APPLICATION

*biodivMapR*: An *R* package for α- and β-diversity mapping using remotely sensed images

Jean-Baptiste Féret | Florian de Boissieu

TETIS, Irstea, CIRAD, CNRS, Université Montpellier, Montpellier, France

Correspondence Jean-Baptiste Féret Email: jb.feret@teledetection.fr

Funding information Agence Nationale de la Recherche, Grant/Award Number: ANR-17-CE32-0001; French Space Agency (CNES)

Handling Editor: Kylie Scales

Abstract

1. The accelerated erosion of biodiversity is a critical environmental challenge. Operational methods for the monitoring of biodiversity taking advantage of remotely sensed data are needed in order to provide information to ecologists and decision-makers.

2. We present an *R* package designed to compute a selection of α- and β-diversity indicators from optical imagery, based on spectral variation hypothesis. This package builds upon previous work on biodiversity mapping using airborne imaging spectroscopy, and has been adapted in order to process broader range of data sources, including Sentinel-2 satellite images.

3. *biodivMapR* is able to produce α-diversity maps including Shannon and Simpson indices, as well as β-diversity maps derived from Bray–Curtis dissimilarity. It is able to process large images efficiently with moderate computational requirements on a personal computer. Additional functions allow computing diversity indicators directly from field plots defined as polygon shapefiles for easy comparison with ground data and validation.

4. The package *biodivMapR* should contribute to improved standards for biodiversity mapping using remotely sensed data. It should also contribute to the identification of relevant Remotely Sensed enabled Essential Biodiversity Variables.

KEYWORDS

alpha diversity, beta diversity, biodiversity mapping, essential biodiversity variables, remote sensing, Sentinel-2

1 | INTRODUCTION

As the urge to scale up ambitions to protect global diversity is now acknowledged, conservation goals need to be implemented efficiently (IPBES, 2019). Earth observation based on airborne and satellite systems is particularly important for biodiversity monitoring, as it allows production of maps for spatially explicit modelling and monitoring from local to global scale (Féret et al., 2017; Rocchini et al., 2018, 2016). Operational methods for biodiversity monitoring taking advantage of remote-sensing data need to be identified, qualified and compared with each other. Optical remote sensing showed very strong potential for the estimation of several components of biodiversity, even for very complex ecosystems such as tropical forests (Baldeck et al., 2015; Féret & Asner, 2014; Leitão et al., 2015; Pettorelli et al., 2016; Rocchini et al., 2018; Somers et al., 2015; Vaglio Laurin et al., 2014). Operational methods should require minimum supervision and should not rely on extensive ground data collection, to ensure applicability even in remote places with no access from the...
ground in tropical environments. Féret and Asner (2014) developed a method meeting these requirements and validated its performances for tropical forest diversity mapping based on very high spatial resolution airborne imaging spectroscopy. Sentinel-2 data now provides global coverage of Earth surface with a revisit time of five days and high spatial resolution multispectral imagery. The package biodivMapR is an adaptation of the method proposed by Féret and Asner (2014) to Sentinel-2 data, and aims at providing operational tools to ecologists and remote-sensing communities in order to accelerate and ease their capacity to assess existing methodologies based on abundant remotely sensed data. Here biodivMapR provides a limited set of functions to process large datasets, from high dimensionality imaging spectroscopy acquired by airborne sensors or satellites, to high spatial coverage multispectral satellites such as Sentinel-2 images. The method is based on the spectral variation hypothesis (Palmer, Earls, Hoagland, White, & Wohlgemuth, 2002; Rocchini et al., 2010; Rocchini, Chiarucci, & Loiselle, 2004), and takes advantage of high spatial resolution multispectral information to differentiate species or groups of species based on the optical traits corresponding to the reflectance of each pixel (Homolová, Malenovský, Clevers, García-Santos, & Schaepman, 2013; Ustin & Gamon, 2010). However, it may be applicable for other types of data than optical data, as long as the hypothesis of spatial heterogeneity is relevant to estimate specific biodiversity components. The package includes a sample of Sentinel-2 image, and a script illustrating the workflow of the method: the optical image is transformed with principal component analysis (PCA), and a subset of selected components is used to produce spectral species maps based on k-means clustering. α- and β-diversity metrics including Shannon index and Bray–Curtis dissimilarity are then computed from this spectral species map at the spatial resolution of a window size defined by user. Some functions are also provided in order to perform direct validation if ground-truth is available.

2 DESCRIPTION OF THE PACKAGE

The workflow of the core processing of the package is presented in Figure 1 and the full pseudo code is proposed in Appendix S1. Preprocessing functions are available in the package and are introduced in the next section. Front-end user interface is limited to core functions, in order to apply this workflow as straightforwardly as possible, and parameterize the different steps of the process flawlessly, even with limited expertise in image processing. Back-end functions are documented within the code, making it easier for a more experienced person to develop upon biodivMapR. We strongly recommend using bottom of atmosphere reflectance instead of top of atmosphere reflectance or radiance.

2.1 Data checking and conversion

The current version of the package requires ENVI HDR format with band-order set to BIL (‘Bands Interleaved by Line’) for all raster data. This decision was initially motivated by computational efficiency, but next versions of the package will accept more raster formats. Please notice that the current version of the code expects spectral bands to be defined in the field ‘wavelength’ of the header file associated with the binary image. The raster format (image and mask) is checked internally when running the functions perform__radiometric__filtering and perform__PCA. A message is return, stating if the image format is as expected, or if it should be modified. If the image format should be modified, the process will automatically stop, in order to let the user convert the image. In this case, the function raster2BIL allows conversion to the expected format. However, raster2BIL is based on the raster package (Hijmans, 2018) and converts any GDAL format into ENVI HDR format with BIL interleave. It also includes additional options such as conversion.
to integers to reduce file volume, and definition of the sensor for automated header formatting.

2.2 Radiometric filtering

The main purpose of the package is biodiversity mapping of vegetated areas, with a focus on forest monitoring, and more specifically tropical forests. Hence, preliminary filtering of irrelevant pixels is crucial to get optimal results from the method. Irrelevant pixels include non-vegetated areas, shaded areas, and cloudy areas. The function `perform_radiometric_filtering` performs a preliminary filtering in order to mask such pixels. This filter applies basic radiometric thresholding, and users may need to define a complementary mask as initial input if the filtering functions available in the package are not sufficient or relevant. The thresholding is only relevant when using optical multi or hyperspectral images, and is performed as follows:

- **Non-vegetated/Dry vegetation pixels**: they are identified based on a low thresholding of the Normalized Difference Vegetation Index.
- **Shaded areas**: these are characterized by low overall reflectance. Pixels characterized by a near infrared (NIR) reflectance inferior to a given threshold are then masked. In temperate and tropical forests, a low threshold of 15%-20% reflectance in the NIR domain (800-1100 nm) usually results in correct detection of shaded areas. However, this does not hold with coniferous forests.
- **Cloudy pixels**: residuals from atmospheric corrections may lead to increased reflectance in the blue domain. Therefore, a high thresholding can be performed on the blue channel, when available.

The blue, red and NIR spectral bands as well as the corresponding thresholds can be adjusted by user. The updated mask is then saved and used during the following steps of the process. This radiometric filtering should not be performed when the input image does not include blue, red, or NIR spectral bands. The above-mentioned criteria may also be irrelevant if both dry and green vegetation should be monitored. Users are invited to judge on the relevance of these filters on their own situation.

2.3 Computation of the PCA

A series of pre-processing steps, including atmospheric water band removal and continuum removal (CR) of the reflectance data and a PCA are applied on the image, by calling the function `perform_PCA`.

*Continuum removal* is applied on high spatial resolution imaging spectroscopy in order to reduce within-crown variability caused by brightness variations, and effects of changing illumination conditions between flightlines. It follows the hypothesis that these effects influence canopy reflectance homogeneously over the full spectrum, and can be assimilated to the application of a multiplicative factor to the canopy reflectance. This CR is computed by fitting a convex hull over the top of the reflectance spectrum, and defining it as a baseline continuum. This continuum is then removed by dividing it into the reflectance spectrum for each pixel in the image. The importance of CR may not be as strong when the spatial resolution is the same order as or larger than the dimension of tree crowns, as within-crown variability is integrated in the pixel reflectance. Users can choose whether to apply CR or not by setting the `Continuum_Removal` argument to `TRUE` or `FALSE` when calling `perform_PCA`.

The illustration in the Example section suggests that the application of CR on Sentinel-2 data leads to relevant results, and more tests should be done.

Principal component analysis prepares for feature extraction in order to reduce noise in the image. It is performed on a random subset of the image to ensure computational efficiency for large images. First, a random selection of pixels is performed, and PCA is computed based on this subset after CR. Once PCA is defined for the subset, it is then applied on the entire continuum removed image.

In some cases, the variability of the information contained in multi-/hyperspectral images may be explained by irrelevant pixels for the analysis, despite efforts to filter out these pixels. For example, fuzzy cloud borders may not be properly detected by standard filtering techniques or cloud detection algorithms. In this case, some principal components (PCs) may highlight these areas, resulting in poor definition of diversity patterns. Therefore, a second filtering based on PC thresholding can be activated. This second filtering discards automatically the pixels showing values beyond the mean PC value ±3 standard deviations for any of the five first components, and the mask is updated accordingly. Then the PCA preprocessing, including random pixel selection and CR is applied a second time with the updated mask. This operation significantly increases computation time and should be applied if the method does not produce satisfying results due to unwanted areas kept in the analysis.

2.4 Component selection

Principal component analysis is used as an image transformation method prior to feature selection. Feature selection allows users to list the most relevant features for biodiversity mapping, and discards those which may result in artifacts in diversity patterns. This feature selection is the only step of the full process requiring a decision from user. The application of PCA on imaging spectroscopy usually leads to components reflecting very different levels of relevance for vegetation study. Some components highlight differences in vegetation properties, characterized by differences among individuals and relatively homogeneous properties within individuals (when the spatial resolution is fine enough to identify individuals), whereas other components show information related to sensor characteristics, environment (slope orientation), or very high noise level. Some components also include a combination of these factors. The Spectral Response Non-uniformity (SRNU) described in the Sentinel-2 Data Quality Report is an example of such artifact (ESA, 2019), although compliant with the mission specifications. The difference in the measured radiometry,
corresponding to up to 2% in worst-cases, is particularly important for forested areas characterized by a reflectance between 2% and 5% in the VIS domain. PCA showed ability to separate radiometric effects caused by SRNU from variations explained by vegetation properties in different study cases. The rank of the components, i.e. the proportion of variance explained, is not systematically relevant to the identification of the components of interest, which means that selecting only the first components usually leads to suboptimal results (Féret & Asner, 2014). Finally, the decision to let user select components can be discussed as a bias in the estimation of biodiversity, as prior expectations may influence this selection (Bongalov et al., 2019). Unsupervised feature extraction methods are currently investigated. The function select_PCA_components creates and opens a text file where the user reports selected spectral bands based on visual interpretation of the PCA file.

2.5 | Spectral species mapping

Spectral species mapping is based on k-means clustering of the components selected from the PCA. First, a subset of pixel is extracted from PCA, equivalent to the number of pixels randomly selected when computing PCA. This pool of pixels is then split into nb_partitions partitions, and k-means clustering is performed on each partition, in order to define nbClusters clusters. The clustering is then applied to the whole PCA image in order to assign a cluster to each pixel based on closest centroid, for each repetition. Therefore, nb_partitions cluster maps result from these independent clusterings. Each cluster map is then processed until the computation of α- and β-diversity metrics, and these diversity metrics are finally averaged. The computation of these diversity metrics is explained in the next section. This operation can be considered as analogous to signal averaging, increasing signal to noise ratio by replicating measurements. The function map_spectral_species performs this clustering and writes the resulting raster file identifying the cluster corresponding to each pixel and each partition: the raster is named SpectralSpecies, with the same dimensions as the original image and nb_partitions bands, each band corresponding to a map of centroids for a specific partition.

Féret and Asner (2014) set nbClusters to 40 clusters, as a trade-off between performances or the estimation of biodiversity and computational efficiency. Here, we set 50 clusters as the default value in the example provided with the package corresponding to tropical forest. Our experience with biodiversity mapping of tropical forests suggests that this is a reasonable default value. However, preliminary studies performed on moderately diverse temperate forested sites showed optimal retrieval of Shannon index when using 20 clusters (results unpublished). This suggests that the number of clusters should be set depending on the level of heterogeneity of the landscape/ecosystem under study. Therefore, we recommend that users test different number of clusters in order to find the best trade-off for their own situation and document the number of clusters when publishing results.

2.6 | α-Diversity mapping

The diversity maps are computed from the SpectralSpecies raster. The α-diversity maps are produced with the function map_alpha_div, and correspond to the computation of an α-diversity metric based on the distribution of clusters for a given window size over the whole image. This window size is expressed in pixels (from the original image), and defined by the input variable window_size. Therefore, the actual surface of the windows used to compute the diversity metrics needs to be adjusted depending on i) the spatial resolution of the image being processed, and ii) the type of ecosystem and spatial scale of the diversity patterns expected. A minimum number of pixels are required to define this distribution: a low number of pixels lead to lower range of variation, and high sensitivity to noise. window_size also needs to be compatible with the surface usually defined in the field to perform species inventories, and needs to include a relevant number of individuals. Surfaces between 0.5 ha and 4 ha are usually compatible with field observations, and correspond to 50–400 pixels of 10 m resolution products from Sentinel-2. window_size is then usually defined in order to fall within this range. If data with coarser spatial resolution are processed (LandSat: 30m, MODIS: 500m), the hypotheses behind the method (pixel size inferior or equivalent to individuals) may not be respected.

biodivMapR 1.0 includes two α-diversity indicators: Shannon and Simpson index. More indices will be available in the next versions of the package.

2.7 | β-Diversity mapping

β-Diversity maps are derived from the pairwise Bray–Curtis dissimilarity (BC) matrix obtained from the spectral species map. In the initial version of the method, the BC matrix was then processed with an ordination method in order to project the n x n dimensional space of the dissimilarity matrix into an n x 3 dimensional space. However, such procedure rapidly becomes unmanageable when processing large datasets. In order to alleviate computational requirements, β-diversity maps are currently produced in two steps. First, ordination aiming at defining the projected 3D space from a dissimilarity matrix is performed on a limited subset of windows. Then, the dissimilarity between each window in the image and the subset of windows used in ordination is computed. Finally, the coordinates of each window in the resulting 3D space are defined based on the coordinates of the three most similar windows used during ordination, using inverse weighted distance. Non-metric multidimensional scaling (NMDS) was used as ordination method used by Féret and Asner (2014). However, NMDS is also particularly computer intensive compared to eigenanalyses such as principal coordinate analysis (PCoA) (Legendre, 1998). Both NMDS and PCoA are available in biodivMapR. More dissimilarity metrics will be implemented in future versions of the package.
2.8 | Computation of α- and β-diversity for a set of plots

Finally, the function `diversity_from_plots` computes diversity indices over each spatial polygon of a shapefile of plots, if available, in order to compare field inventories with diversity indices estimated from remotely-sensed images. The biodiversity indices are then directly computed from the `SpectralSpecies` raster, in order to keep the exact contour of the plot in the native spatial resolution of the image. The resulting diversity indices (α-diversity indices and BC matrix) are then saved in csv format in a dedicated 'VALIDATION' subdirectory.

**FIGURE 2** Example of Sentinel-2 image subset acquired over tropical forest in Cameroon and field plots with corresponding vegetation type

**FIGURE 3** α-Diversity (left) and β-diversity (right) maps produced with `biodivMapR` from the Sentinel-2 image subset displayed in Figure 2
EXAMPLE

The package includes a 10 × 10 km Sentinel-2 subset with associated plot network shapefiles. The plots were defined based on visual interpretation of the image, therefore values for α- and β-diversity metrics are not provided. Samples were selected over very diverse forest, moderately diverse forest, monodominant forest, and degraded forest/low vegetation close to the main roads (Figure 2). The script used to process the data is provided in Appendix S2 of SI.

Figure 3 shows the α- and β-diversity maps produced when processing this image and Appendix S3 of SI displays the eight components resulting from PCA. These results show that the different types of vegetation cover identified in Figure 2 can also be identified in the diversity maps.

Figure 4 summarizes α- and β-diversity in scatterplots and illustrates that the combination of the three components computed with PCoA allows proper differentiation among vegetation types:

- PCoA#1 allows differentiating medium and high diversity forests from low diversity forest and low vegetation, but does not discriminate medium and high diversity forests.
- PCoA#2 allows differentiating low diversity forest from medium/high diversity forests and low vegetation.
- PCoA#3 allows differentiating medium diversity forests from high diversity forests and low vegetation.

ACKNOWLEDGMENTS

We acknowledge financial support from Agence Nationale de la Recherche (BioCop project—ANR-17-CE32-0001) and TOSCA program grant of the French Space Agency (CNES) (HyperTropik/HyperBIO project). The authors have no conflicts of interest to declare.

AUTHORS’ CONTRIBUTIONS

J.-B.F. conceived the ideas and designed the methodology; J.-B.F. and F.B. created the package; J.-B.F. and F.B. wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

The current stable version of the package can be downloaded from the following URL: https://github.com/jbferet/biodivMapR

It can be installed with this command line:

devtools::install_github('https://github.com/jbferet/biodivMapR.git')

Zenodo ID: https://doi.org/10.5281/zenodo.3463499 (de Boissieu, 2019).

ORCID

Jean-Baptiste Féret https://orcid.org/0000-0002-0151-1334

Florian de Boissieu https://orcid.org/0000-0002-2185-9952

REFERENCES

Baldeck, C. A., Asner, G. P., Martin, R. E., Anderson, C. B., Knapp, D. E., Kelner, J. R., & Wright, S. J. (2015). Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. PLoS ONE, 10, e0118403. https://doi.org/10.1371/journal.pone.0118403

Bongalov, B., Burslem, D., Jucker, T., Thompson, S., Rosindell, J., Swinfield, T.,...Coomes, D. (2019). Reconciling the contribution of environmental and stochastic structuring of tropical forest diversity through the lens of imaging spectroscopy. Apollo - Univ. Camb. Repos. https://doi.org/10.17863/cam.41527

de Boissieu, F. (2019, September 27). jbferet/biodivMapR: v1.0.1 (Version v1.0.1). Zenodo. http://doi.org/10.5281/zenodo.3463499

ESA. (2019). L1C Data Quality Report (No. S2-PDGS-MPC-DQR). ESA.

Féret, J.-B., & Asner, G. P. (2014). Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. Ecological Applications, 24, 1289–1296. https://doi.org/10.1890/13-1824.1

Féret, J.-B., Rocchini, D., He, K. S., Nagendra, H., & Luque, S. (2017). Forest species mapping. In GOFC-GOLD & GEO BON (Eds.), A sourcebook of methods and procedures for monitoring essential biodiversity variables...
in tropical forests with remote sensing. Report version UN CBD COP-13, GOFC-GOLD Land Cover Project Office, Wageningen, The Netherlands: Wageningen University.

Hijmans, R. J. (2018). raster: Geographic Data Analysis and Modeling. Homolová, L., Malenovský, Z., Clevers, J. G. P. W., García-Santos, G., & Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. Ecological Complexity, 15, 1-16. https://doi.org/10.1016/j.ecocom.2013.06.003

IPBES. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In IPBES. IPBES Secretariat, Bonn, Germany.

Legendre, P., & Legendre, L. (1998). Numerical ecology (2nd ed.). Burlington, VT: Elsevier.

Leitão, P., Schwieder, M., Suess, S., Okujeni, A., Galvão, L., Linden, S., & Hostert, P. (2015). Monitoring natural ecosystem and ecological gradients: Perspectives with EnMAP. Remote Sensing, 7, 13098-13119. https://doi.org/10.3390/rs71013098

Palmer, M. W., Earls, P. G., Hoagland, B. W., White, P. S., & Wohlgemuth, T. (2002). Quantitative tools for perfecting species lists. Environmetrics, 13, 121-137. https://doi.org/10.1002/env.516

Pettorelli, N., Wegmann, M., Skidmore, A., Müller, S., Dawson, T. P., Fernandez, M., ... Geller, G. N. (2016). Framing the concept of satellite remote sensing essential biodiversity variables: Challenges and future directions. Remote Sensing in Ecology and Conservation, 2, 122-131. https://doi.org/10.1002/rse2.15

Rocchini, D., Balkenhol, N., Carter, G. A., Foody, G. M., Gillespie, T. W., He, K. S., ... Neteler, M. (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. Ecological Informatics, 5, 318–329. https://doi.org/10.1016/j.ecoinf.2010.06.001

Rocchini, D., Boyd, D. S., Féret, J.-B., Foody, G. M., He, K. S., Lausch, A., ... Pettorelli, N. (2016). Satellite remote sensing to monitor species diversity: Potential and pitfalls. Remote Sensing in Ecology and Conservation, 2, 25-36. https://doi.org/10.1002/rse2.9

Rocchini, D., Chiarucci, A., & Loiselle, S. A. (2004). Testing the spectral variation hypothesis by using satellite multispectral images. Acta Oecologica, 26, 117-120. https://doi.org/10.1016/j.actao.2004.03.008

Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., ... Nagendra, H. (2018). Measuring β-diversity by remote sensing: A challenge for biodiversity monitoring. Methods in Ecology and Evolution, 9, 1787-1798. https://doi.org/10.1111/2041-210X.12941

Somers, B., Asner, G. P., Martin, R. E., Anderson, C. B., Knapp, D. E., Wright, S. J., & Van De Kerchove, R. (2015). Mesoscale assessment of changes in tropical tree species richness across a bioclimatic gradient in Panama using airborne imaging spectroscopy. Remote Sensing of Environment, 167, 111-120. https://doi.org/10.1016/j.rse.2015.04.016

Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. New Phytologist, 186, 795–816. https://doi.org/10.1111/j.1469-8137.2010.03284.x

Vaglio Laurin, G., Chan, J.-C.-W., Chen, Q., Lindsell, J. A., Coomes, D. A., Guerriera, L., ... Valentini, R. (2014). Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. PLoS ONE, 9, e97910. https://doi.org/10.1371/journal.pone.0097910

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.