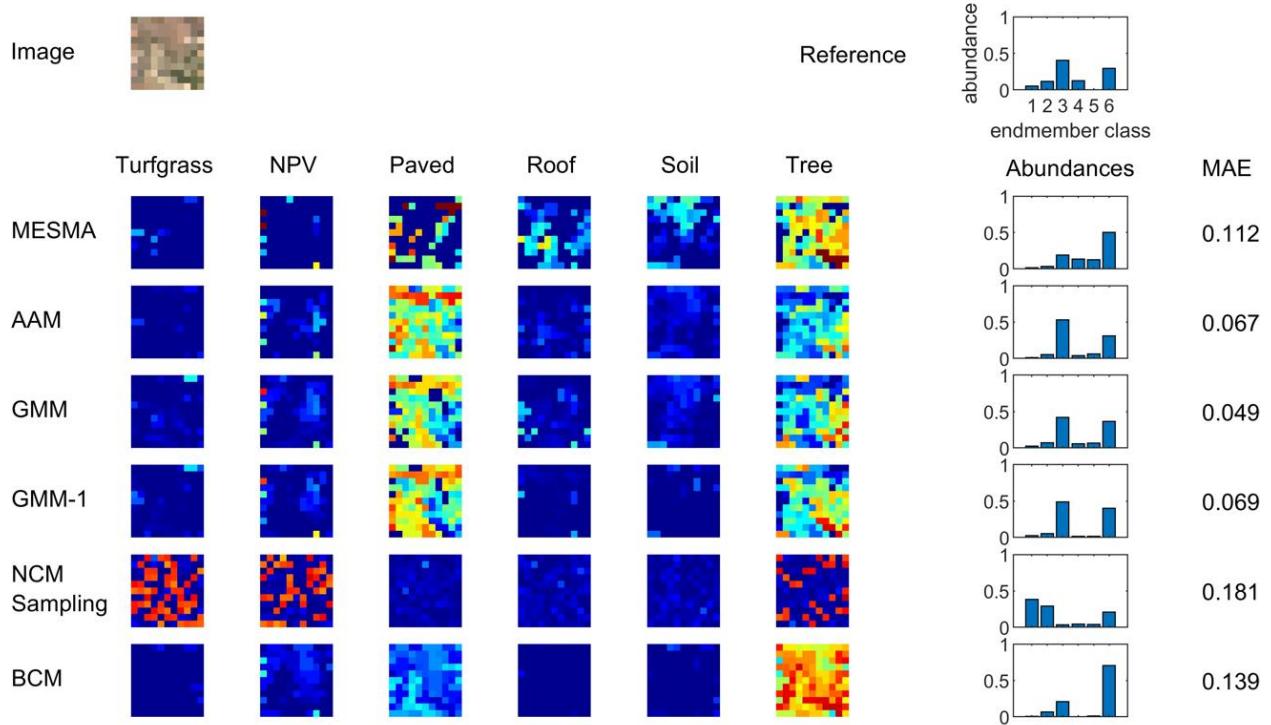


Figure S41: Abundance maps for polygon No. 41.

16 m



4 m

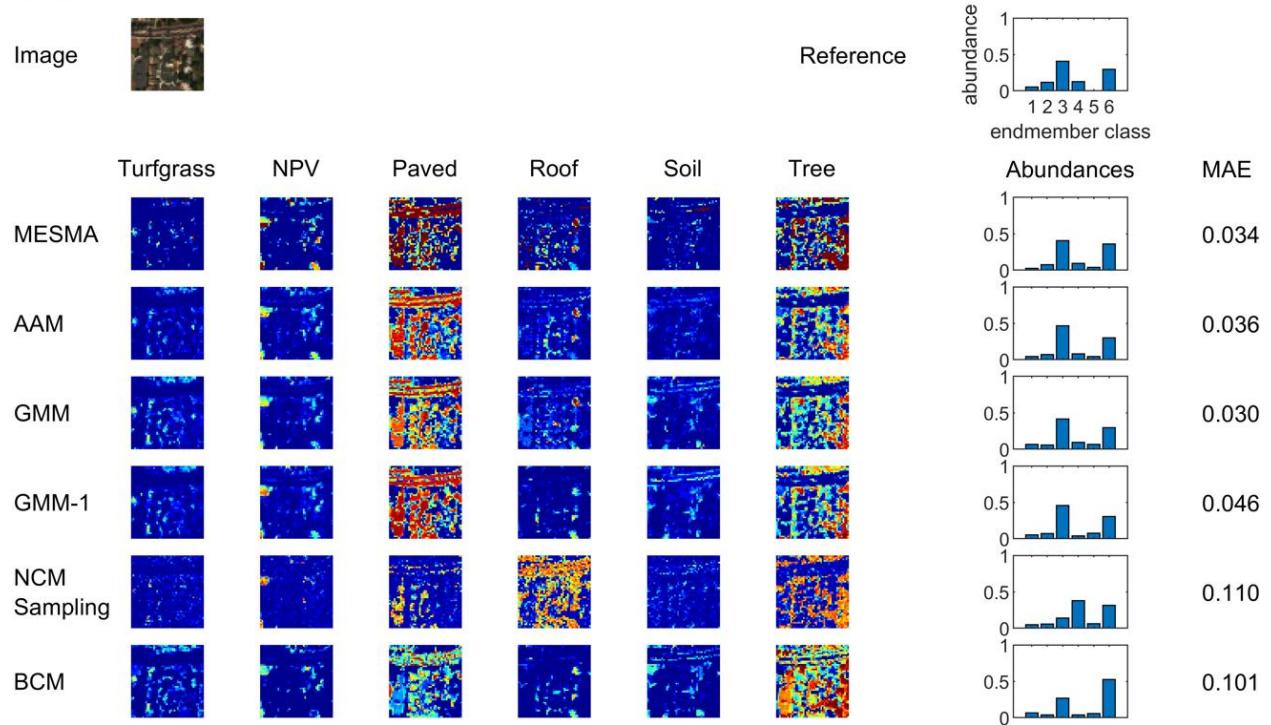
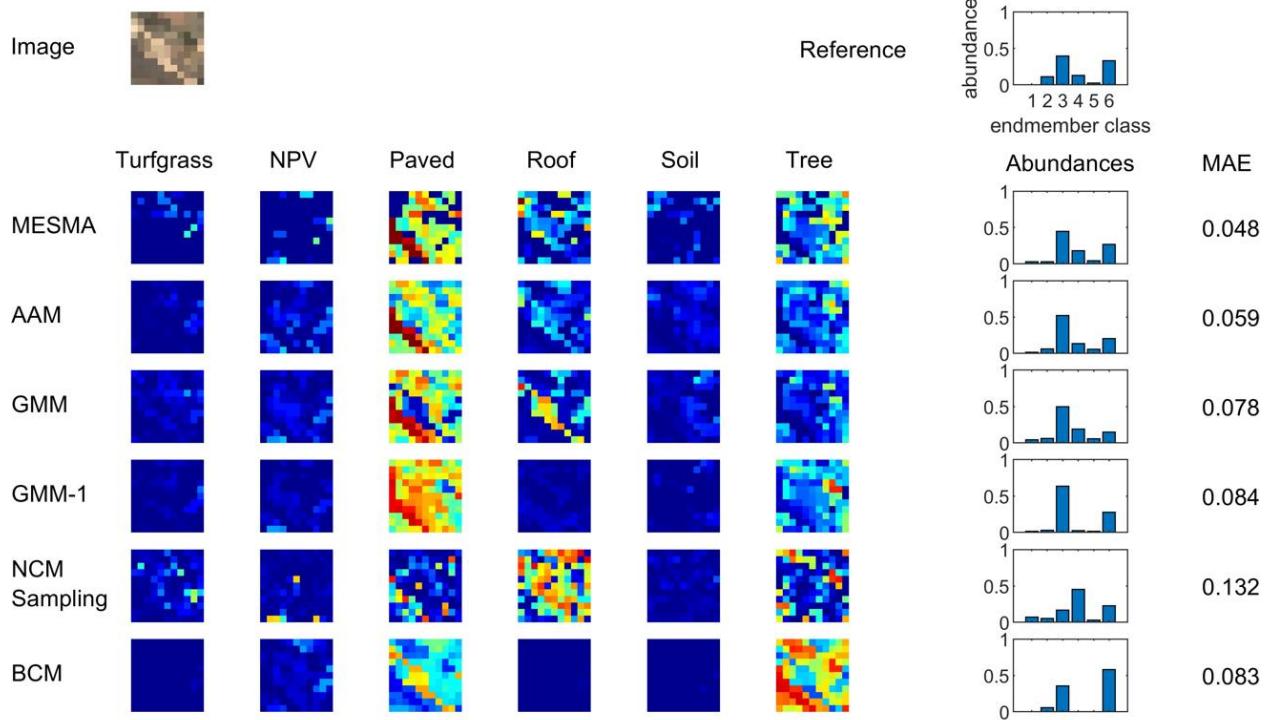


Figure S42: Abundance maps for polygon No. 42.

16 m



4 m

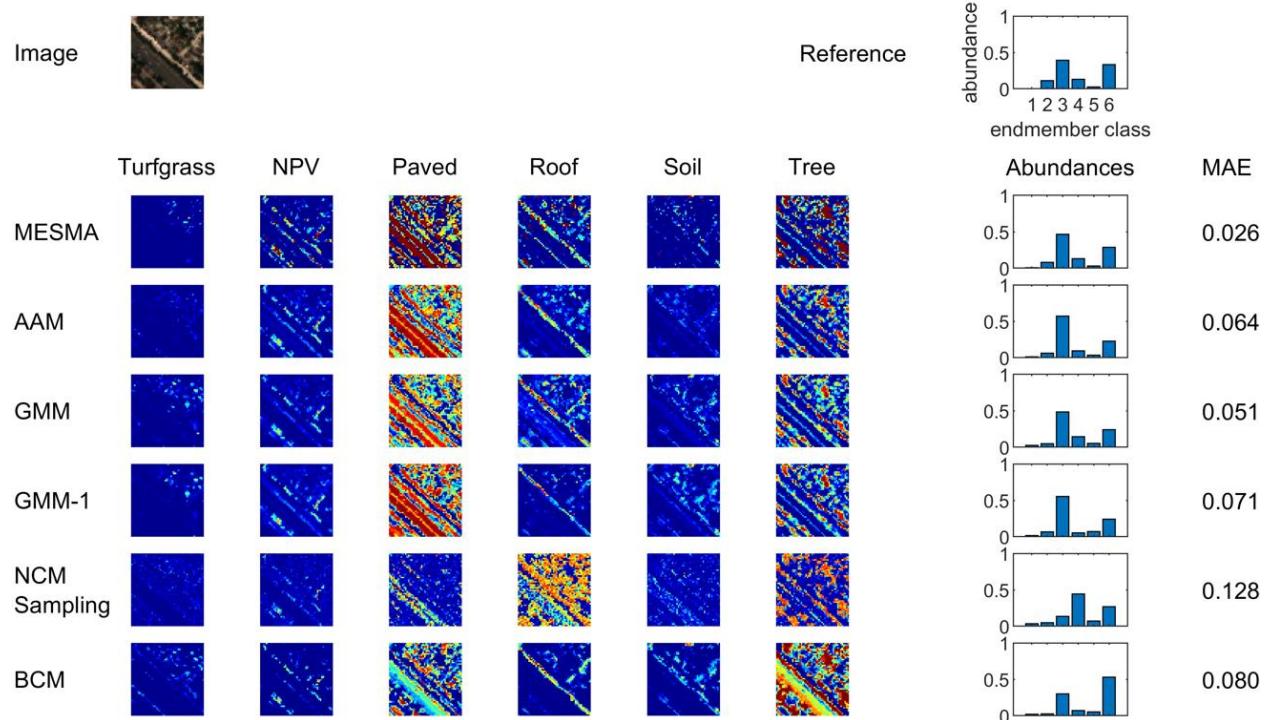
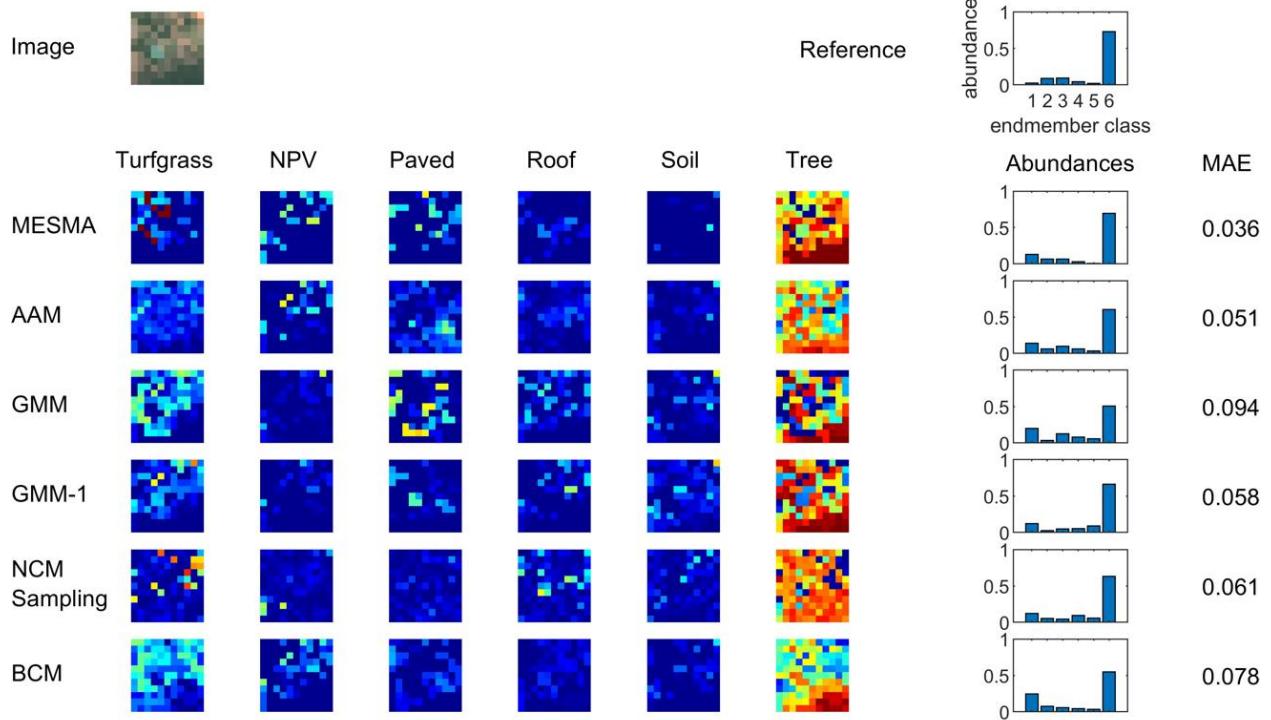


Figure S43: Abundance maps for polygon No. 43.

16 m



4 m

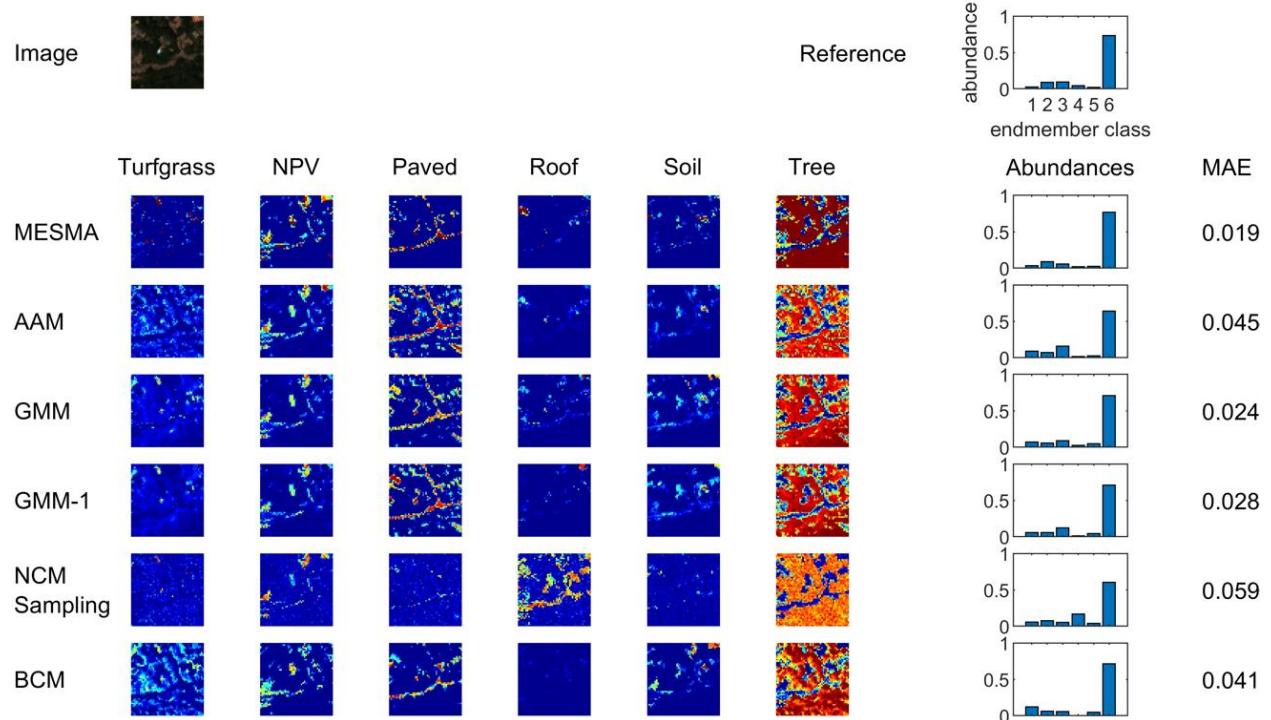
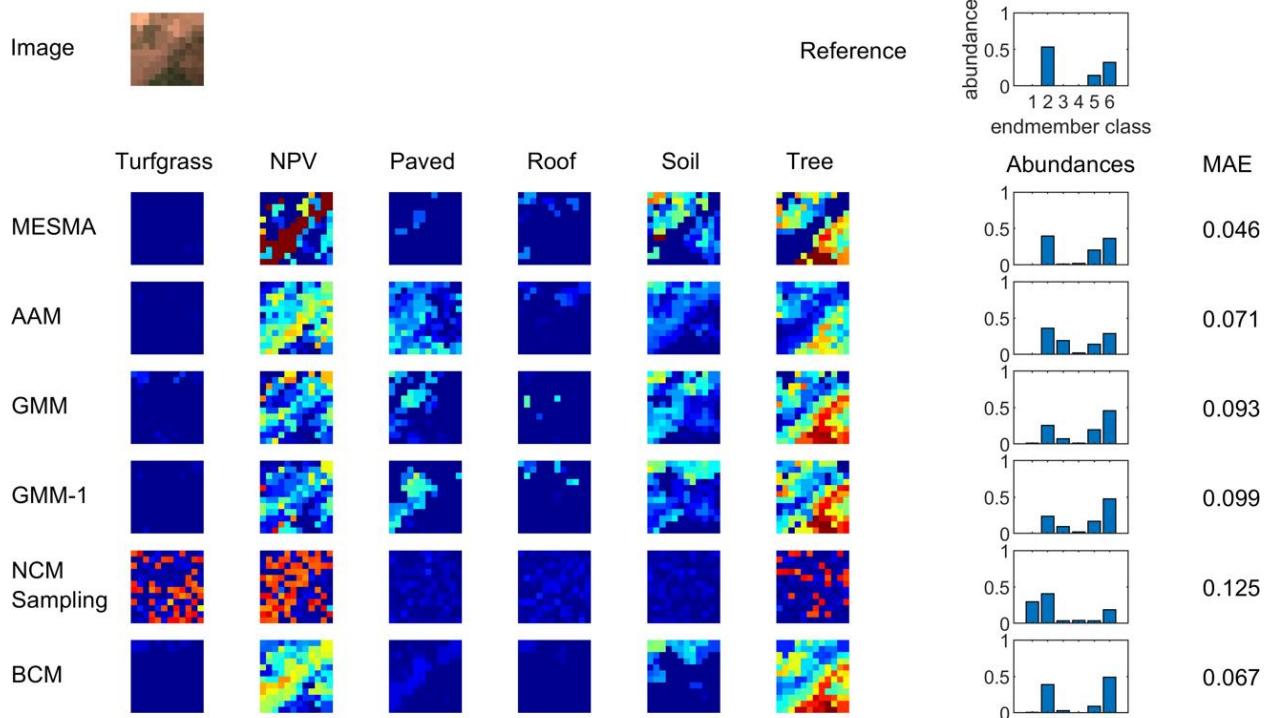


Figure S44: Abundance maps for polygon No. 44.

16 m



4 m

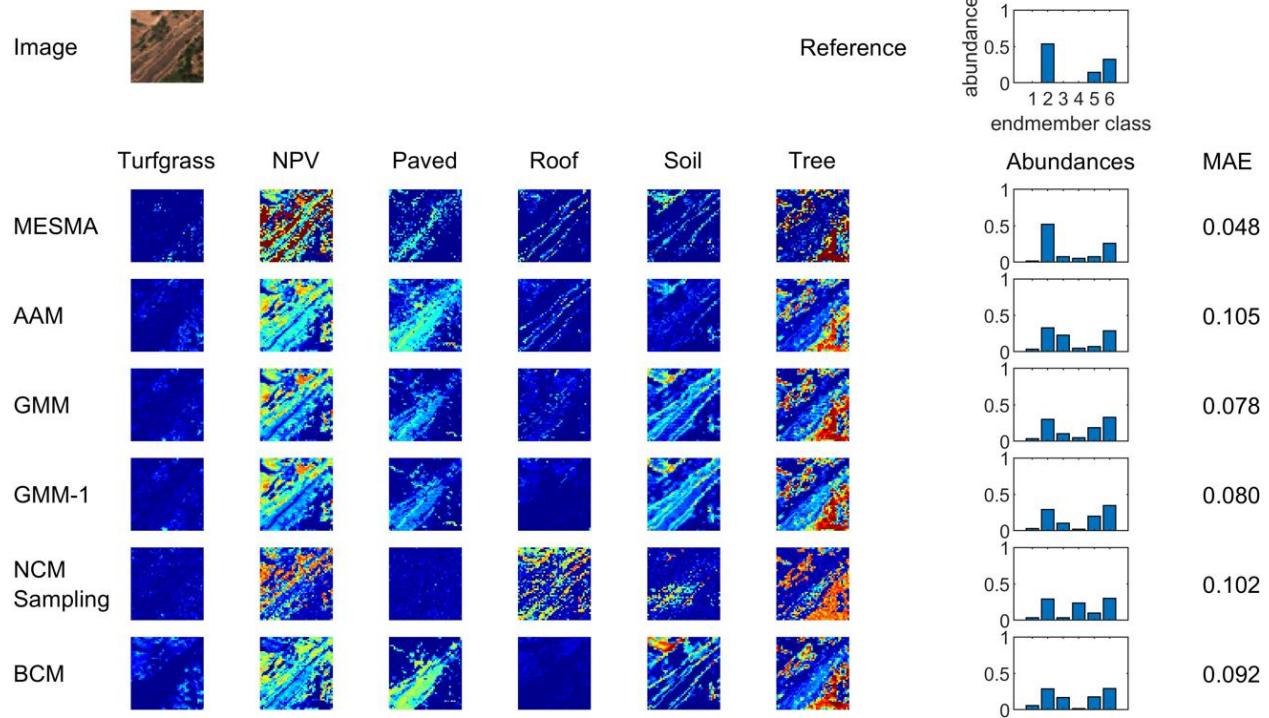
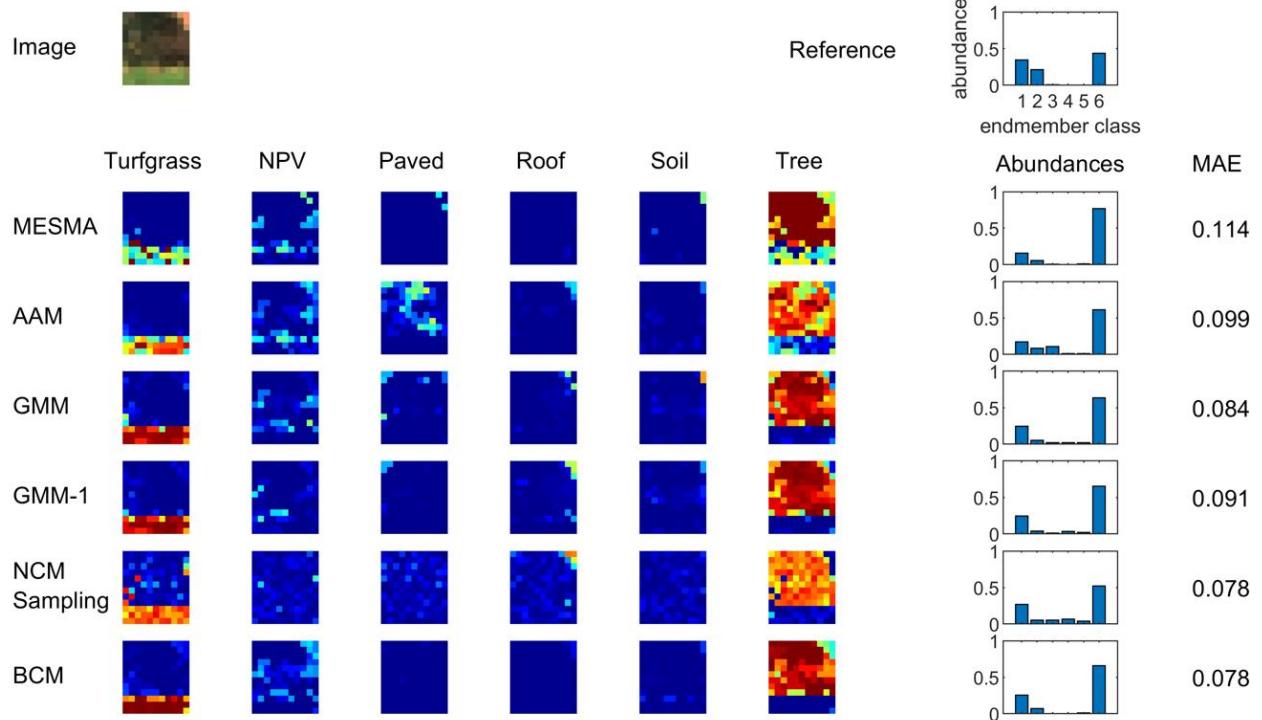


Figure S45: Abundance maps for polygon No. 45.

16 m



4 m

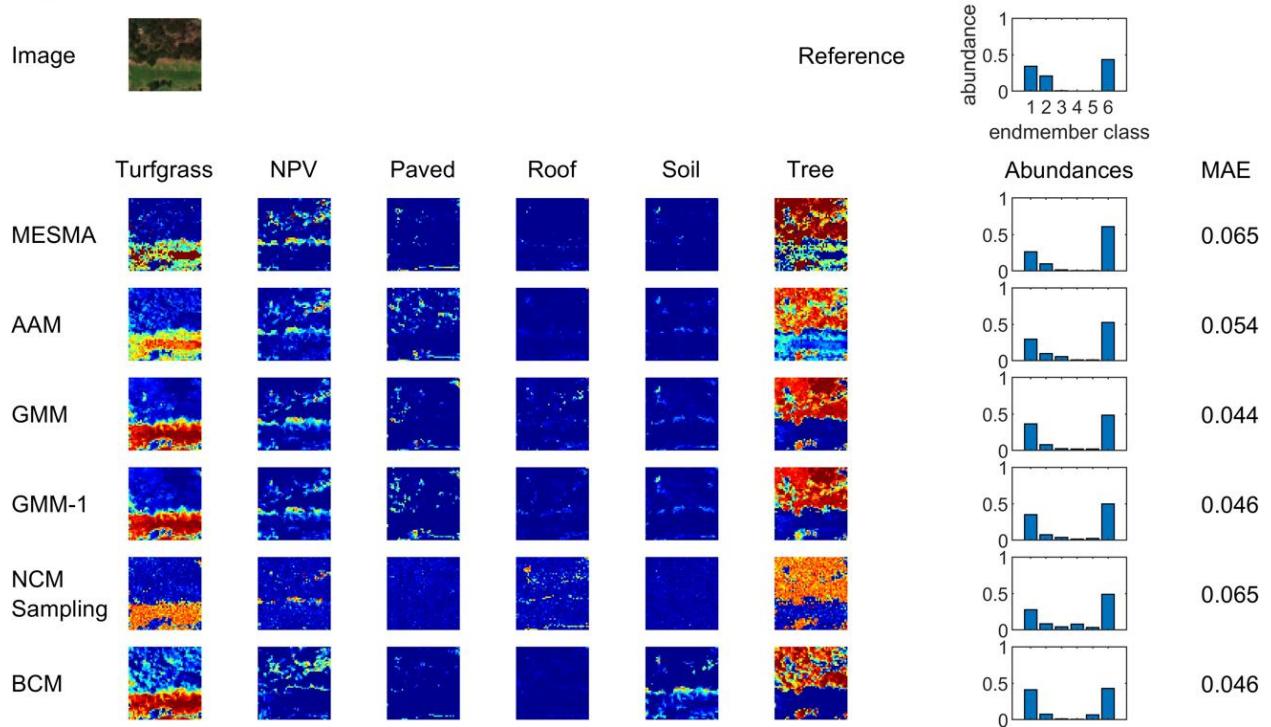
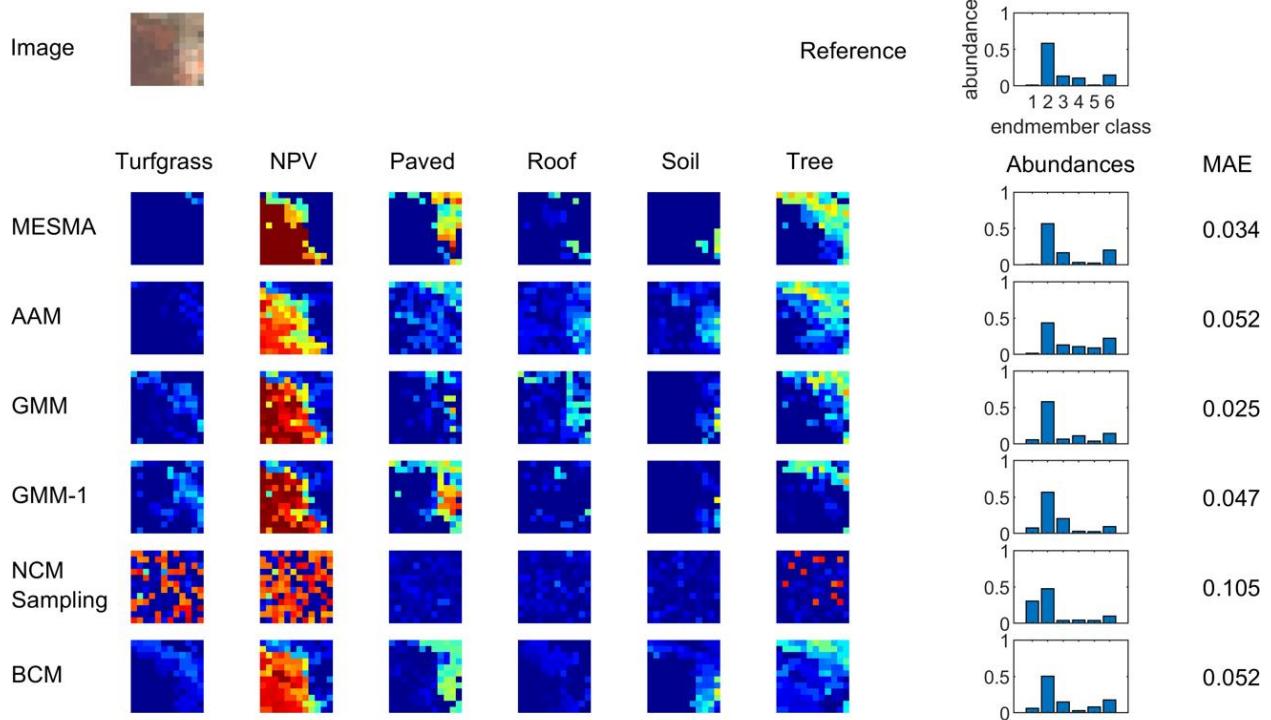


Figure S46: Abundance maps for polygon No. 46.

16 m



4 m

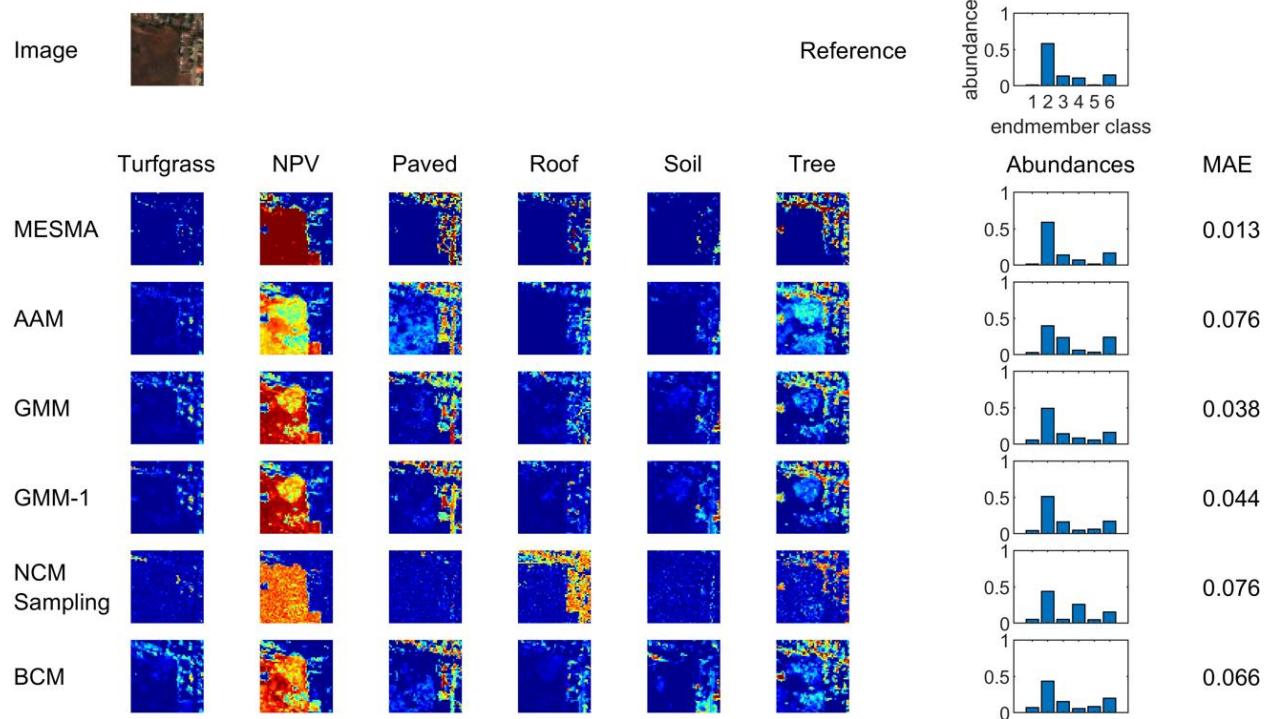
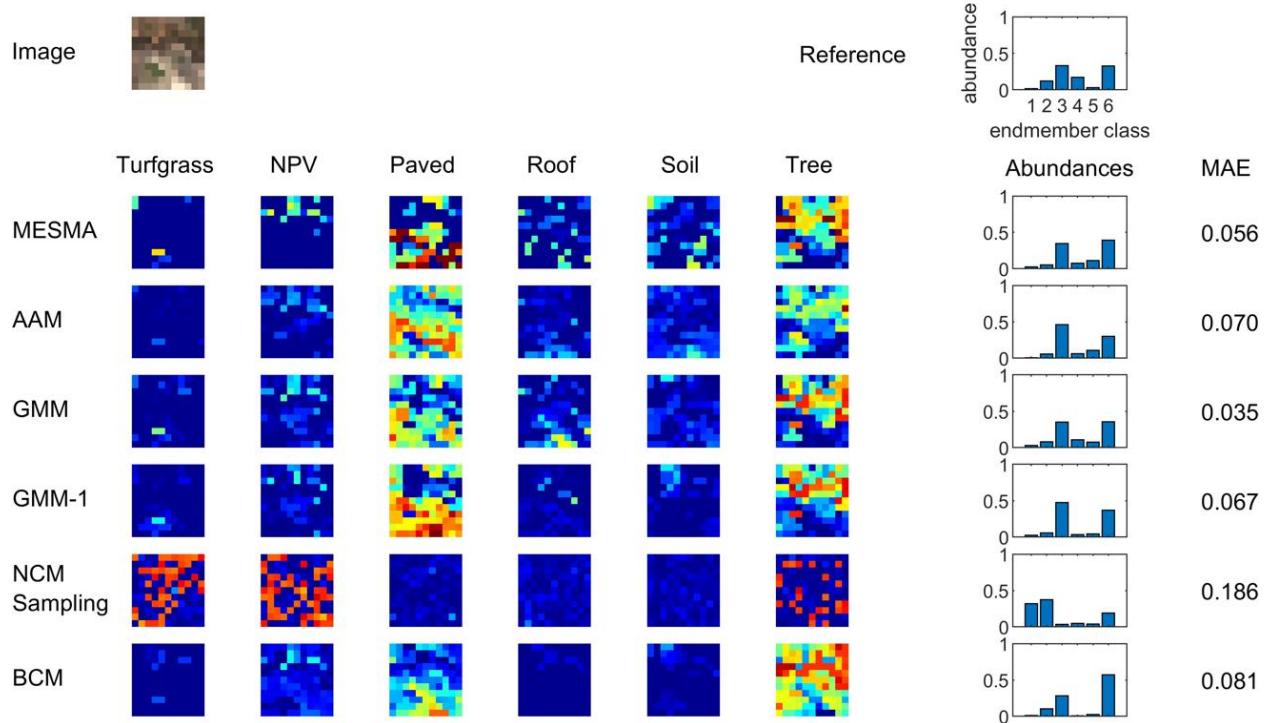


Figure S47: Abundance maps for polygon No. 47.

16 m



4 m

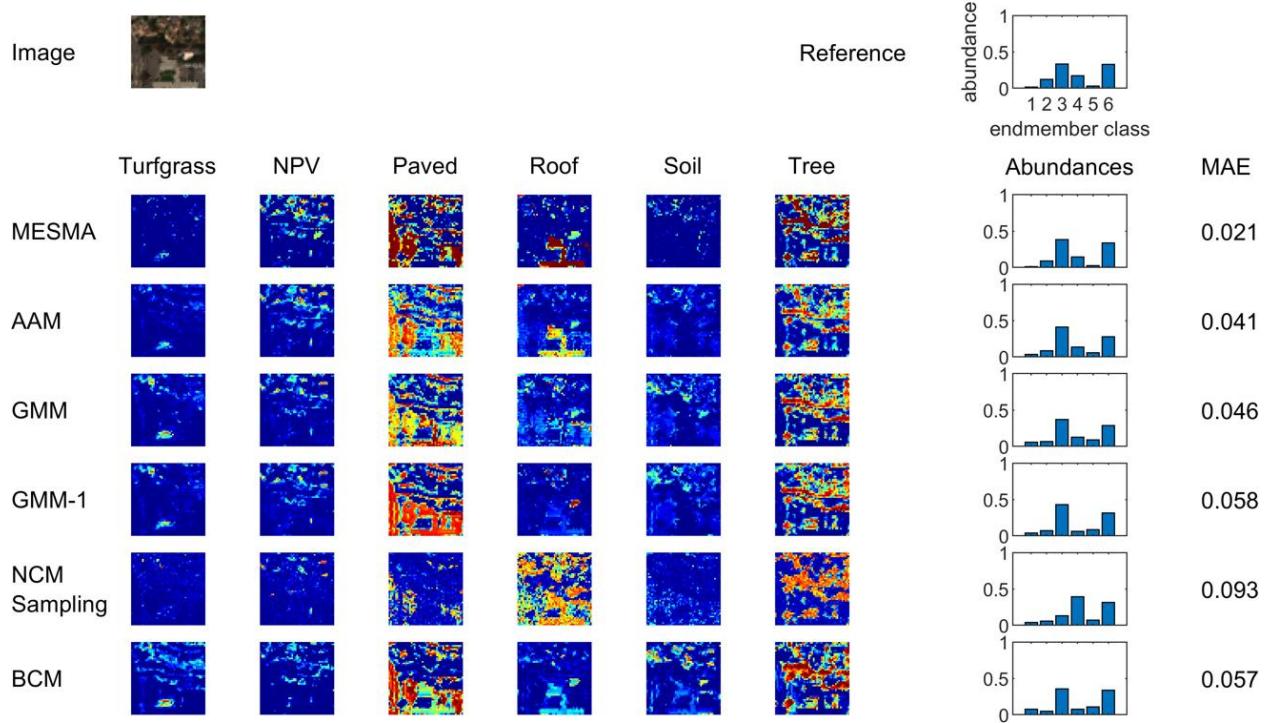
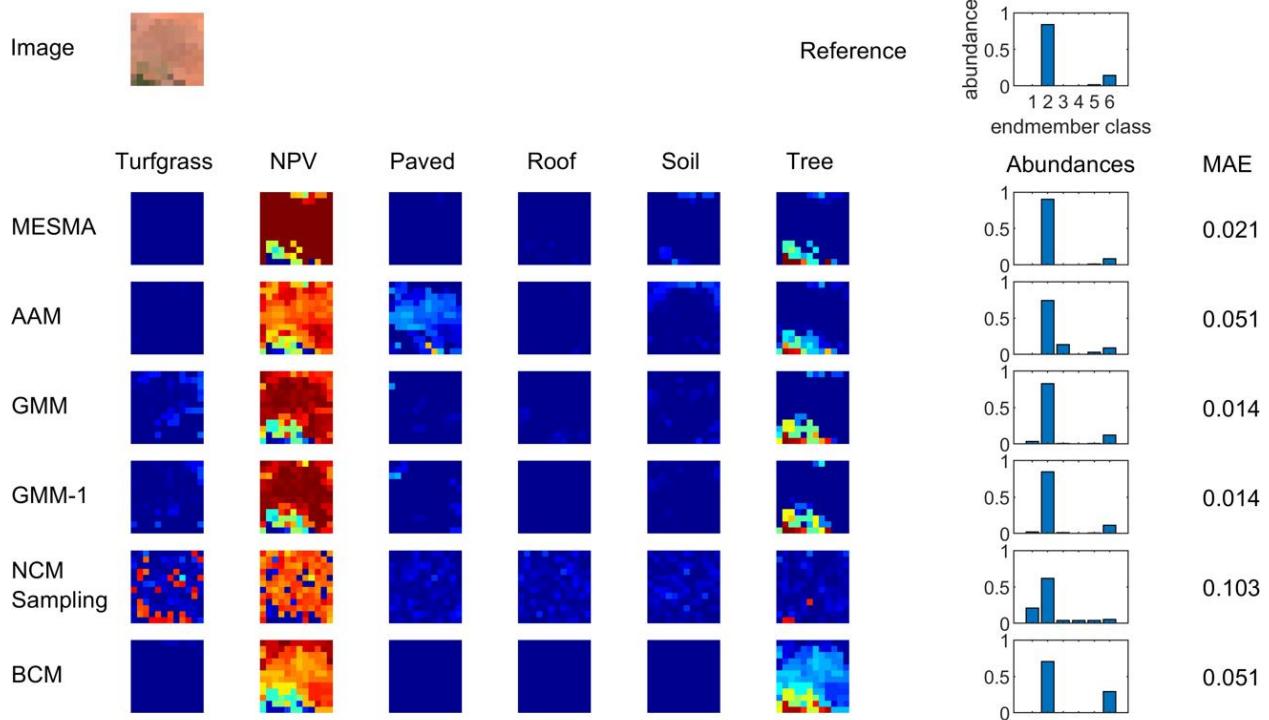


Figure S48: Abundance maps for polygon No. 48.

16 m



4 m

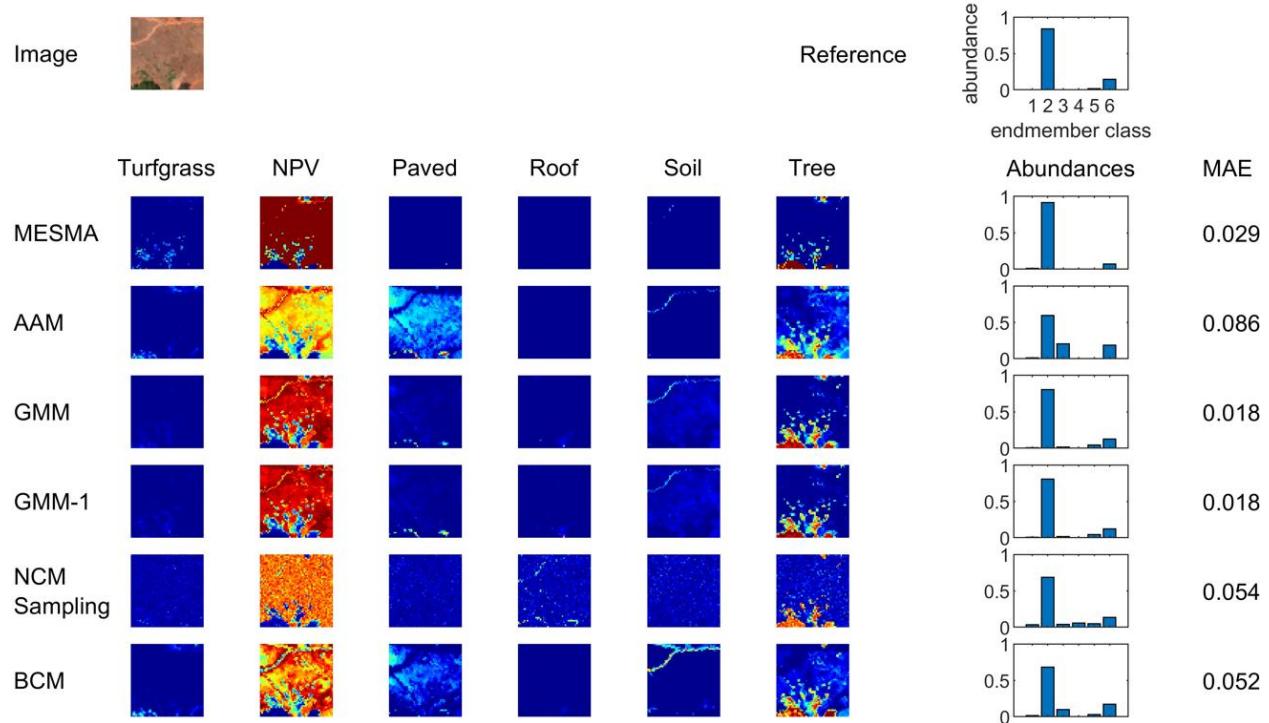
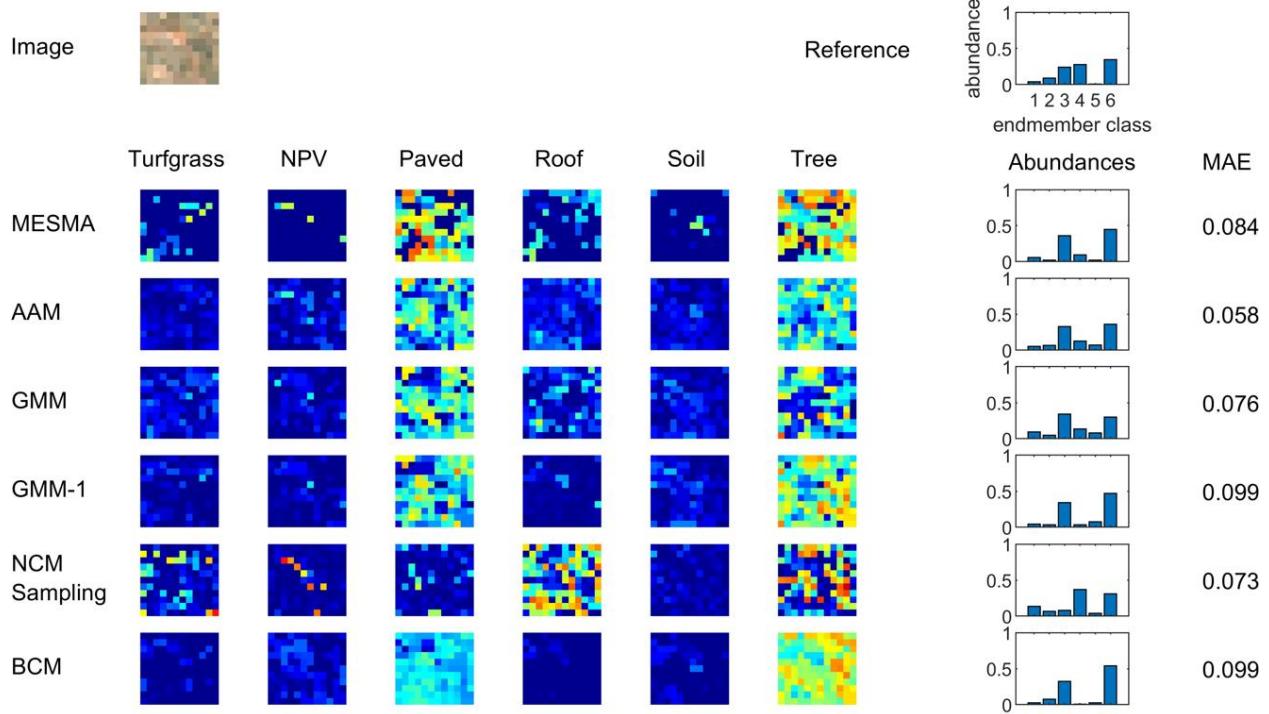


Figure S49: Abundance maps for polygon No. 49.

16 m



4 m

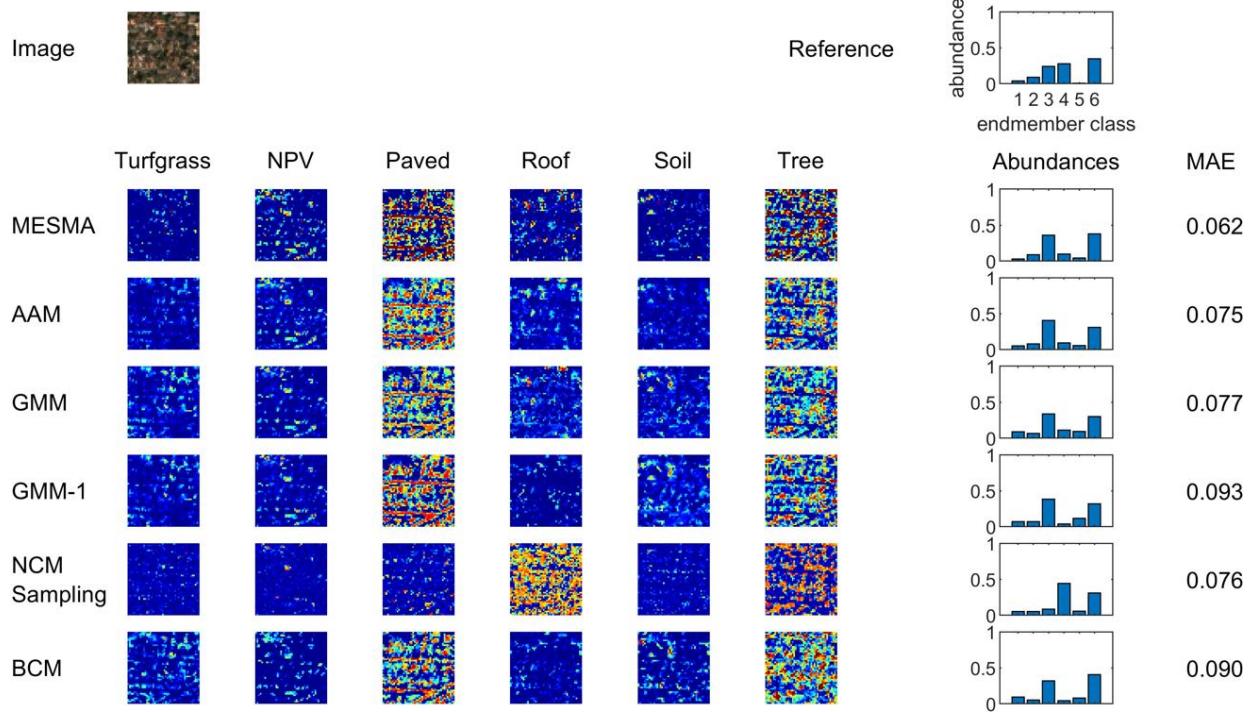
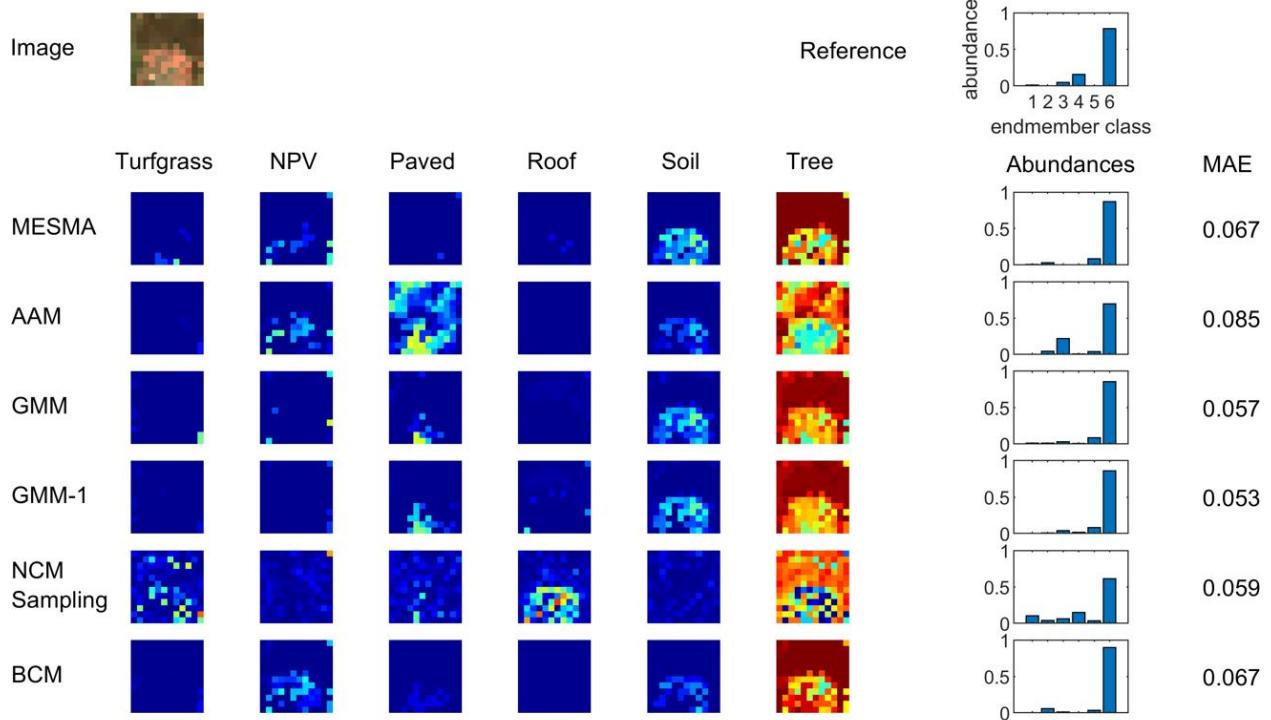


Figure S50: Abundance maps for polygon No. 50.

16 m



4 m

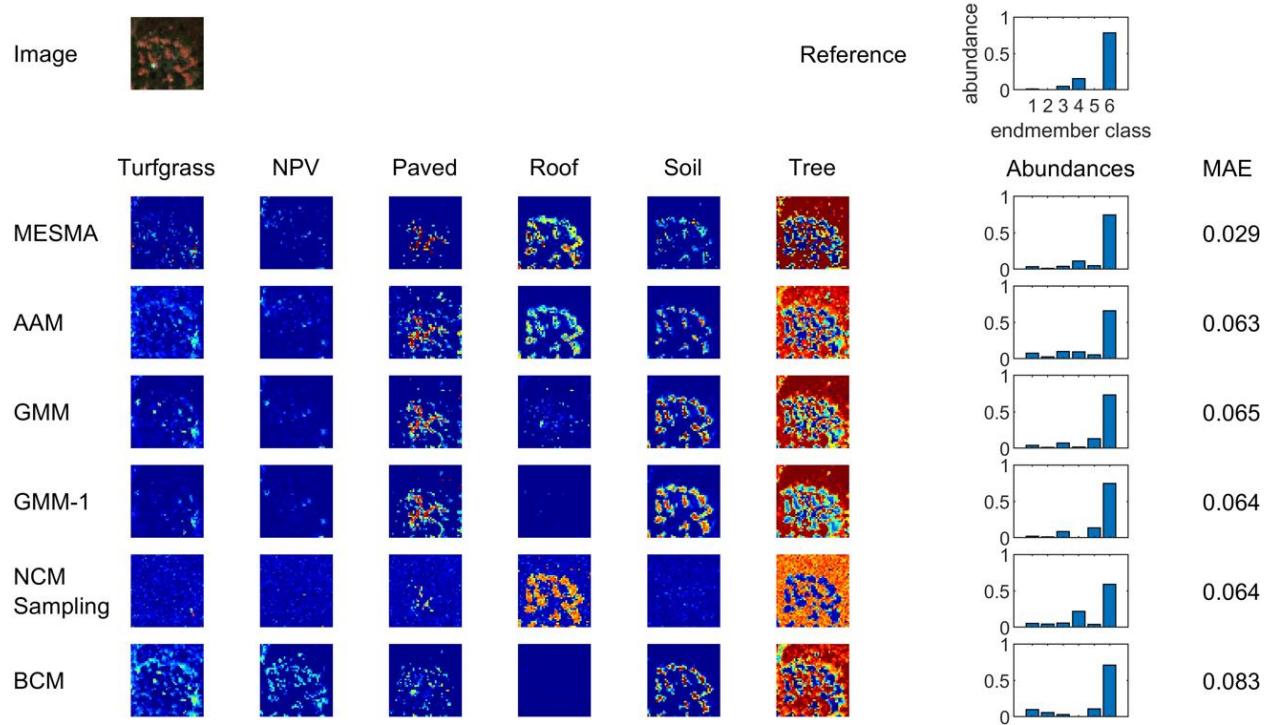
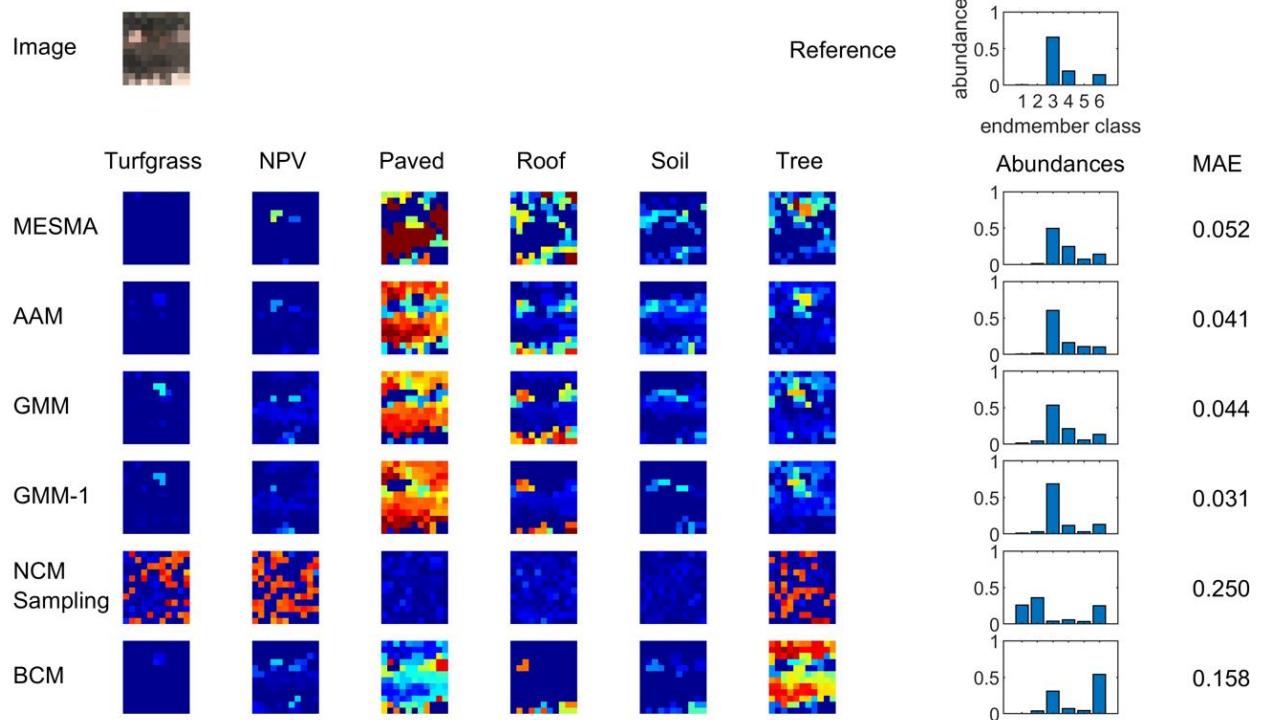


Figure S51: Abundance maps for polygon No. 51.

16 m



4 m

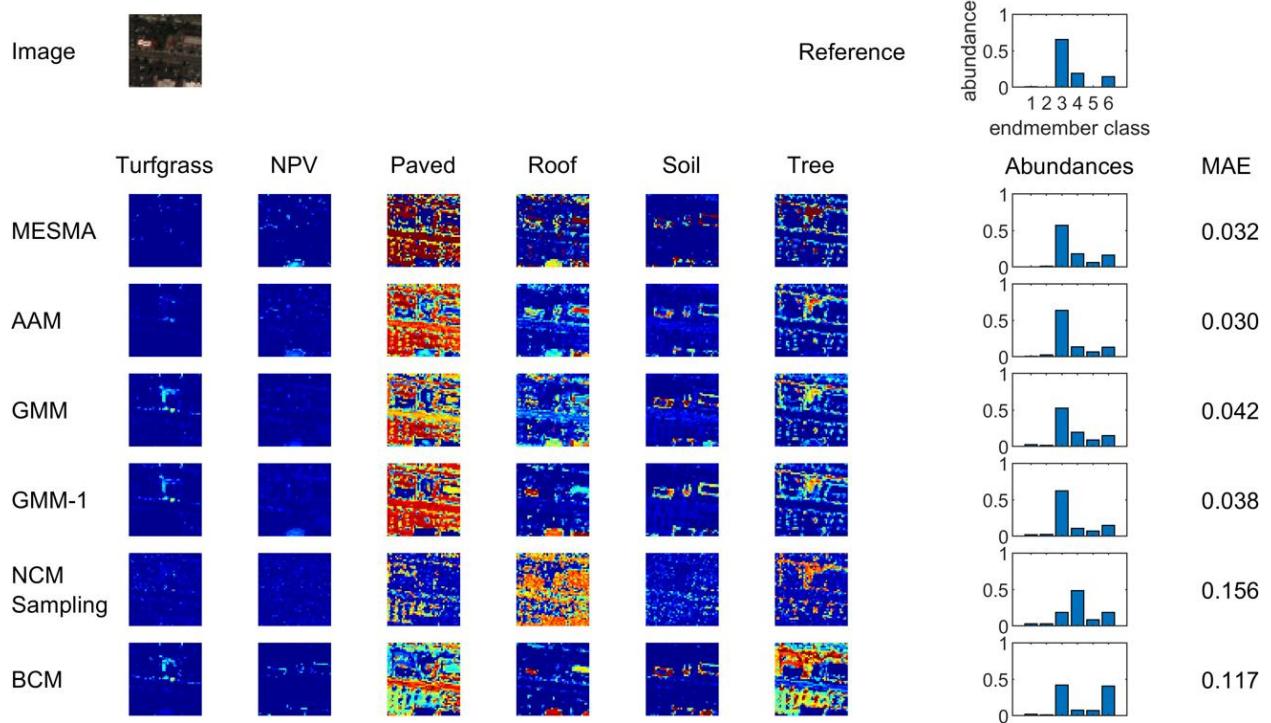
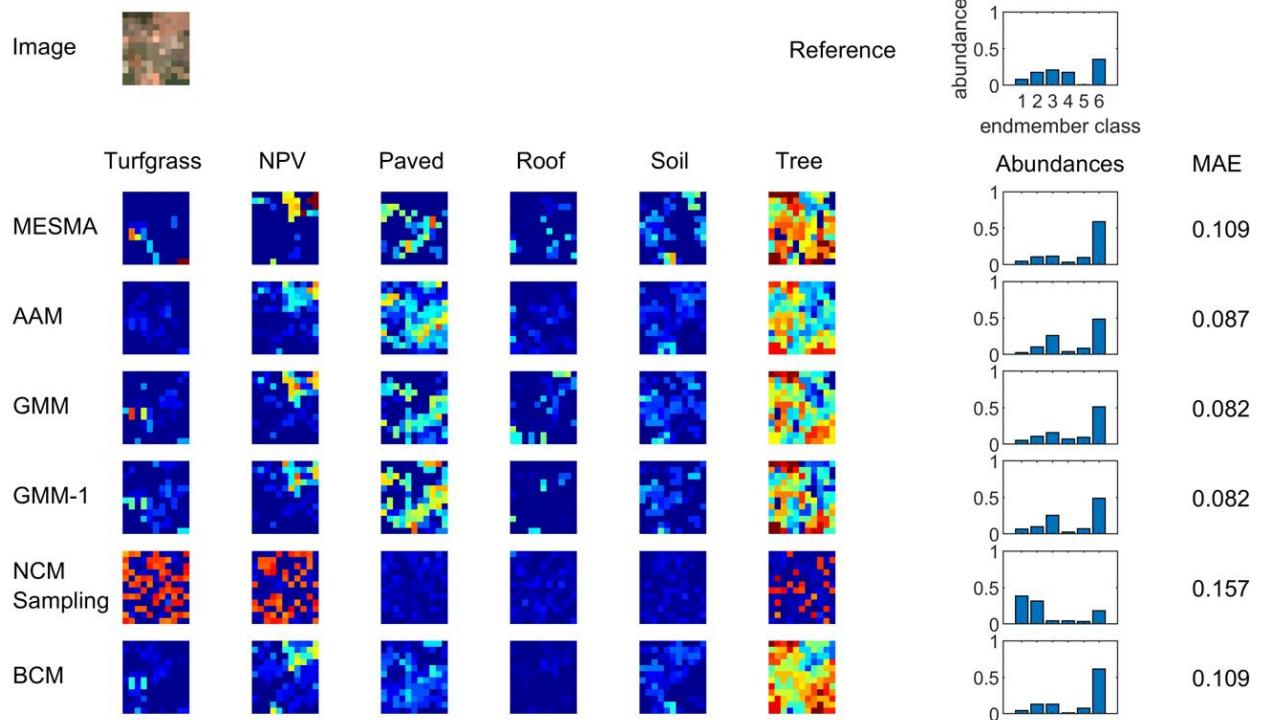


Figure S52: Abundance maps for polygon No. 52.

16 m



4 m

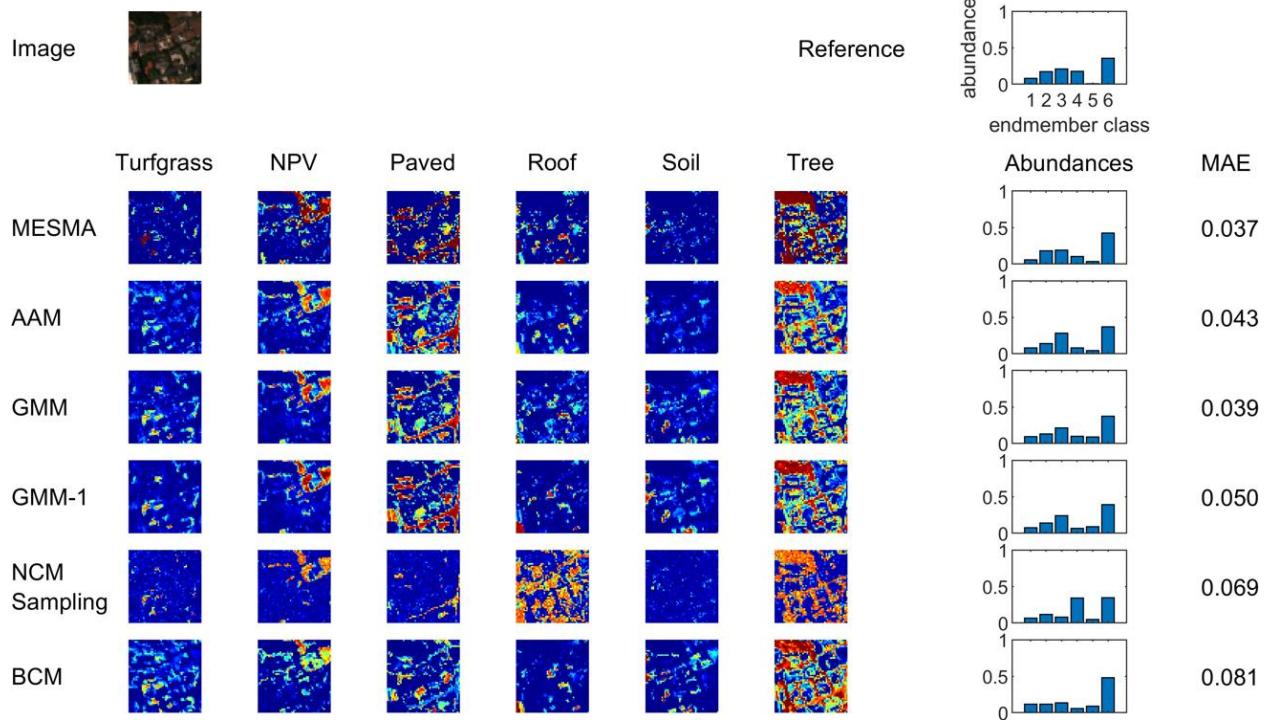
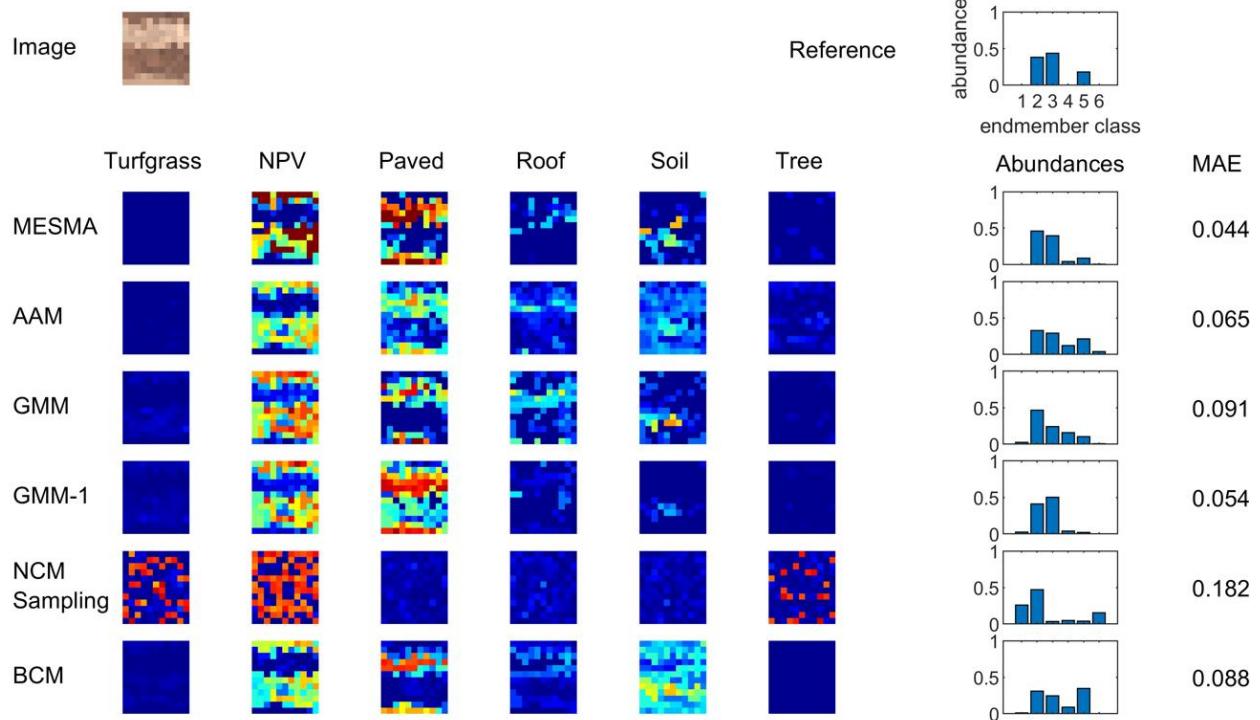


Figure S53: Abundance maps for polygon No. 53.

16 m



4 m

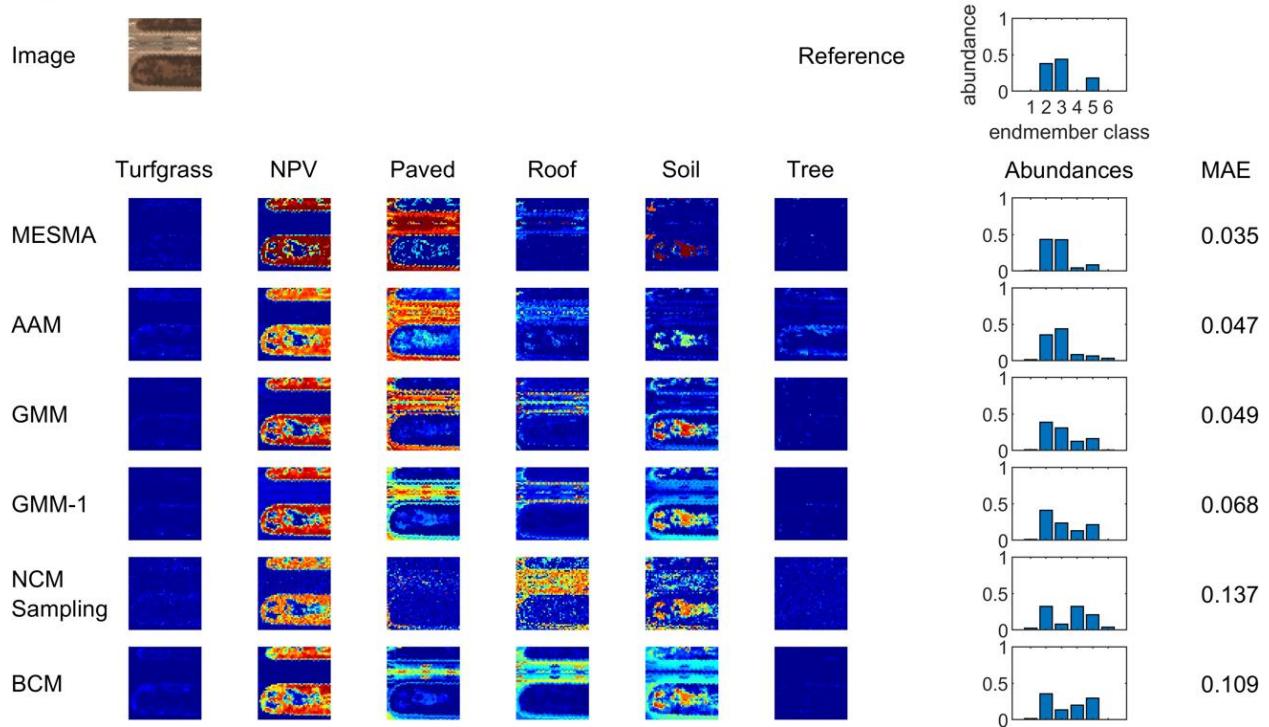
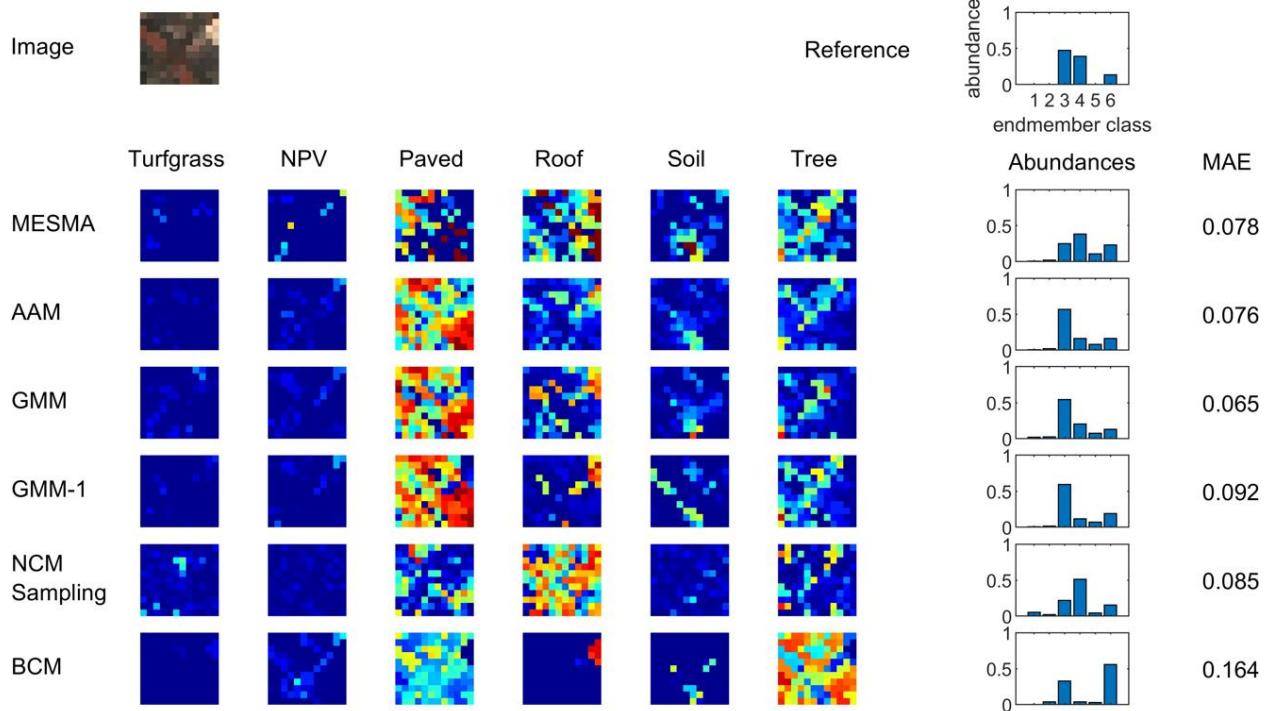


Figure S54: Abundance maps for polygon No. 54.

16 m



4 m

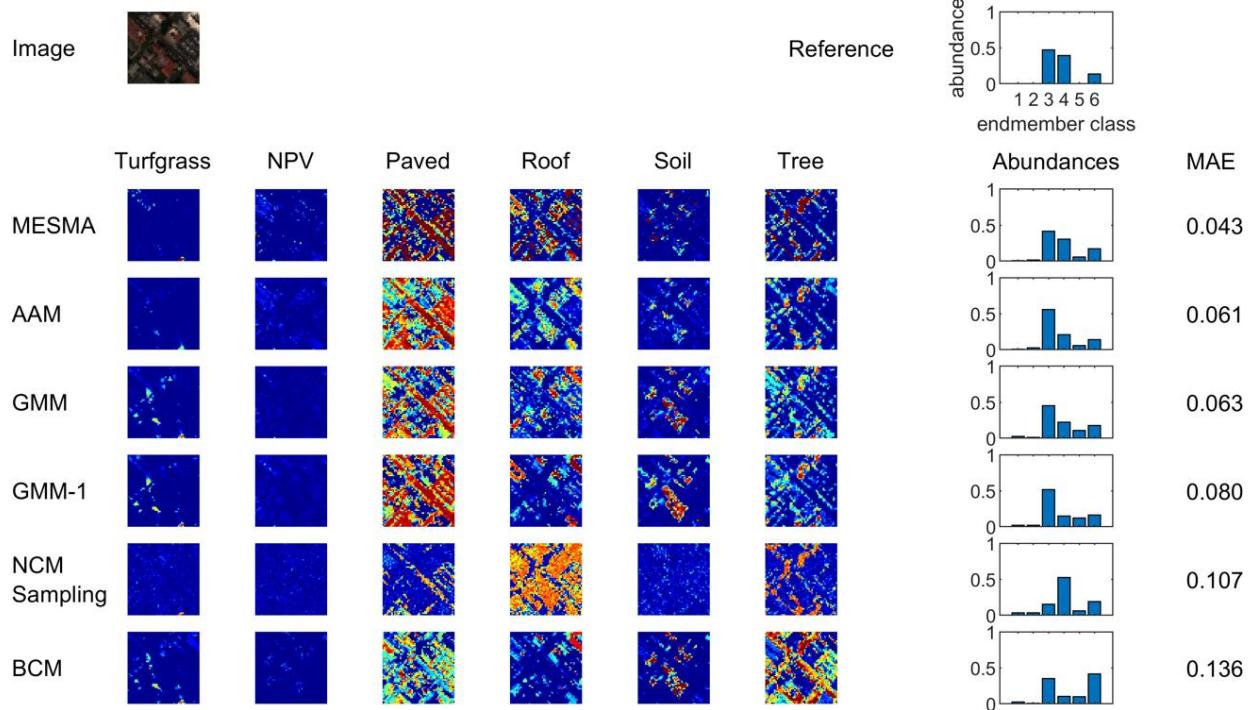
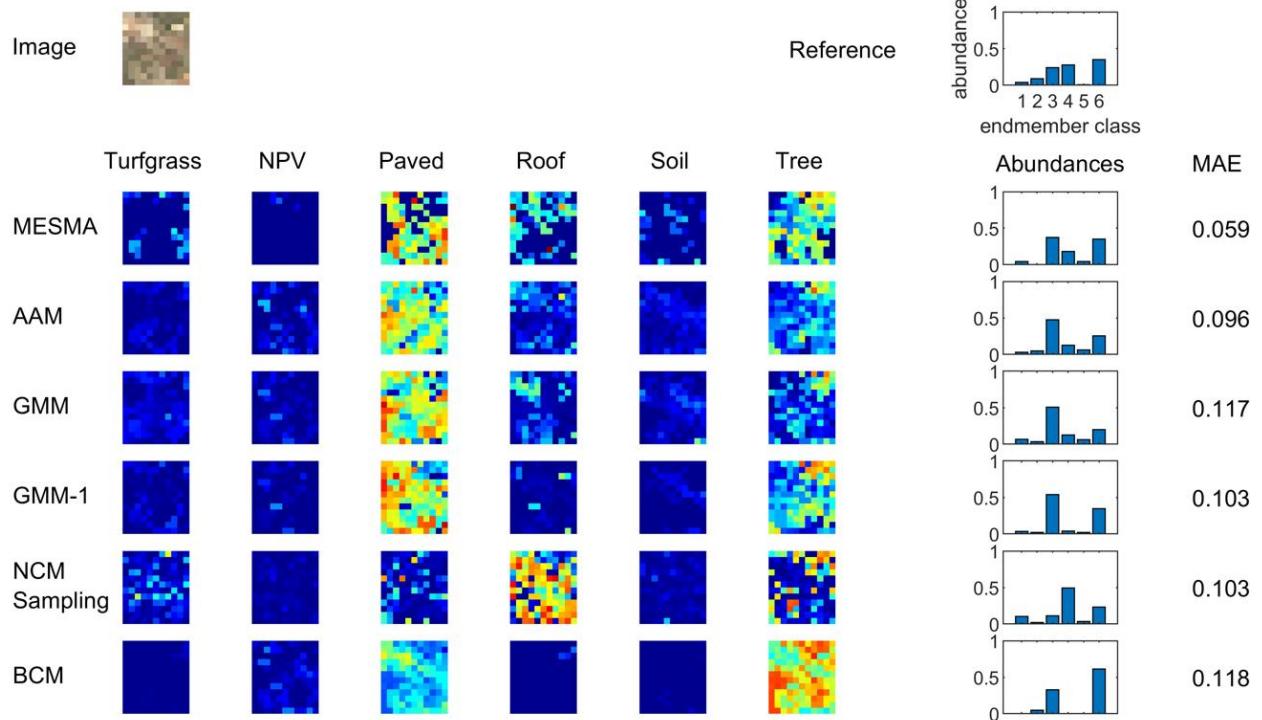


Figure S55: Abundance maps for polygon No. 55.

16 m



4 m

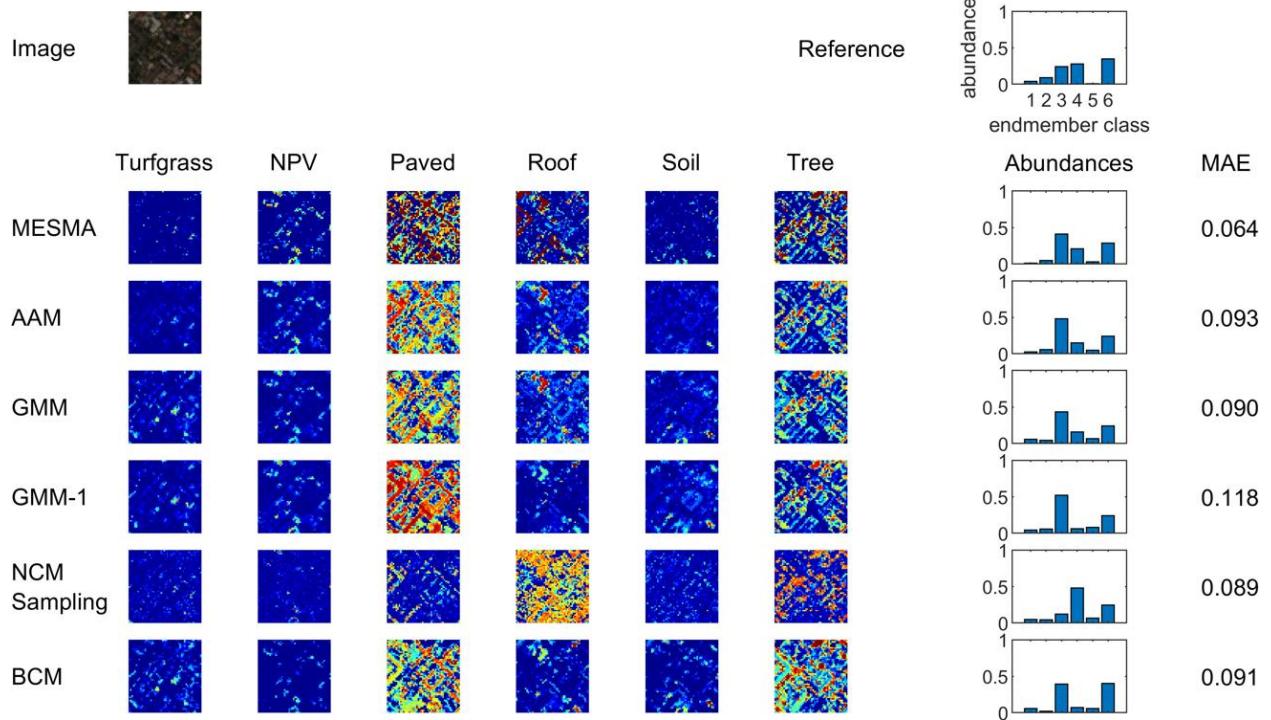
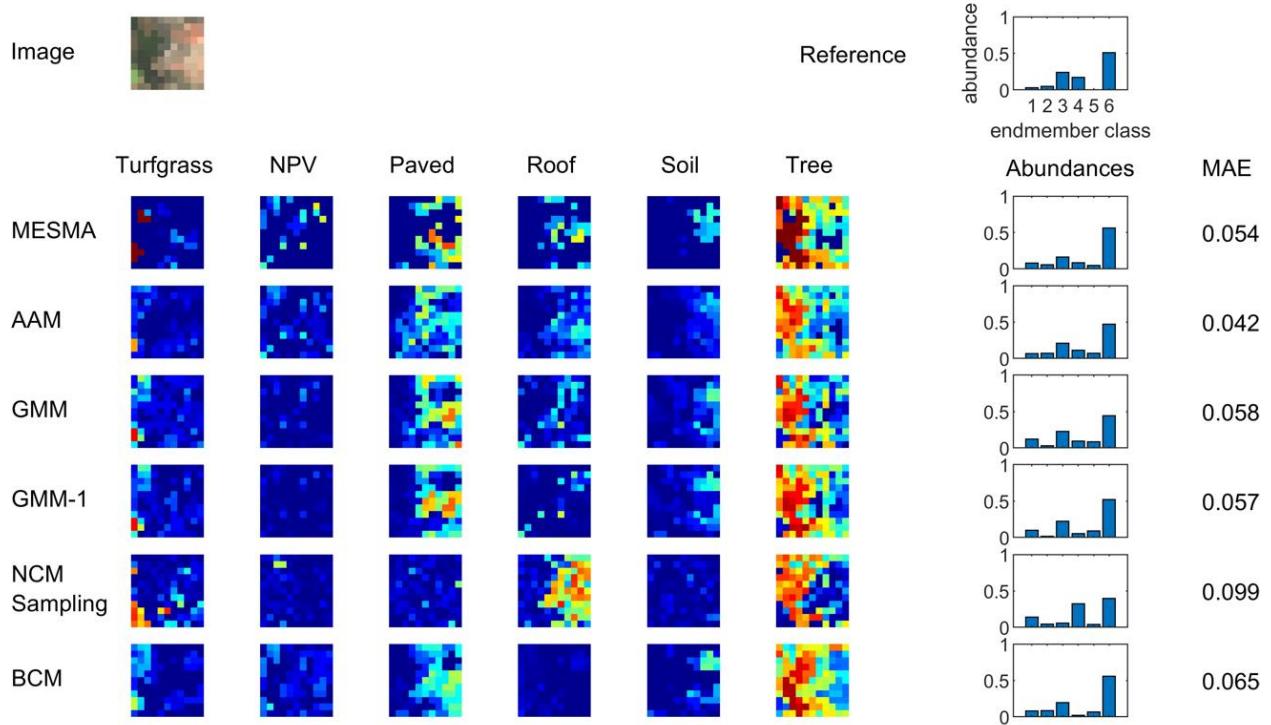


Figure S56: Abundance maps for polygon No. 56.

16 m



4 m

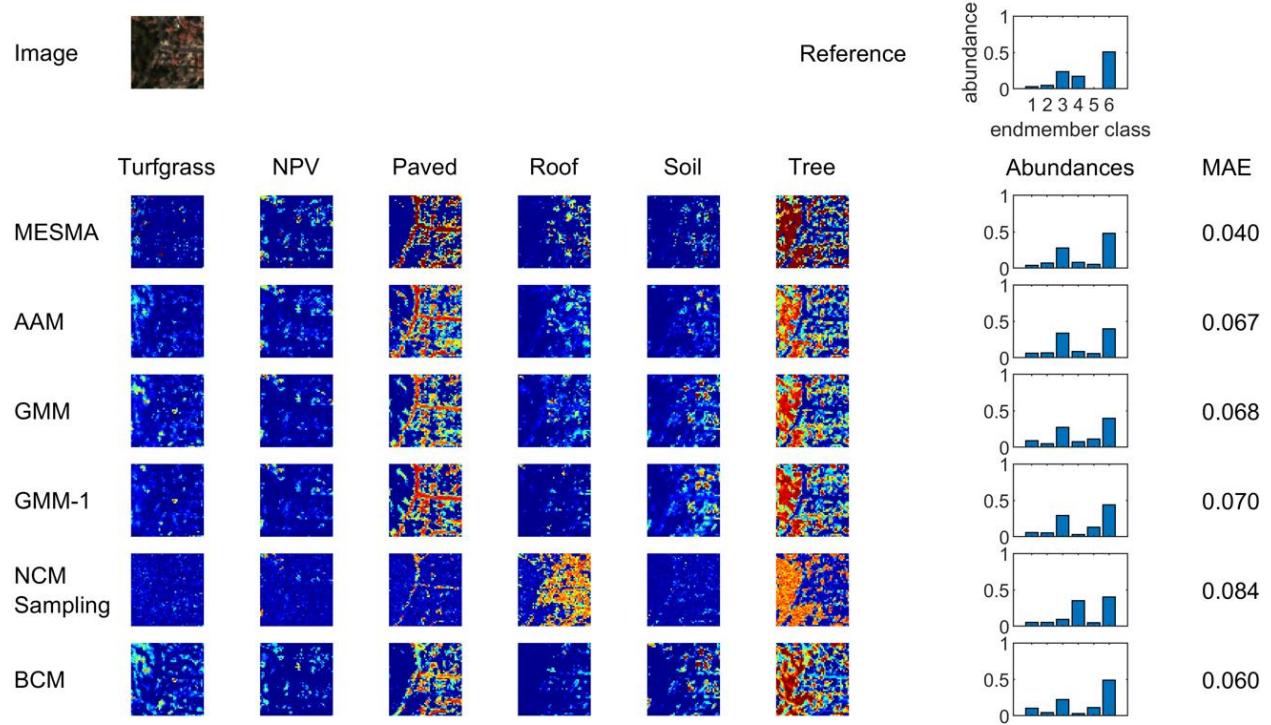
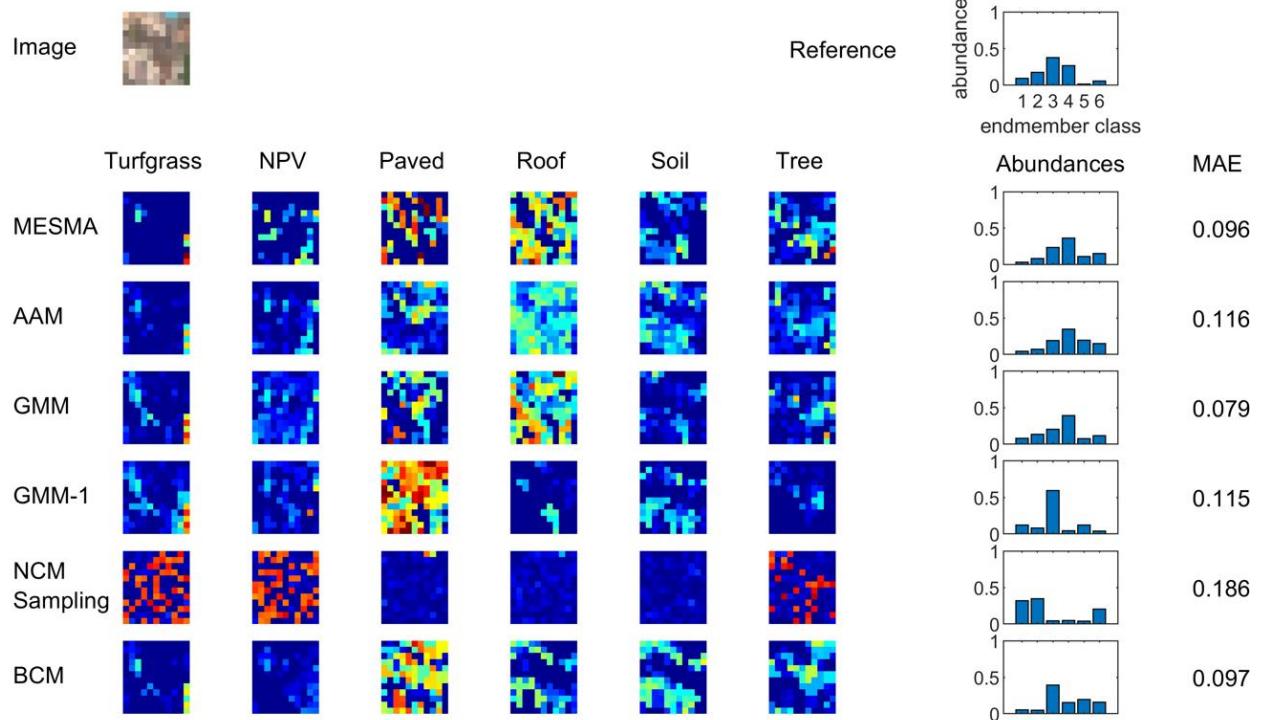


Figure S57: Abundance maps for polygon No. 57.

16 m



4 m

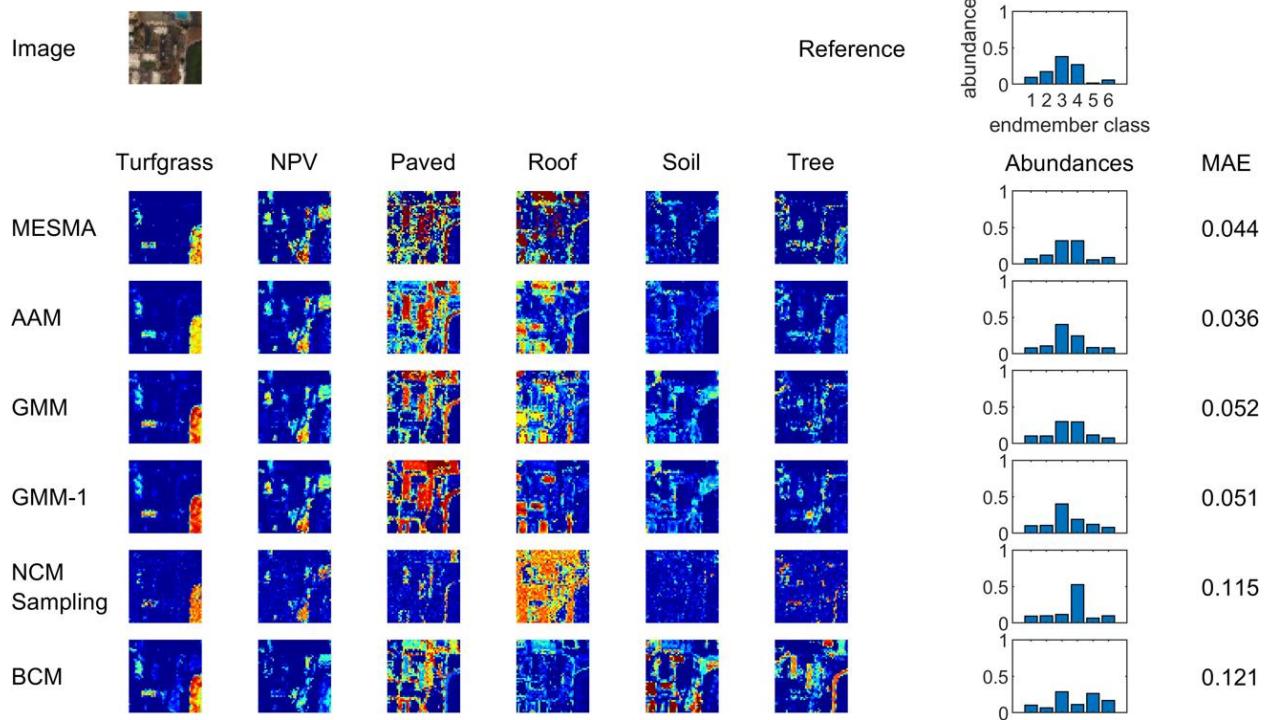
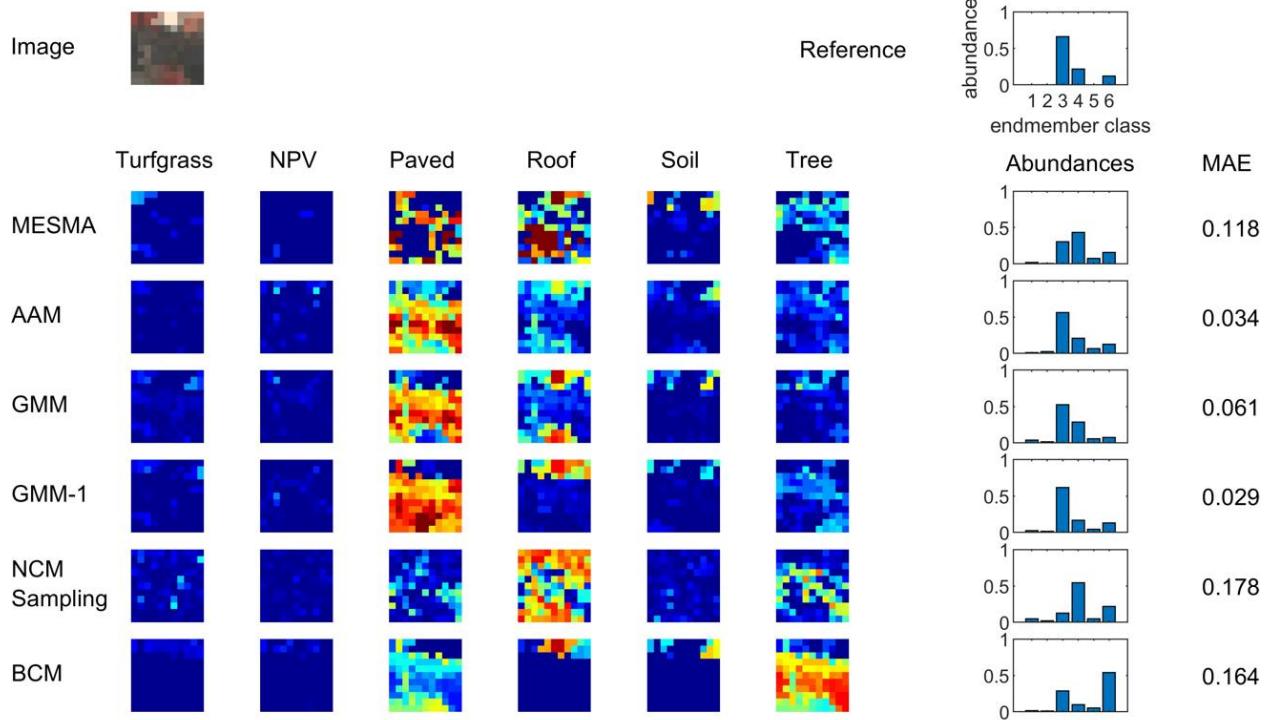


Figure S58: Abundance maps for polygon No. 58.

16 m



4 m

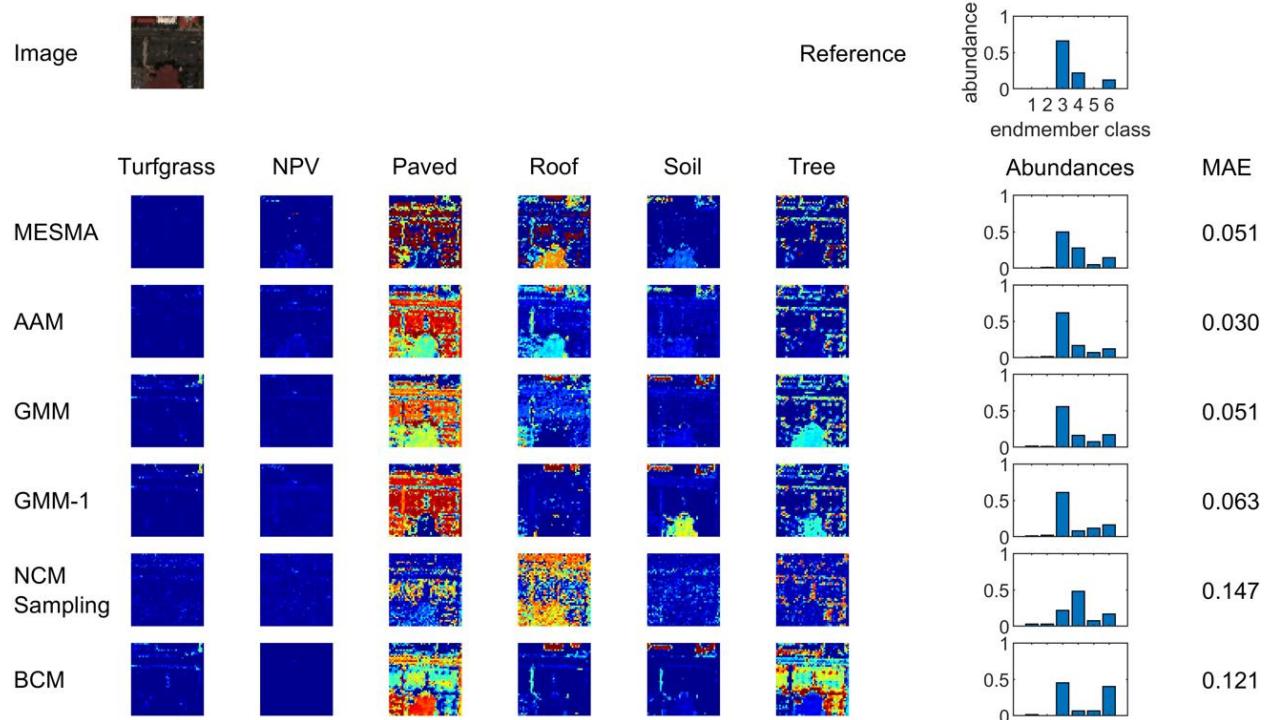
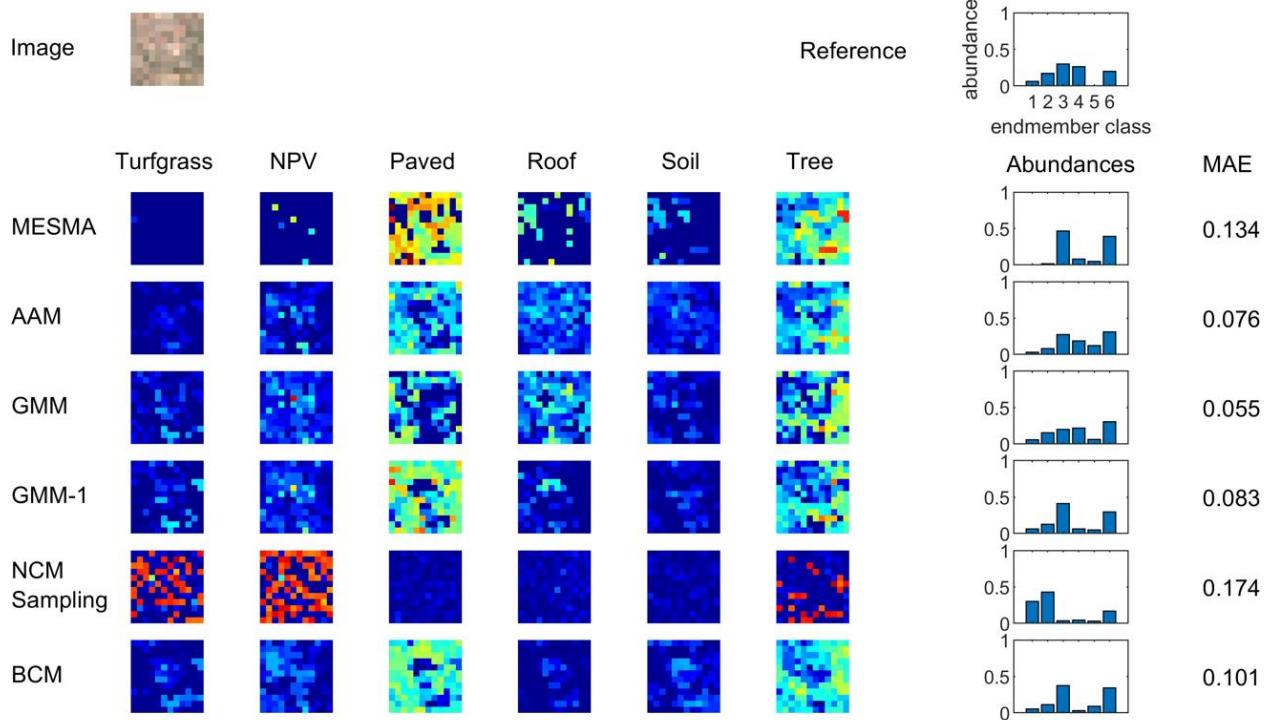


Figure S59: Abundance maps for polygon No. 59.

16 m



4 m

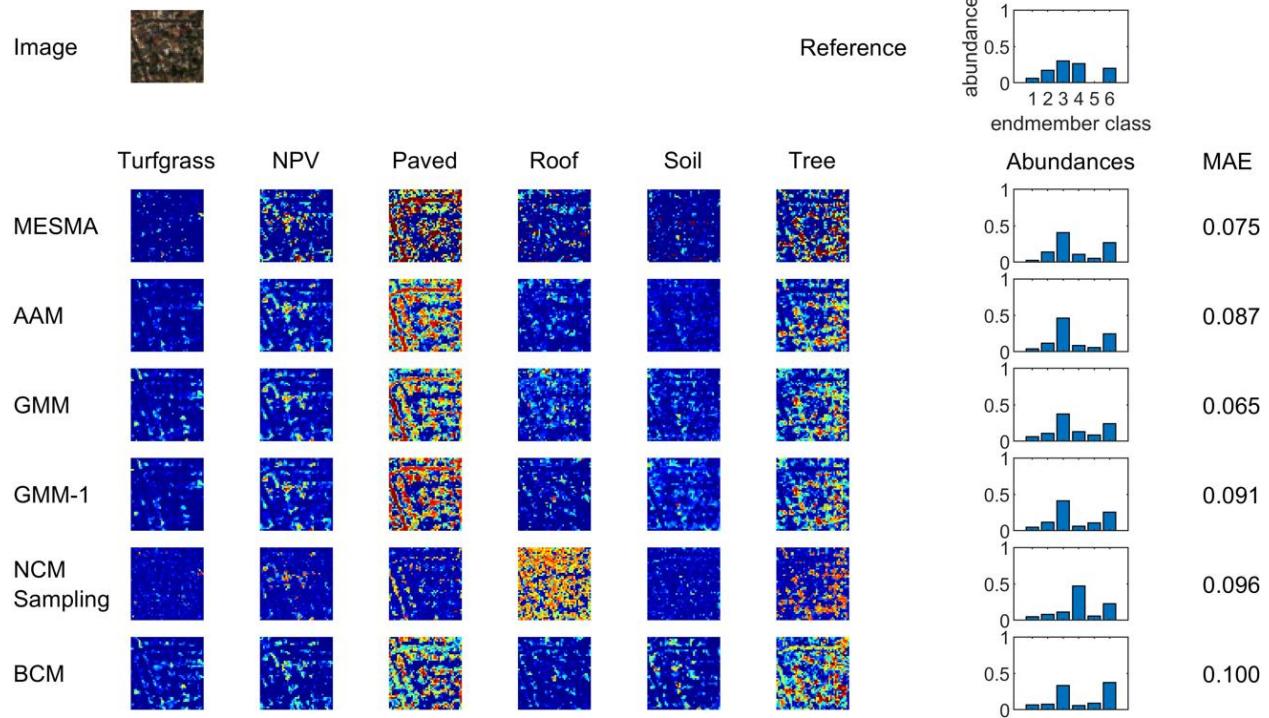
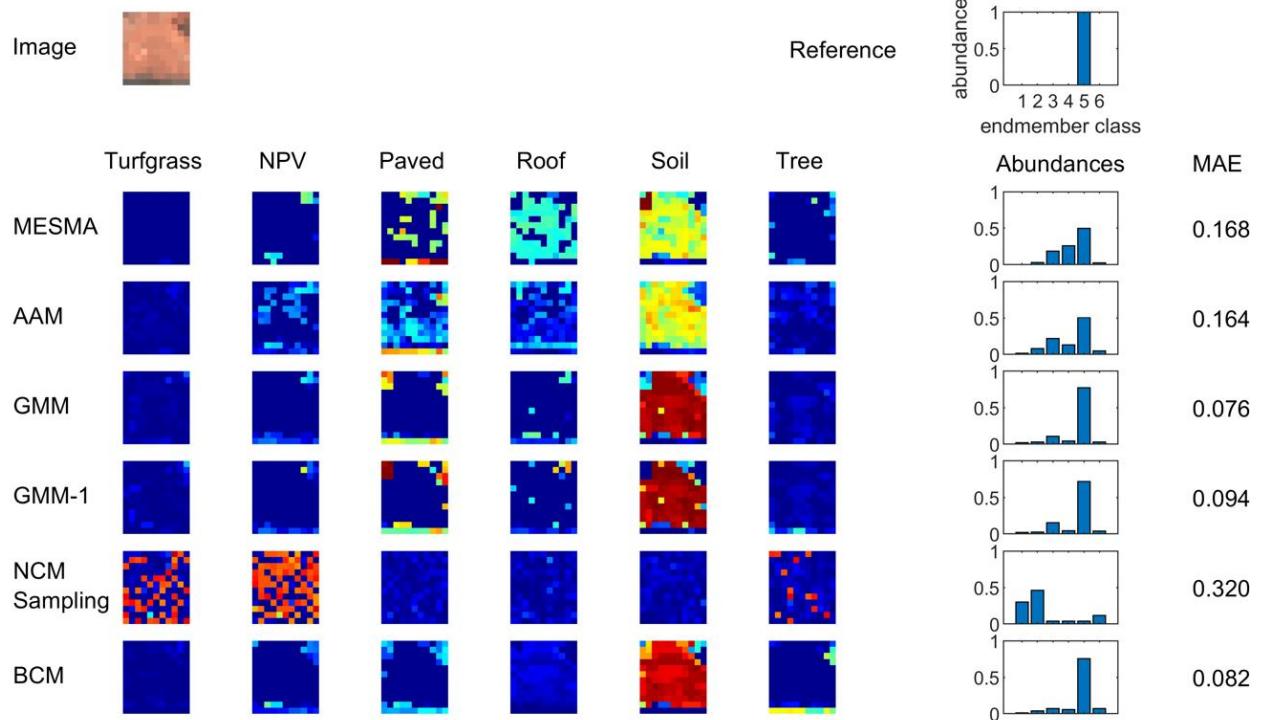


Figure S60: Abundance maps for polygon No. 60.

16 m



4 m

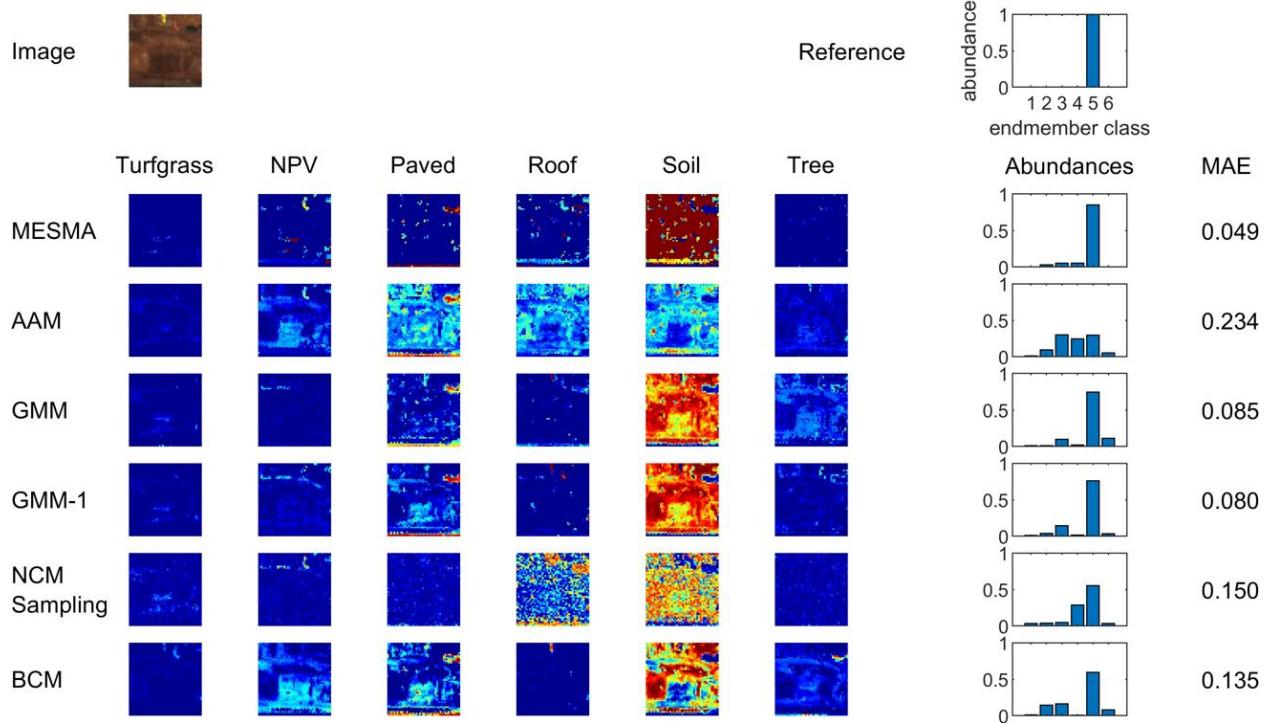
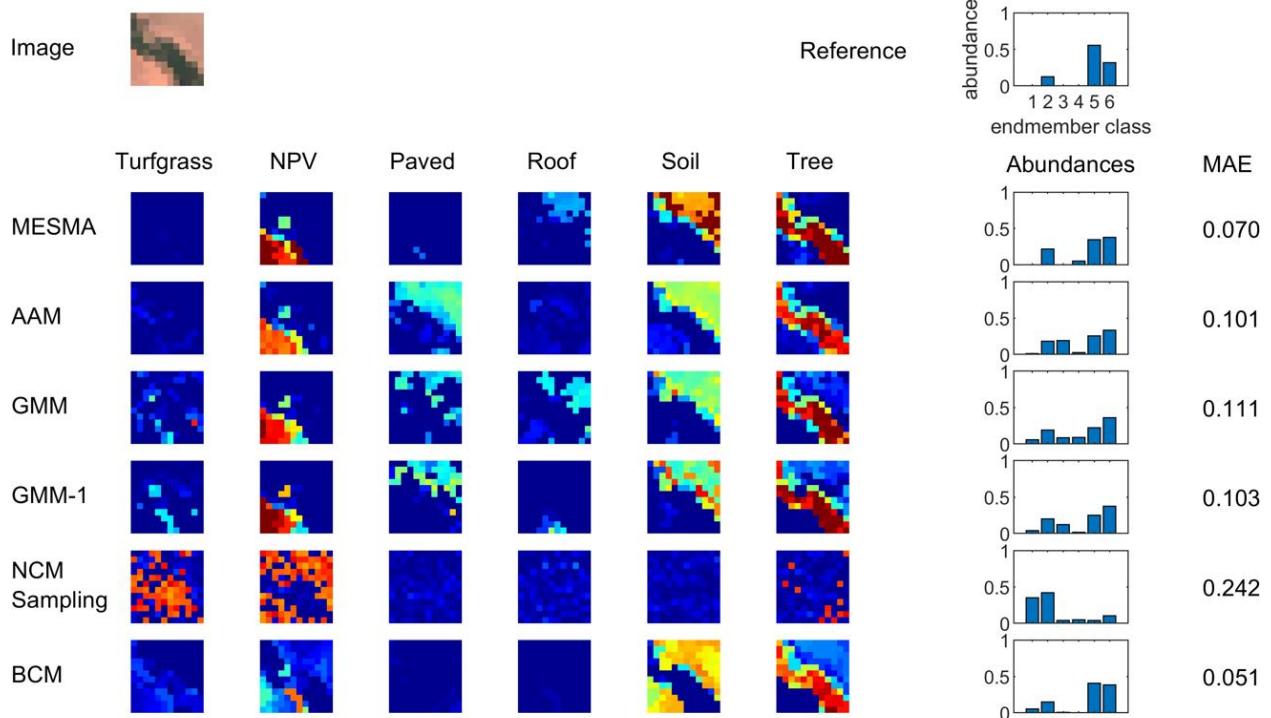


Figure S61: Abundance maps for polygon No. 61.

16 m



4 m

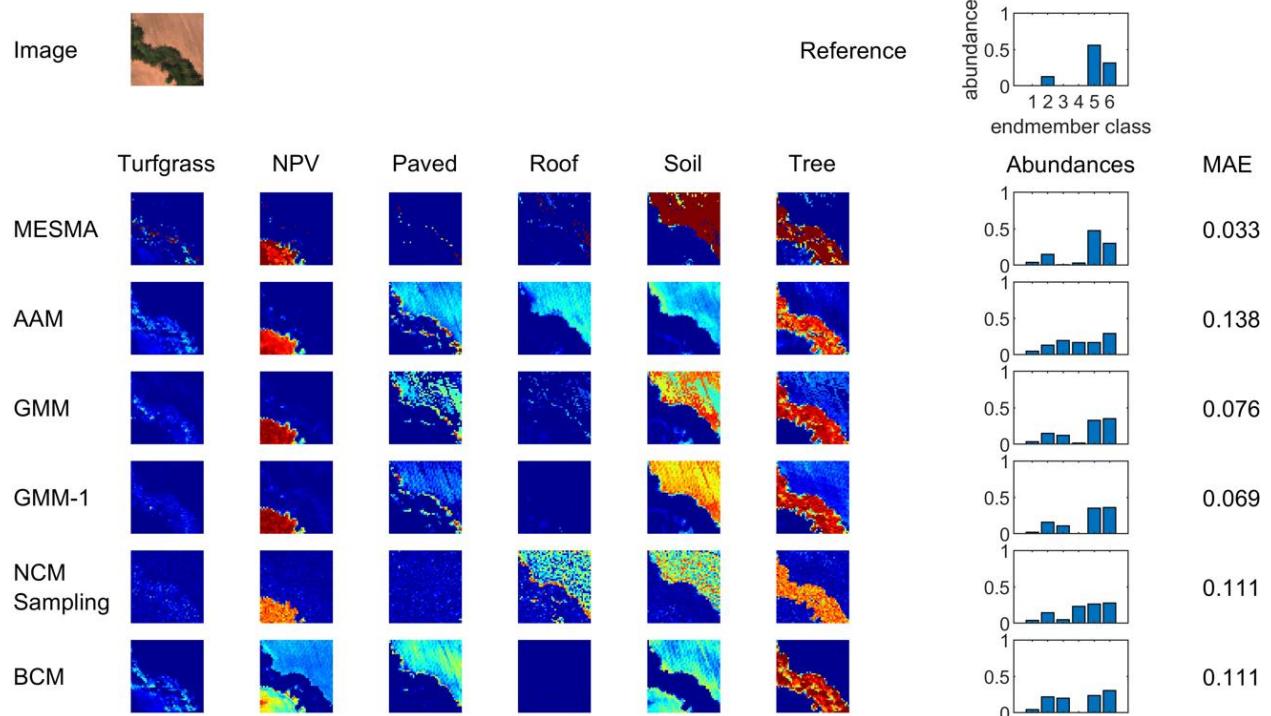
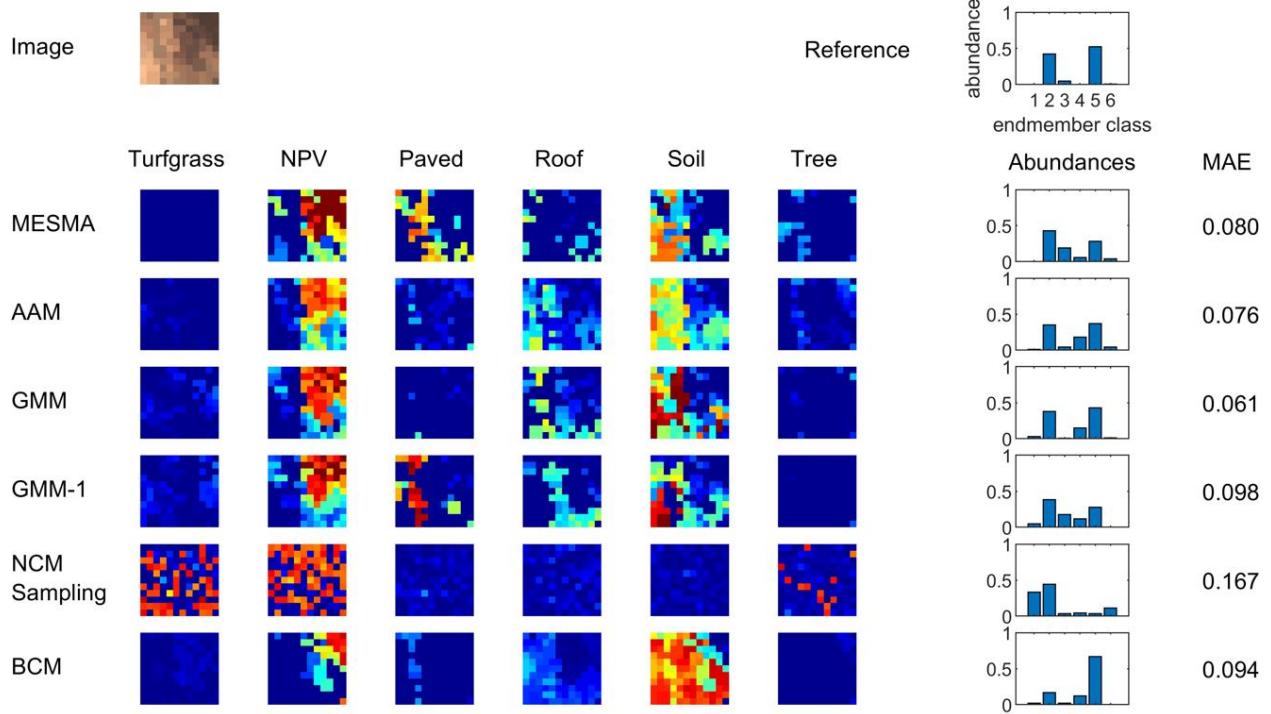


Figure S62: Abundance maps for polygon No. 62.

16 m



4 m

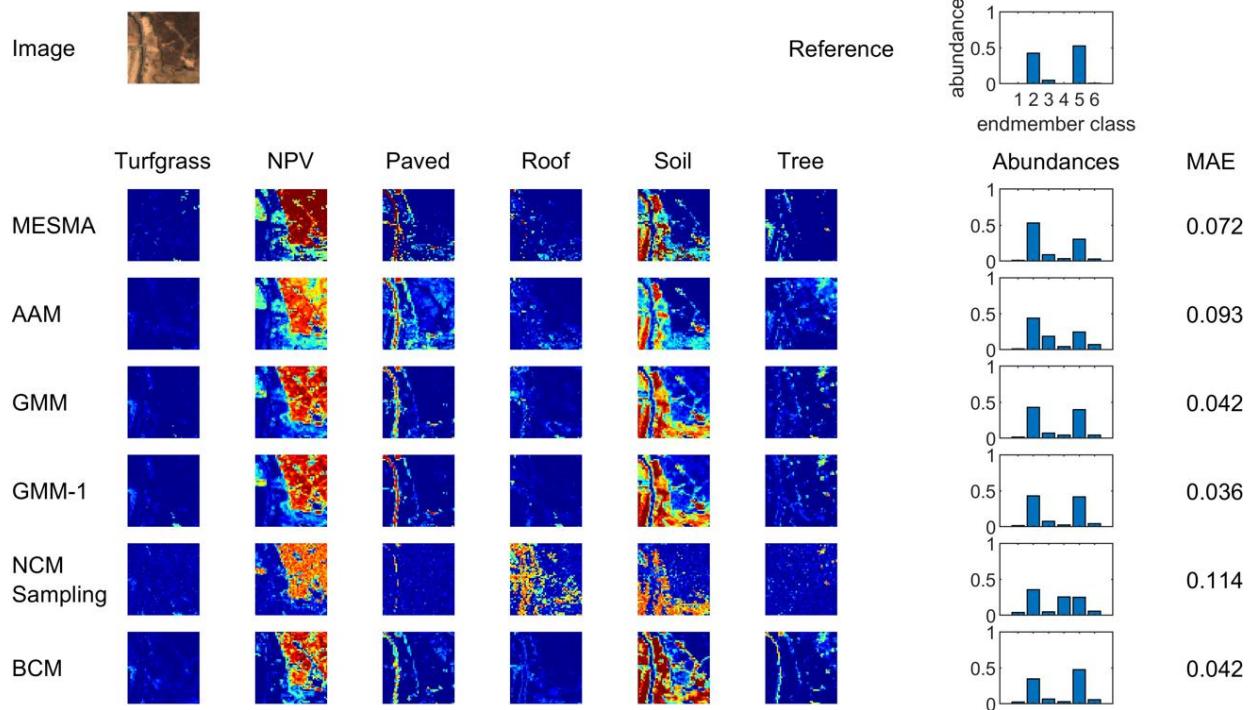
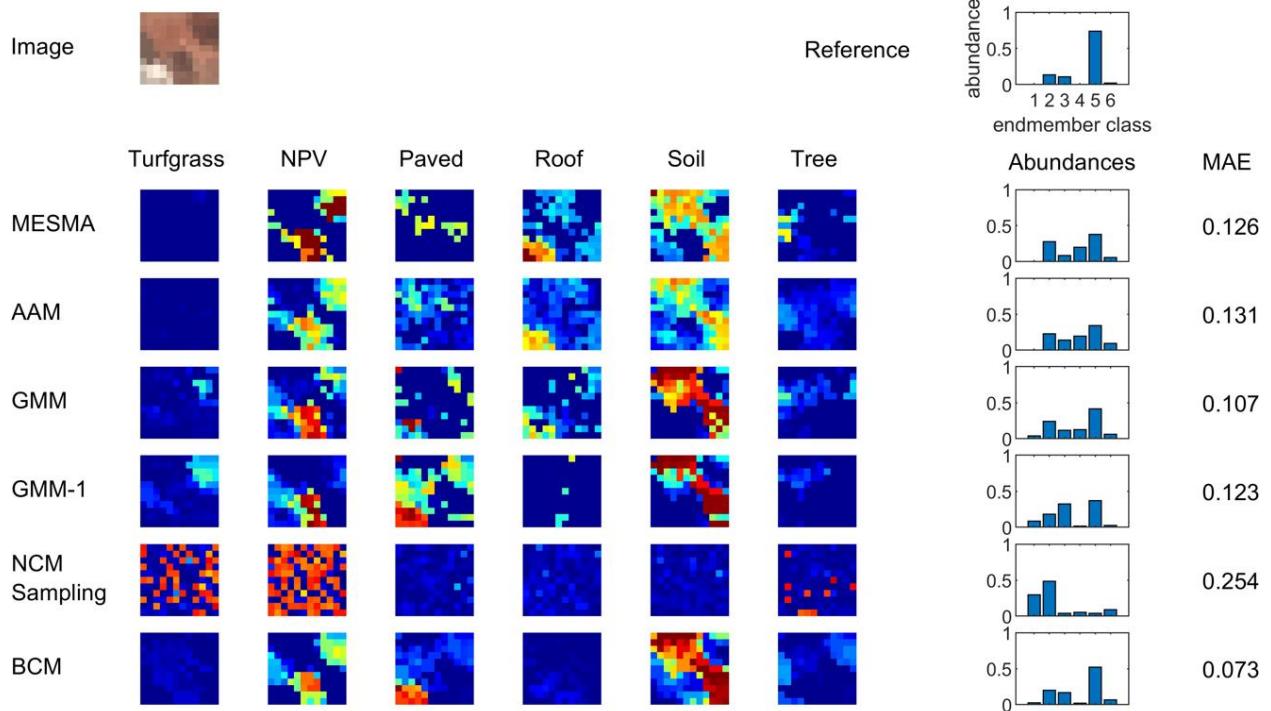


Figure S63: Abundance maps for polygon No. 63.

16 m



4 m

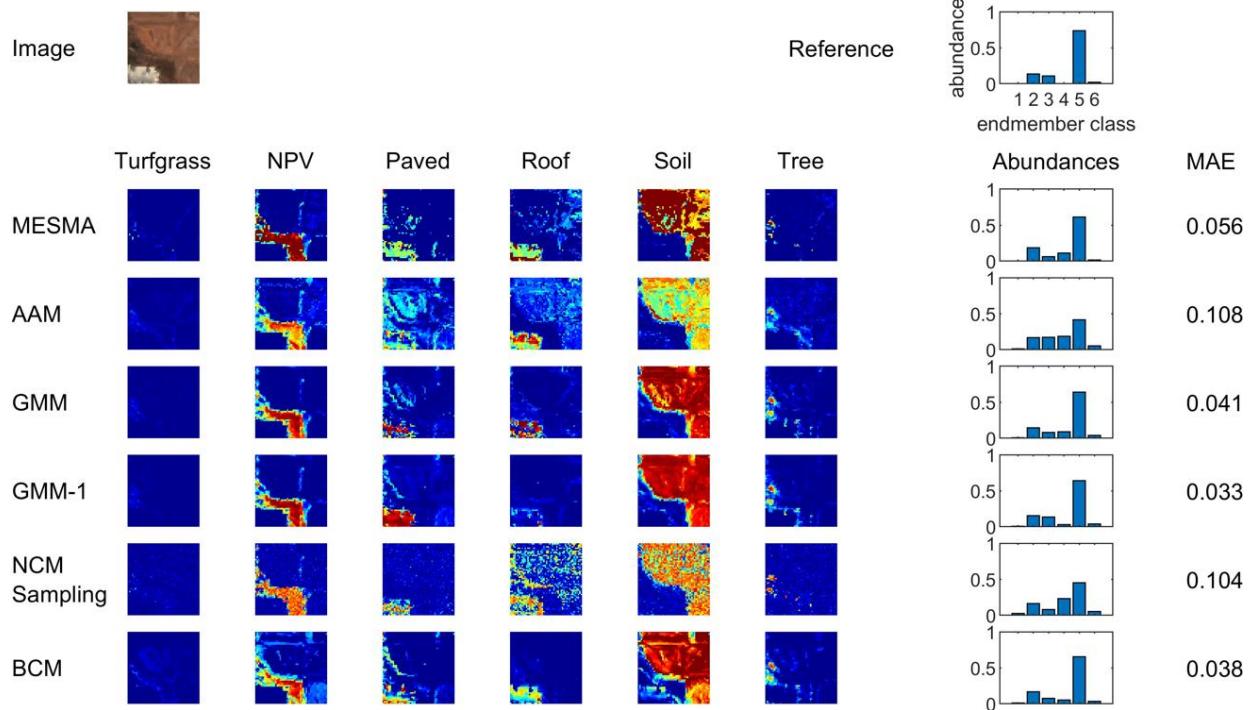


Figure S64: Abundance maps for polygon No. 64.

## 2. Additional Results

We also attached some additional results for the real dataset. First we show the scatter plots of abundance values for the 4 m data, which are missing in the main text.

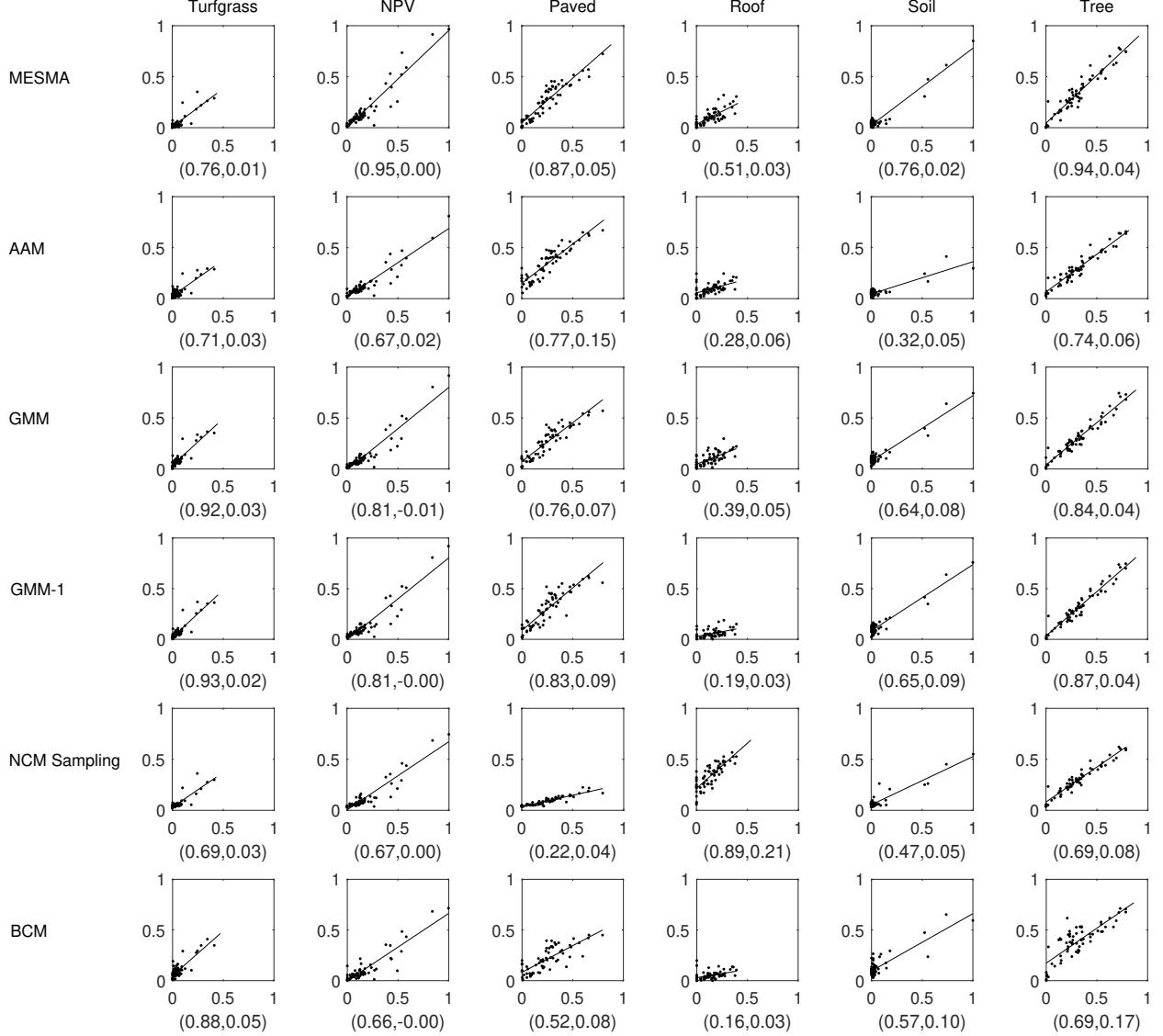


Figure S65: Scatter plots of 64 abundance values in 4 m for the ground truth (x-axis) and the estimated (y-axis). The parenthesis below the plot contains the slope and intercept.

Then we show the scatter plots and Bland-Altman plots for merged abundances in 16 m. The scatter plots (Fig. S66 and Fig. S68) show that the increase in  $R^2$  value from merged abundances is mainly caused by the dispersion of abundances.

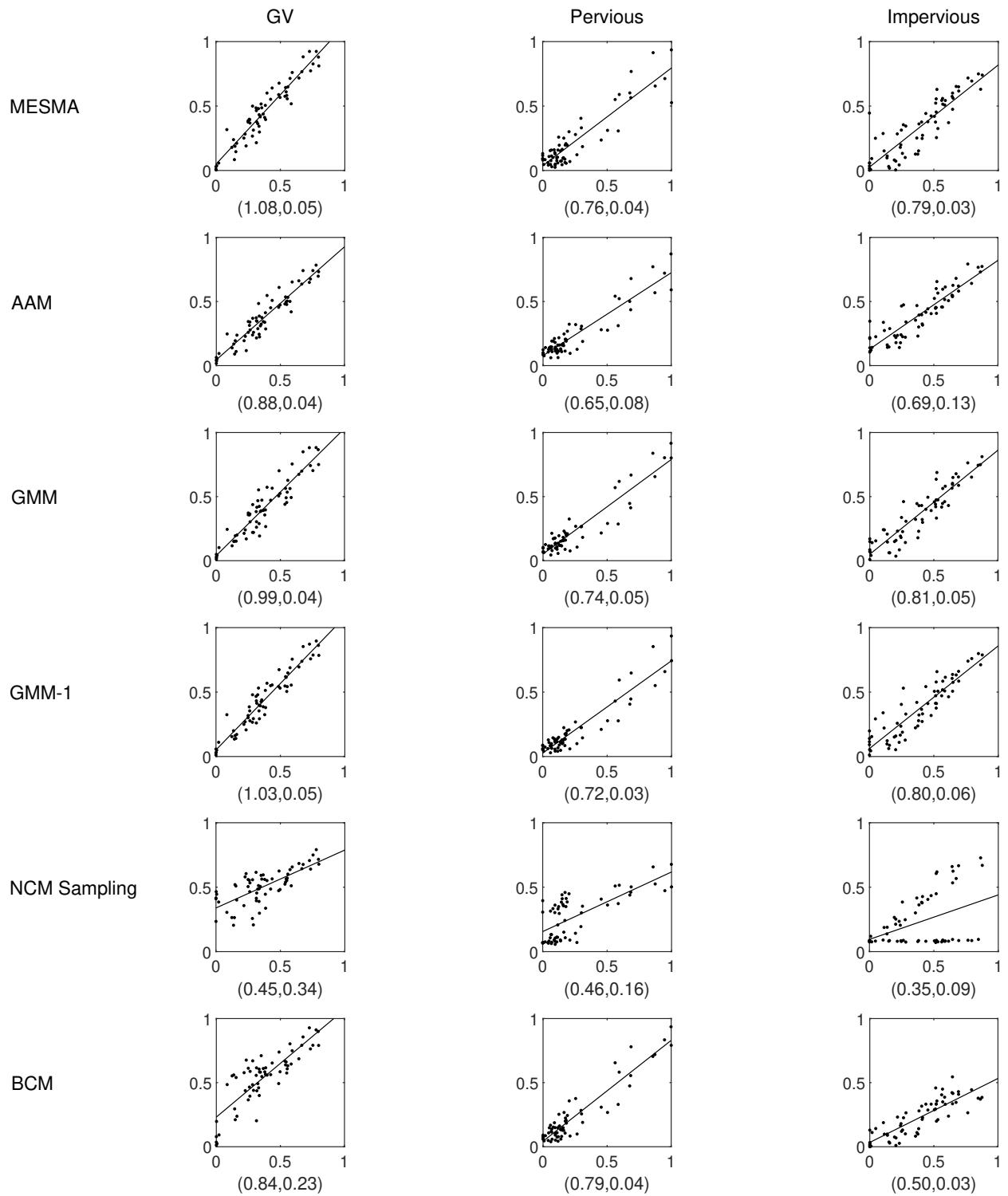


Figure S66: Scatter plots of 64 merged abundance values in 16 m for the ground truth (x-axis) and the estimated (y-axis). The parenthesis below the plot contains the slope and intercept.

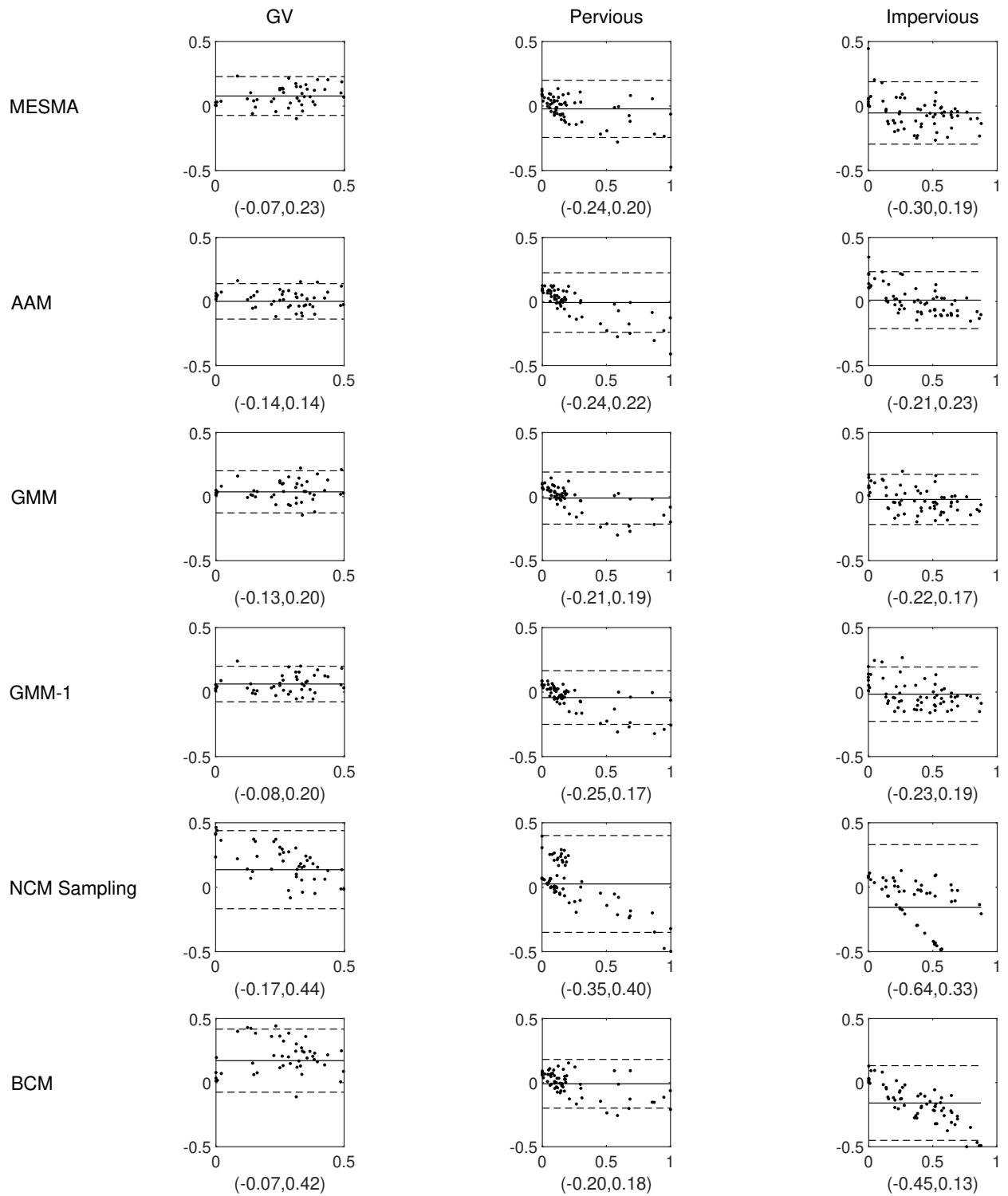


Figure S67: Bland-Altman plots of 64 merged abundance values in 16 m for the ground truth (x-axis) and the difference (y-axis). The parenthesis below the plot contains the 95% interval.

Finally we show the scatter plots and Bland-Altman plots for merged abundances in 4 m.

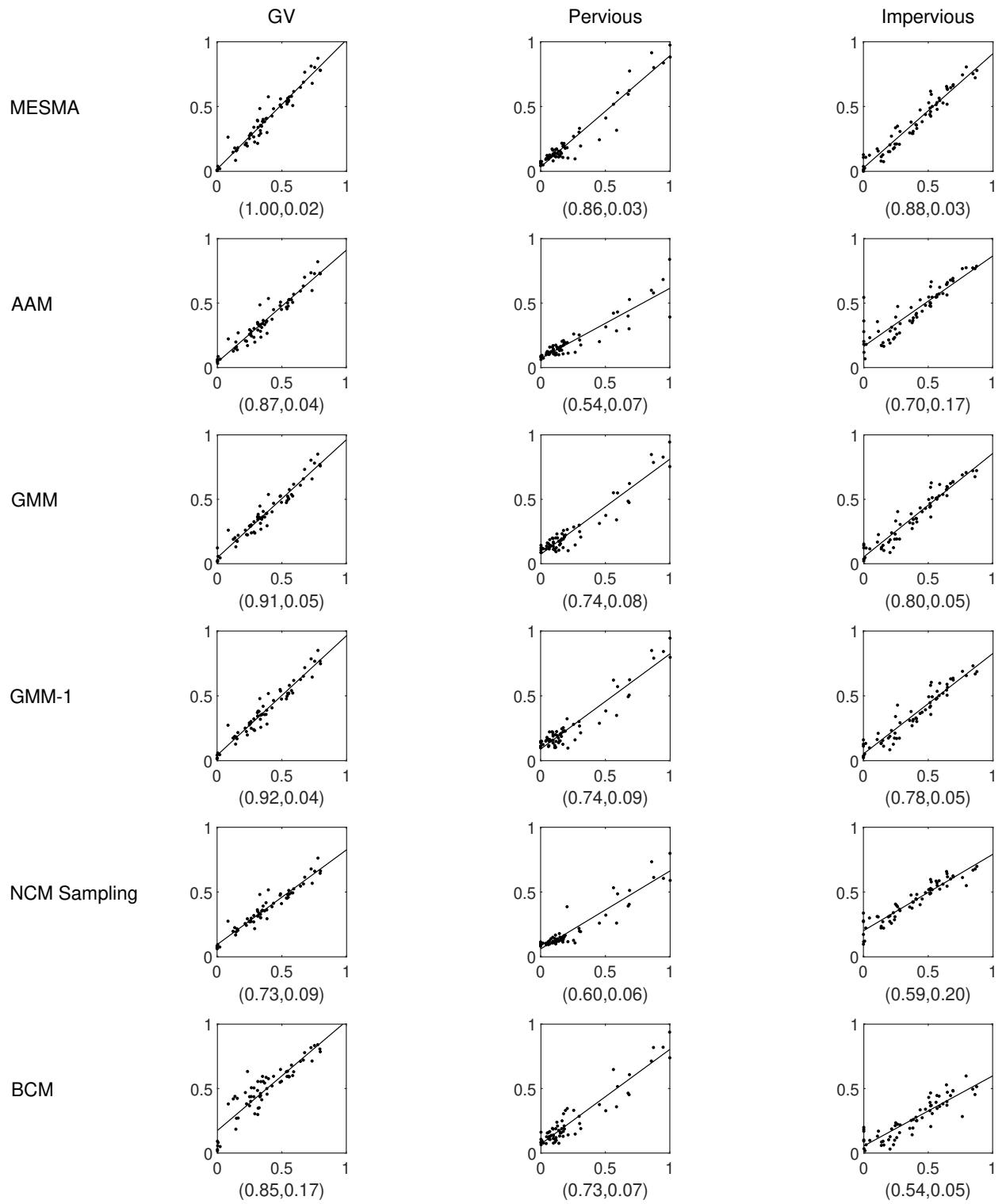


Figure S68: Scatter plots of 64 merged abundance values in 4 m for the ground truth (x-axis) and the estimated (y-axis). The parenthesis below the plot contains the slope and intercept.

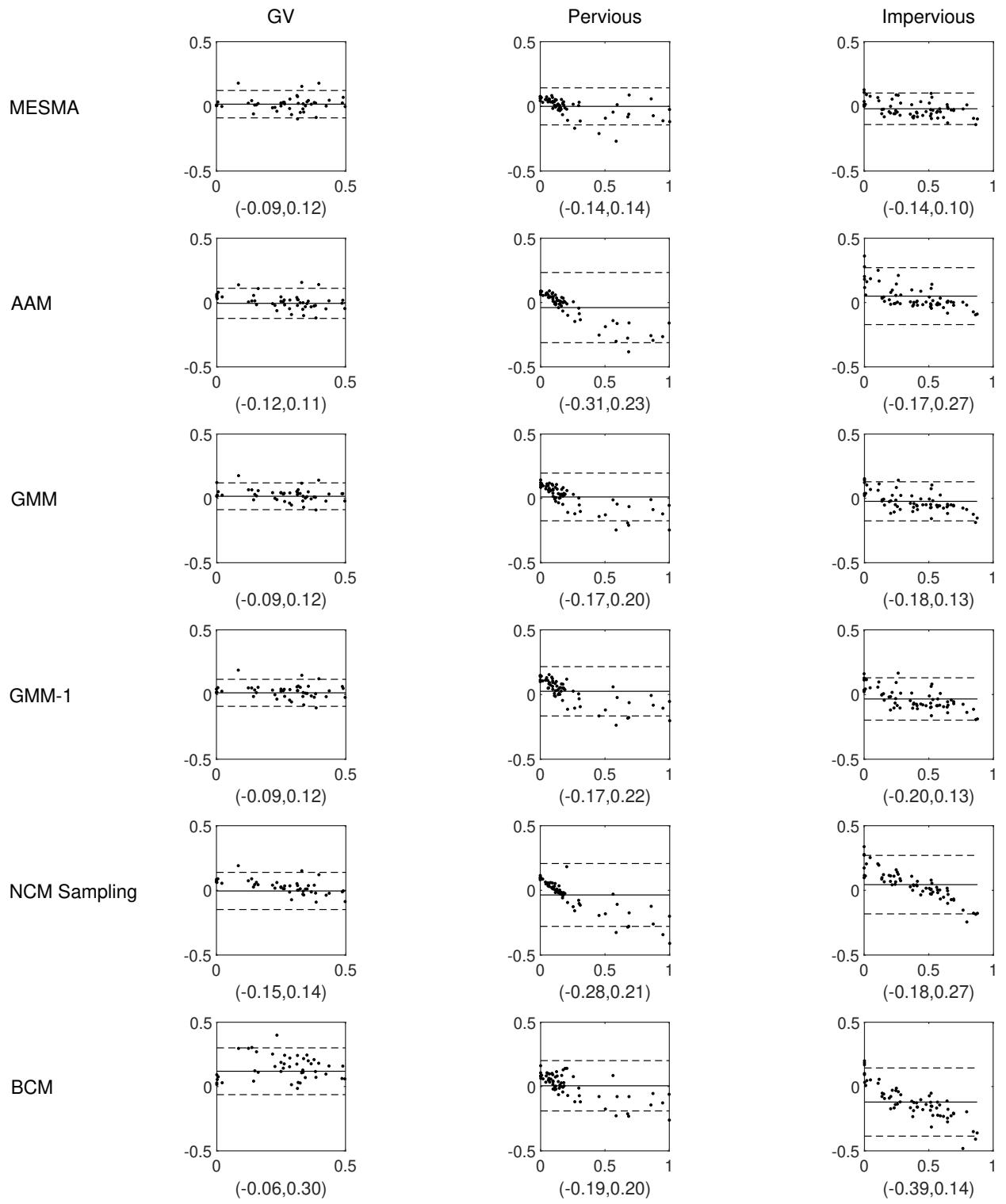


Figure S69: Bland-Altman plots of 64 merged abundance values in 4 m for the ground truth (x-axis) and the difference (y-axis). The parenthesis below the plot contains the 95% interval.