The segmetric Package: Metrics for Assessing Segmentation Accuracy for Geospatial Data

Rolf Simoes, Alber Sanchez, Michelle C. A. Picoli and Patrick Meyfroidt

Abstract Segmentation methods are a valuable tool for exploring spatial data by identifying objects based on images' features. However, proper segmentation assessment is critical to improve results and tune segmentation algorithms. Usually, various metrics are used to inform different types of errors that dominate the results. We describe a new R package, **segmetric**, for assessing and analyzing the geospatial segmentation of satellite images. This package unifies code and knowledge spread across programming languages and scientific journals to provide a variety of supervised segmentation metrics available in the literature. It also allows users to create their own metrics to evaluate the accuracy of segmented objects based on reference polygons. We hope this package helps to fulfill some of the needs of the R community that works with Earth Observation data.

Introduction

Earth Observation aims to collect data regarding the Earth's systems at several spatio-temporal resolutions. These data allow scientists to understand Earth's processes, such as greenhouse gas emissions and land cover change. Segmentation is among the most used unsupervised processing methods to extract information from satellite imagery (Hossain and Chen, 2019). Segmentation is the process by which we extract objects based on features in images. This process consists of delineating groups of adjacent pixels with similar characteristics such as intensity, color, and texture. Numerous segmentation algorithms are available in the Remote Sensing scientific literature (Kotaridis and Lazaridou, 2021).

Assessing segmentation results is harder than it looks due to different types of errors. To assess these errors we can use two approaches: unsupervised and supervised quality metrics (Costa et al., 2018). Unsupervised metrics measure segment properties such as their inter-class heterogeneity and intra-class homogeneity (Jozdani and Chen, 2020; Zhang et al., 2008). On the other hand, supervised metrics compare segments to reference data measuring their similarity or discrepancy in terms of under and over-segmentation (Clinton et al., 2010). Undersegmentation occurs when the segmentation algorithm fails to separate a contiguous pixel group while oversegmentation is the opposite, that is, the segmentation algorithm unnecessarily splits a pixel group (Costa et al., 2018). Both under and over-segmentation come with assessment metrics that target segments' characteristics such as area, shape, and position.

One significant challenge in the domain of Earth Observation data is the scarcity of software tools specifically designed for segmentation assessment. While several packages such as **imageseg** (Niedballa et al., 2022), **ExpImage** (Azevedo, 2022), **SuperpixelImageSegmentation** (Mouselimis, 2022a), **OpenImageR** (Mouselimis, 2022b), and **image.Otsu** (Wijffels, 2020) enable users to segment images but, when provided, they offer a limited set of facilities to assess accuracy segmentation and some of them are not necessarily tailored to the needs of Earth Observation data or applications. This often requires users to write their code for this purpose, which can be time-consuming and unrelated to their primary research goals.

In this paper, we introduce the **segmetric** package, which addresses the lack of R tools for assessing segmentation of Earth Observation data and provides a coherent set of metrics that can be used to compare and contrast different assessment methods for evaluating segmentation. Additionally, **segmetric** provides innovative visualization tools to assist qualitative spatial analysis as well as metrics that can be used to tune and assess segmentation algorithms.

Supervised segmentation metrics

Supervised metrics use reference data to assess segmentation accuracy. These metrics are grouped into two categories: geometric, which uses the geometry of objects and polygons to determine the similarity between them, and non-geometric, which uses instead the objects' contents (on the land use and land cover classes associated with the objects) (Costa et al., 2018).

segmetric focuses on geometric methods. In this category, supervised segmentation metrics take two sets of polygons as inputs. The first is the results of a segmentation algorithm ($Y = \{y_j : j = 1, ..., m\}$) and the second are the reference polygons ($X = \{x_i : i = 1, ..., n\}$) [1]. Implemented metrics are defined by considering different subsets of X and Y to compute its values. The subsets of the Y used to compute metrics for each reference i are defined as follows:

- $\tilde{Y}_i \subset Y$, where $\tilde{Y}_i = \{y_j : \text{area}(\text{x}_\text{i} \cap \text{y}_\text{j}) \neq 0\}$
- $Y'_i \subset Y$, where $Y'_i = \{y_j : max(\text{area}(\text{x}_\text{i} \cap \text{y}_\text{j}))\}$
- $Ya_i \subset \tilde{Y}_i$, where $Ya_i = \{y_j : \text{centroid}(\text{x}_\text{i}) \text{ in } \text{y}_\text{j}\}$
- $Yb_i \subset \tilde{Y}_i$, where $Yb_i = \{y_j : \text{centroid}(\text{y}_\text{j}) \text{ in } \text{x}_\text{i}\}$
- $Yc_i \subset \tilde{Y}_i$, where $Yc_i = \{y_j : \text{area}(\text{x}_\text{i} \cap \text{y}_\text{j})/\text{area}(\text{y}_\text{j}) > 0.5\}$
- $Yd_i \subset \tilde{Y}_i$, where $Yd_i = \{y_j : \text{area}(\text{x}_\text{i} \cap \text{y}_\text{j})/\text{area}(\text{x}_\text{i}) > 0.5\}$
- Y^*_i, where $Y^*_i = Ya_i \cup Yb_i \cup Yc_i \cup Yc_i$
- Ycd_i, where $Ycd_i = Yc_i \cup Yd_i$
- $Ye_i \subset \tilde{Y}_i$, where $Ye_i = \{y_j : \text{area}(\text{x}_\text{i} \cap \text{y}_\text{j})/\text{area}(\text{y}_\text{j}) = 1\}$
- $Yf_i \subset \tilde{Y}_i$, where $Yf_i = \{y_j : \text{area}(\text{x}_\text{i} \cap \text{y}_\text{j})/\text{area}(\text{y}_\text{j}) > 0.55\}$
- $Yg_i \subset \tilde{Y}_i$, where $Yg_i = \{y_j : \text{area}(\text{x}_\text{i} \cap \text{y}_\text{j})/\text{area}(\text{y}_\text{j}) > 0.75\}$

Likewise, subsets of X used to compute metrics for each segment j are defined as:

- $\tilde{X}_j \subset X$, where $\tilde{X}_j = \{x_i : \text{area}(\text{y}_\text{j} \cap \text{x}_\text{i}) \neq 0\}$
- $X'_j \subset X$, where $X'_j = \{x_i : max(\text{area}(\text{y}_\text{j} \cap \text{x}_\text{i}))\}$

To illustrate these subsets definition, we depict some of them in Figure 1. Subsets contains all elements for which a metric value has to be computed. To obtain a single metric value, a summary function can be applied on all values, typically a mean or a weighted mean. The range of possible values can differ from metric to metric. Also, the optimal value varies for each metric. Table 1 lists all implemented metrics in **segmetric**, their ranges and optimal values. The corresponding subsets used to compute values are shown in their formulas definition.

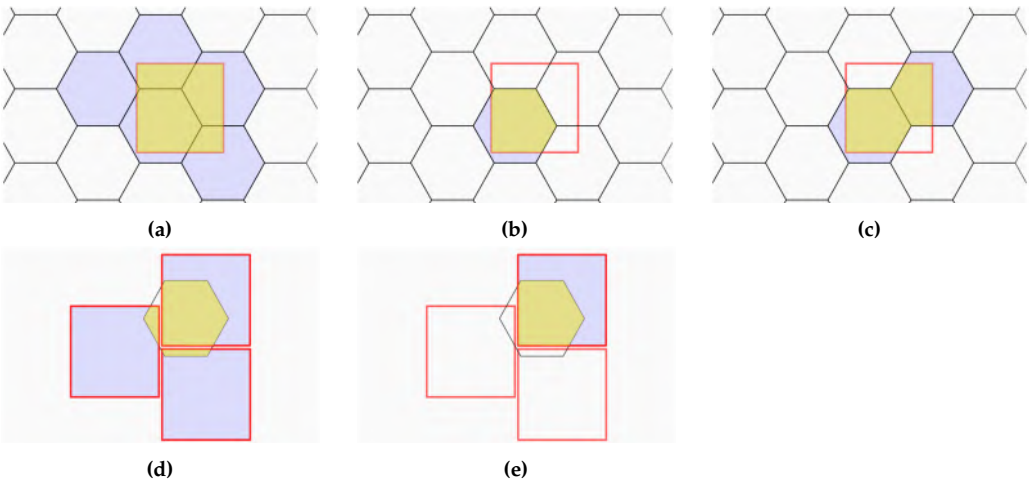

Figure 1: Subsets are used to compute segmentation metrics. The red squares and black hexagons represent reference and segmented polygons, respectively. (a) $\tilde{Y}$ is made of segmentation polygons overlapping any reference polygon. (b) Y' is made of the segmentation polygon overlapping the most of a reference polygon. (c) Y^* is made of polygons in the segmentation that either their centroids fall inside the reference, or they cover or overlap more than half of the reference polygon, or the reference polygon centroid falls inside them. (d) $\tilde{X}$ is made of the reference polygons overlapping any segmentation polygon. (e) X' is made of the reference polygons, which overlap the most with the segmentation polygons. Yellow represents the intersection between the references and segments included in a subset. The magenta areas are excluded from the reference-segmentation intersection but included in the subset. Adapted from Jozdani and Chen (2020).

[1] We are following the notation used by Clinton et al. (2010) and Costa et al. (2018)

Table 1: Metrics implemented

Metric	Range	Opt.	References
$OS1_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(x_i)},\ y_j \in Y^*{}_i$	$[0,1]$	0	Clinton et al. (2010)
$OS2_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(x_i)},\ y_j \in Y'{}_i$	$[0,1]$	0	Persello and Bruzzone (2010)
$OS3_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(x_i)},\ y_j \in Ycd_i$	$[0,1]$	0	Yang et al. (2014)
$US1_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(y_i)},\ y_j \in Y^*{}_i$	$[0,1]$	0	Clinton et al. (2010)
$US2_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(y_i)},\ y_j \in Y'{}_i$	$[0,1]$	0	Persello and Bruzzone (2010)
$US3_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(y_i)},\ y_j \in Ycd_i$	$[0,1]$	0	Yang et al. (2014)
$AFI_{ij} = \frac{area(x_i) - area(y_j)}{area(x_i)},\ y_j \in Y'{}_i$	$(-\infty,1]$	0	Lucieer and Stein (2002); Clinton et al. (2010)
$QR_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(x_i \cup y_j)},\ y_j \in Y^*{}_i$	$[0,1]$	0	Weidner (2008); Clinton et al. (2010)
$D_{ij} = \sqrt{\frac{OS_{ij}^2 + US_{ij}^2}{2}}$	$[0,1]$	0	Levine and Nazif (1982); Clinton et al. (2010)
$precision_{ij} = \frac{area(x_i \cap y_j)}{area(y_i)},\ y_j \in Y'{}_i$	$[0,1]$	1	van Rijsbergen (1979); Zhang et al. (2015)
$recall_{ij} = 1 - \frac{area(x_i \cap y_j)}{area(x_i)},\ y_j \in Y'{}_i$	$[0,1]$	1	van Rijsbergen (1979); Zhang et al. (2015)
$UMerging_{ij} = \frac{area(x_i) - area(x_i \cap y_j)}{area(x_i)},\ y_j \in Y^*{}_i$	$[0,1]$	0	Levine and Nazif (1982); Clinton et al. (2010)
$OMerging_{ij} = \frac{area(y_j) - area(x_i \cap y_j)}{area(x_i)},\ y_j \in Y^*{}_i$	$[0,\infty)$	0	Levine and Nazif (1982); Clinton et al. (2010)
$M_{ij} = \sqrt{\frac{area(x_i \cap y_j)^2}{area(x_i)area(y_j)}},\ y_j \in Y_i'$	$[0,1]$	1	Janssen and Molenaar (1995); Feitosa et al. (2010)
$E_{ij} = \frac{area(y_j) - area(x_i \cap y_j)}{area(y_i)} \times 100,\ x_i \in X'{}_j$	$[0,100]$	0	Carleer et al. (2005)
$RAsub_{ij} = \frac{area(x_i \cap y_j)}{area(x_i)},\ y_j \in \tilde{Y}_i$	$[0,1]$	1	Moller et al. (2007); Clinton et al. (2010)
$RAsuper_{ij} = \frac{area(x_i \cap y_j)}{area(y_i)},\ y_j \in \tilde{Y}_i$	$[0,1]$	1	Moller et al. (2007); Clinton et al. (2010)
$PI_i = \sum_{j=1}^{m} \frac{area(x_i \cap y_j)^2}{area(x_i)area(y_i)},\ y_j \in \tilde{Y}_i$	$[0,1]$	1	Van Coillie et al. (2008)
$Fitness_{ij} = \frac{area(x_i) + area(y_i) - 2 \times area(x_i \cap y_j)}{area(y_i)},\ x_i \in X_i'$	$[0,\infty)$	0	Costa et al. (2008)
$ED3_{ij} = \sqrt{\frac{OS3_{ij}^2 + US3_{ij}^2)}{2}}$	$[0,1]$	0	Yang et al. (2014)
$F\text{-}measure_{ij}{}^* = \frac{1}{\frac{\alpha}{precision} + \frac{(1-\alpha)}{recall}}$	$[0,1]$	1	van Rijsbergen (1979); Zhang et al. (2015)
$IoU_{ij} = \frac{area(x_i \cap y_j)}{area(x_i \cup y_j)},\ y_j \in Y'{}_i$	$[0,1]$	1	Jaccard (1912); Rezatofighi et al. (2019)
$SimSize_{ij} = \frac{\min(area(x_i), area(y_j))}{\min(area(x_i), area(y_j))},\ y_j \in Y^*{}_i$	$[0,1]$	1	Zhan et al. (2005)
$qLoc_{ij} = dist(centroid(x_i), centroid(y_j)),\ y_j \in Y^*{}_i$	$[0,\infty)$	0	Zhan et al. (2005)
$RPsub_{ij} = dist(centroid(x_i), centroid(y_j)),\ y_j \in \tilde{Y}_i$	$[0,\infty)$	0	Moller et al. (2007); Clinton et al. (2010)
$RPsuper_{ij} = \frac{dist(centroid(x_i), centroid(y_j))}{\max_j(dist(centroid(x_i), centroid(y_j)))},\ y_j \in Y^*{}_i$	$[0,1]$	0	Moller et al. (2007); Clinton et al. (2010)
$OI2_i = \max_j \left(\frac{area(x_i \cap y_j)}{area(x_i)} * \frac{area(x_i \cap y_j)}{area(y_j)} \right),\ y_j \in \tilde{Y}_i$	$[0,1]$	1	Yang et al. (2017)
$Dice_i = \frac{2 * area(x_i \cap y_j)}{area(x_i) + area(y_j)},\ y_j \in Y'{}_i$	$[0,1]$	1	Dice (1945)

* It takes the optional weight argument $\alpha \in [0,1]$ (the default is 0.5).

Metrics could be computed from scratch using subsets or by combining other metrics. Examples of the first type of metrics include: Oversegmentation (OS), Undersegmentation (US), Area Fit Index (AFI), Quality Rate (QR), Precision, Recall, Undermerging (UMerging), Overmerging (OMerging), Match (M), Evaluation measure (E), Relative area (RAsub and RAsuper), Purity Index (PI), and Fitness Function (Fitness). Examples of metrics computed by combining other metrics include: Index D (D), Euclidean Distance (ED3), and F-measure (F_measure). Some of these metrics are not intended to be summarized such as Relative position (RPsub and RPsuper). In that cases, when users try to summarize them using `summary()` a warning is thrown.

The segmetric package

Installation

The stable release of **segmetric** package can be installed from CRAN, using:

```
install.packages("segmetric")
```

Computing metrics

segmetric depends on the **sf** package (Pebesma, 2018) to open and manipulate these data sets. **sf** is an implementation of a standard issued by the Open Geospatial Consortium (OGC, 2011), which was further formalized in ISO 19125-1 (2004). This standard defines a common way to store and access spatial data in the context of geographic information systems.

To start with **segmetric**, users should create a `segmetric` object using `sm_read(ref_sf,seg_sf)` passing to it a reference spatial data set and a segmentation spatial data set. The parameters `ref_sf` and `seg_sf` should be either `sf` objects or paths to a supported file vector format (e.g., '`shapefile`').

```
library(segmetric)

# load example data sets
data("sample_ref_sf", package = "segmetric")
data("sample_seg_sf", package = "segmetric")

# create a segmetric object
m <- sm_read(ref_sf = sample_ref_sf, seg_sf = sample_seg_sf)
```

To compute a metric, users should run the function `sm_compute(m,metric_id,...)`, where `m` is a `segmetric` object and `metric_id` is the identification of a metric in **segmetric**. Any extra parameter necessary to compute metrics can be informed using the `ellipsis` parameter. The list of available metrics can be obtained using `sm_list_metrics()` which returns a `character` vector listing all registered metrics.

The `sm_compute()` function can compute a set of metrics by passing a vector of values to the `metric_id` parameter or making a sequence of function calls using a `pipe` operator. The two examples below produce equivalent results:

```
# compute three metrics
sm_compute(m, c("AFI", "OS1", "US1"))

# compute the same three metrics as above
sm_compute(m, "AFI") %>%
  sm_compute("OS1") %>%
  sm_compute("US1")
```

Most metrics are computed by feature (i.e., by reference or segment). To summarize the values of a set of metrics, users can run the function `summary(object,...)`, which computes aggregated values for the metrics returned by `sm_compute()`.

```
# compute three metrics
sm_compute(m, c("AFI", "OS1", "US1")) %>%
  summary()
```

Once created, a `segmetric` object stores in the cache every computed subset. Further subset requests are retrieved from the cache, speeding up the computation.

How to extend segmetric

The **segmetric** package is extensible by providing functions to implement new metrics. To implement a new metric, users can use `sm_new_metric()` to create a new metric object and register it using `sm_reg_metric()` function. Users can type `?sm_reg_metric()` to find more details on how new metrics can be implemented. The following example implements the Jaccard index (Jaccard, 1912), also known as Intersection over Union (*IoU*) (Rezatofighi et al., 2019), which is defined between 0 and 1 (optimal):

```
# register 'IoU' metric
sm_reg_metric(
    metric_id = "IoU",
    entry = sm_new_metric(
        fn = function(m, s, ...) {
            # m is the metric object, s is the subset
            #  for IoU, s is equivalent to sm_yprime(m)
            sm_area(s) / sm_area(sm_subset_union(s))
        },
        fn_subset = sm_yprime,
        name = "Intersection over Union",
        optimal = 1,
        description = "Values from 0 to 1 (optimal)",
        reference = "Jaccard (1912); Rezatofighi et al. (2019)"
    )
)

# describes the 'IoU' metric
sm_desc_metric("IoU")
#> * IoU (Intersection over Union)
#>   Values from 0 to 1 (optimal)
#>   reference: Jaccard (1912); Rezatofighi et al. (2019)
```

Contributions to the package are welcome at GitHub [2] and more details on how to contribute can be find in segmetric home-page at https://michellepicoli.github.io/segmetric.

Package segmetric in action

The specific steps involved in a segmentation workflow can vary depending on researcher goals, and characteristics of the input data. Depending on task requirements, the workflow may also involve other steps. However, in general, a segmentation workflow typically includes (Figure 2):

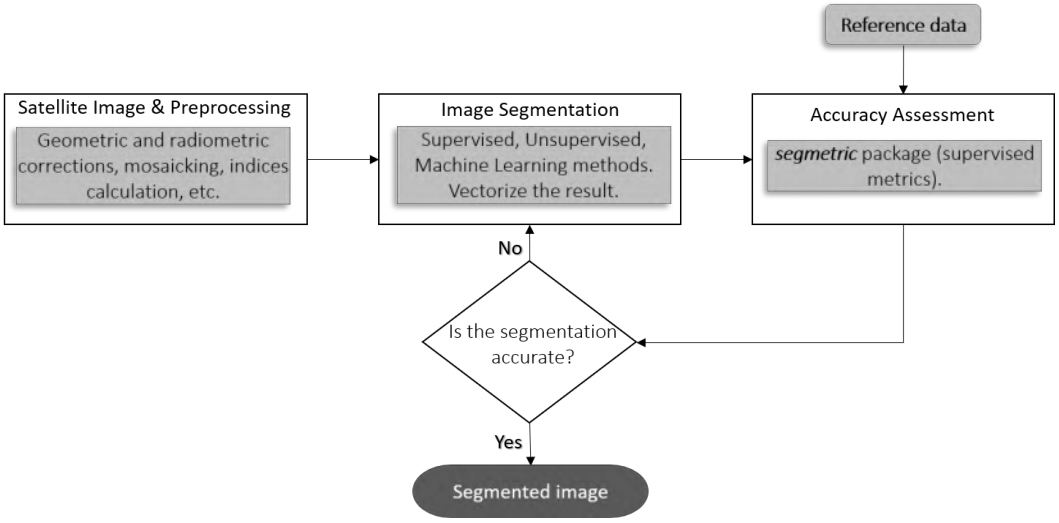

Figure 2: General steps of segmentation workflow.

[2] https://github.com/michellepicoli/segmetric

1. Preprocessing: This includes radiometric and geometric corrections, image mosaicking, cloud masking, indices computation, and texture extraction.
2. Image segmentation: This involves applying a specific segmentation algorithm based on the spectral bands, indices, and textures, to divide the image into different regions or objects.
3. Accuracy assessment: This involves the quality evaluation of the segmentation using metrics to determine the effectiveness of the segmentation algorithm and identify any potential areas for improvement. The **segmetric** can be used in this step to make a supervised evaluation on segmentation.
4. Repeat steps 2-3: Based on the evaluation results, users can repeat steps 2-3 by adjusting the parameters of the segmentation algorithm to improve its quality until the desired level of accuracy is achieved.

In what follows, we demonstrate an application of the **segmetric** package to assess several segmentation parameters and guide users to select the most accurate one.

Data

In agriculture studies, mapping characteristics such as the size and number of fields can provide important information about productivity and other food security, socioeconomic and environmental variables. To demonstrate **segmetric**, we used data on the Luís Eduardo Magalhães (LEM) municipality, west of Bahia state, Brazil. This municipality belongs to the Brazilian agricultural frontier known as MATOPIBA, which includes the states of Maranhão (MA), Tocantins (TO), Piauí (PI), and Bahia (BA) (Figure 3).

Figure 3: Study area in Luís Eduardo Magalhães municipality, west of Bahia state, Brazil (Google Earth imagery). Reference data (in red) was provided by Oldoni et al. (2020).

We used three PlanetScope images acquired on Feb 18, 2020, with a 3.7-meter resolution and four spectral bands (blue, green, red, and near-infrared). Radiometric and geometric corrections were applied to the image (level 3B) (Planet Team, 2017). The images were in the same projection (UTM zone 23S) and we mosaicked them.

We segmented the image applying a multi-resolution segmentation approach (Baatz and Schape, 2000). We tested four scale parameters (SP) to segment the image: 200, 500, 800, and 1000; shape parameter: 0.9; and compactness: 0.1. The resulting polygons were simplified using the Douglas-Peucker algorithm (Douglas and Peucker, 1973) (distance parameter: 10 meters) in QGIS software

(version 3.22.2). The Self-intersections were removed using SAGA's Polygon Self-Intersection tool (version 7.8.2). The final segmentation set is composed of polygons intersecting the reference data with an area-perimeter ratio above 25. We stored the segmentation results in the package.

The reference data set (ref_sf), provided by Oldoni et al. (2020), was collected in two fieldwork campaigns in March and August 2020. Oldoni et al. (2020) draw the field boundaries in-situ on top of images Sentinel-2, with a spatial resolution of 10 meters. **segmetric** includes only a portion of this data set. The spatial data sets can be loaded into R using **sf** objects. To create a **segmetric** object, use function `sm_read()`:

```
library(segmetric)

# load data sets
data("ref_sf", package = "segmetric")
data("seg200_sf", package = "segmetric")
data("seg500_sf", package = "segmetric")
data("seg800_sf", package = "segmetric")
data("seg1000_sf", package = "segmetric")

# create a segmetric object
m200 <- sm_read(ref_sf = ref_sf, seg_sf = seg200_sf)
m500 <- sm_read(ref_sf = ref_sf, seg_sf = seg500_sf)
m800 <- sm_read(ref_sf = ref_sf, seg_sf = seg800_sf)
m1000 <- sm_read(ref_sf = ref_sf, seg_sf = seg1000_sf)
```

Analysis

This analysis assesses four different segmentations with different Scale Parameters (SP) to verify which one fits better with the reference polygons. First, we visualize the reference polygons and the four segmentations individually using the plot() function (Figure 4).

```
# plot layers
plot(m200, layers = "ref_sf", plot_centroids = FALSE)
plot(m200, layers = "seg_sf", plot_centroids = FALSE)
plot(m500, layers = "seg_sf", plot_centroids = FALSE)
plot(m800, layers = "seg_sf", plot_centroids = FALSE)
plot(m1000, layers = "seg_sf", plot_centroids = FALSE)
```

The metrics available in the package can be consulted using the function `sm_list_metrics()`. In this example, the metrics chosen to evaluate the accuracy of the segmentations and verify the best value of the scale parameter were: Area Fit Index (Carleer et al., 2005), F-measure (van Rijsbergen, 1979) (Zhang et al., 2015), Quality Rate (Weidner, 2008) (Clinton et al., 2010), Oversegmentation (Clinton et al., 2010), and Undersegmentation (Clinton et al., 2010).

```
# compute all metrics
metrics <- c("QR", "F_measure", "IoU", "M", "OS2", "US2")
m200 <- sm_compute(m200, metrics)
m500 <- sm_compute(m500, metrics)
m800 <- sm_compute(m800, metrics)
m1000 <- sm_compute(m1000, metrics)

# results
summary(m200)
#>        QR F_measure       IoU         M       OS2       US2
#> 0.7394817 0.6988555 0.4988198 0.6569973 0.2948025 0.2585708

summary(m500)
#>         QR  F_measure        IoU          M        OS2        US2
#> 0.50380348 0.80671198 0.56837519 0.70140431 0.07982693 0.37207120

summary(m800)
#>         QR  F_measure        IoU          M        OS2        US2
#> 0.47487615 0.78764418 0.54923433 0.68297970 0.04300207 0.43014287
```

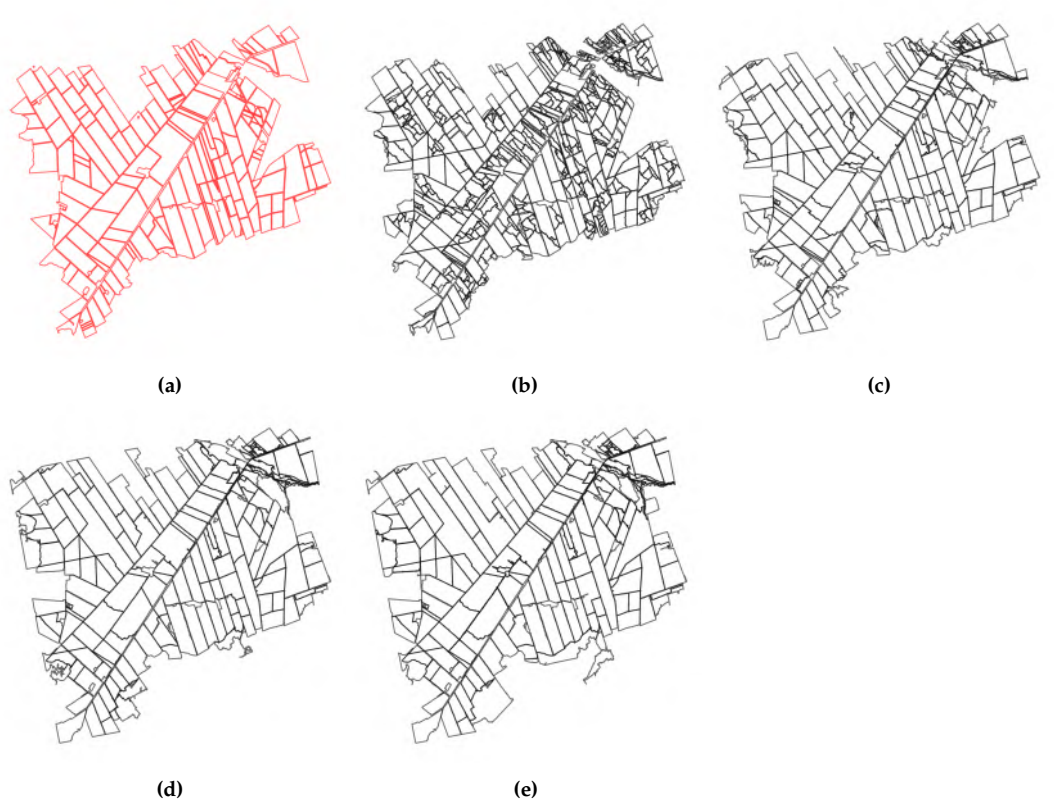

Figure 4: (a) reference polygons; (b) segmentation using SP = 200; (c) segmentation using SP = 500; (d) segmentation using SP = 800; (e) segmentation using SP = 1000.

```
summary(m1000)
#>         QR  F_measure        IoU          M        OS2        US2
#> 0.50311268 0.75742922 0.51745883 0.65548524 0.03679037 0.46524463
```

The computed metrics are presented in Table 2; the optimal value of QR, OS2, and US2 is 0, and 1 for F-measure, M, and IoU. These results indicate that segmentation using SP equal to 200 had the highest oversegmentation while using SP equal to 1000 had the highest undersegmentation. Observing the metrics F-measure, IoU, and M, we conclude that the best SP is 500.

Users must pay attention to which metric better fits their goals of accuracy assessment. For more information, we suggest the user consult comparative studies dedicated to geometric metrics such as Clinton et al. (2010), Räsänen et al. (2013), Yang et al. (2015), Costa et al. (2018), and Jozdani and Chen (2020).

Table 2: Accuracy metrics of Quality Rate (QR), F-measure, Intersection over Union (IoU), Match (M), Oversegmentation (OS2), and Undersegmentation (US2) for four segmentations with different Scale Parameters (SP).

	QR	F_measure	IoU	M	OS2	US2
seg 200	0.739	0.699	0.499	0.657	0.295	0.259
seg 500	0.504	0.807	0.568	0.701	0.080	0.372
seg 800	0.475	0.788	0.549	0.683	0.043	0.430
seg 1000	0.503	0.757	0.517	0.655	0.037	0.465

segmetric allows users to visualize subsets used to compute metrics. The example in Figure 5 shows the results of the function to plot the subset `Y_tilde` over the reference and the segmentation polygons (SP = 500). This allows analyzing the overlap between the reference and segmentation polygons visually.

```
plot(
```

```
    x = m500,
    type = "subset",
    subset_id = "Y_tilde",
    plot_centroids = FALSE,
    plot_legend = TRUE,
    extent = sm_seg(m500)
)
```

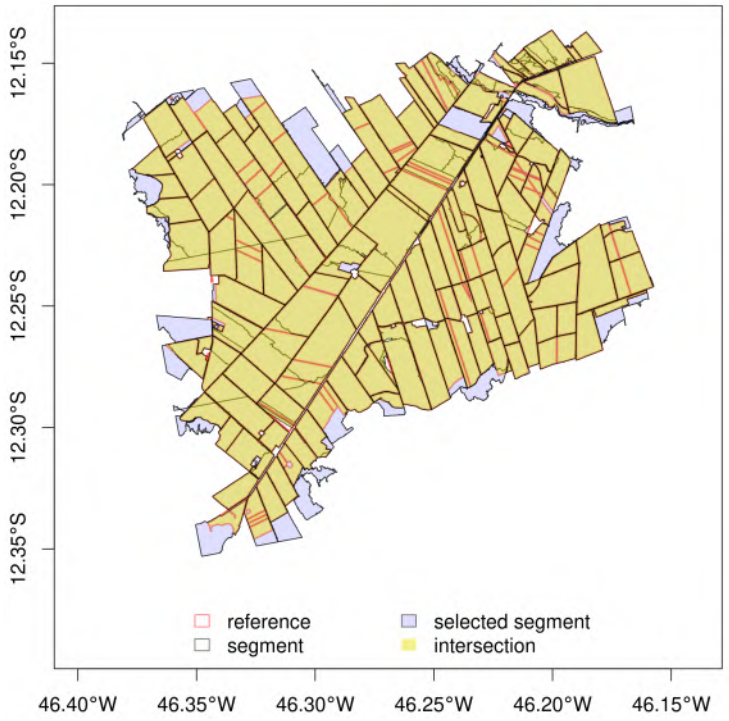

Figure 5: Overlapping between reference polygons and segmentation objects (SP = 500).

It is also possible to visualize the metrics for each segment in choropleth maps using the function:

```
plot(
    x = m500,
    type = "choropleth",
    metric_id = c("QR", "IoU", "M", "OS2", "US2"),
    break_style = "jenks",
    choropleth_palette = "RdYlBu",
    plot_centroids = FALSE
)
```

The legend bar of the choropleth maps is generated automatically, and users can provide some configuration such as palette and break options. The legend consistently uses the same color for the optimal metric value (for example, in Figure 6, blue is better while red is worse), except for those metrics in which the optimal value is in the middle of the color scale (e.g., AFI). The size and number of intervals in each color scale change accordingly to the metric values present in dataset. Users can choose the method to compute the intervals. To check available options use `?plot` and see `break_style` parameter.

Figure 6 presents the spatialized results of the calculated metrics. The F-measure metric was not plotted because it is a global metric with a single value for all objects. Figures 6a, d, and e show the similarity between the QR, OS, and US metrics results, for which the ideal value is zero. In the three plots of these metrics, it is noted that the objects with the best results (close to zero) are located southeast of the study area. The IoU and M metric maps (Figure 6b and c), for which the ideal value is 1, are also similar. It is also observed that in these two metrics, the objects located southwest of the study area have values close to 1. The figure shows differences in the number of objects plotted in each of the metrics, as the subsets used to calculate each of the metrics are different.

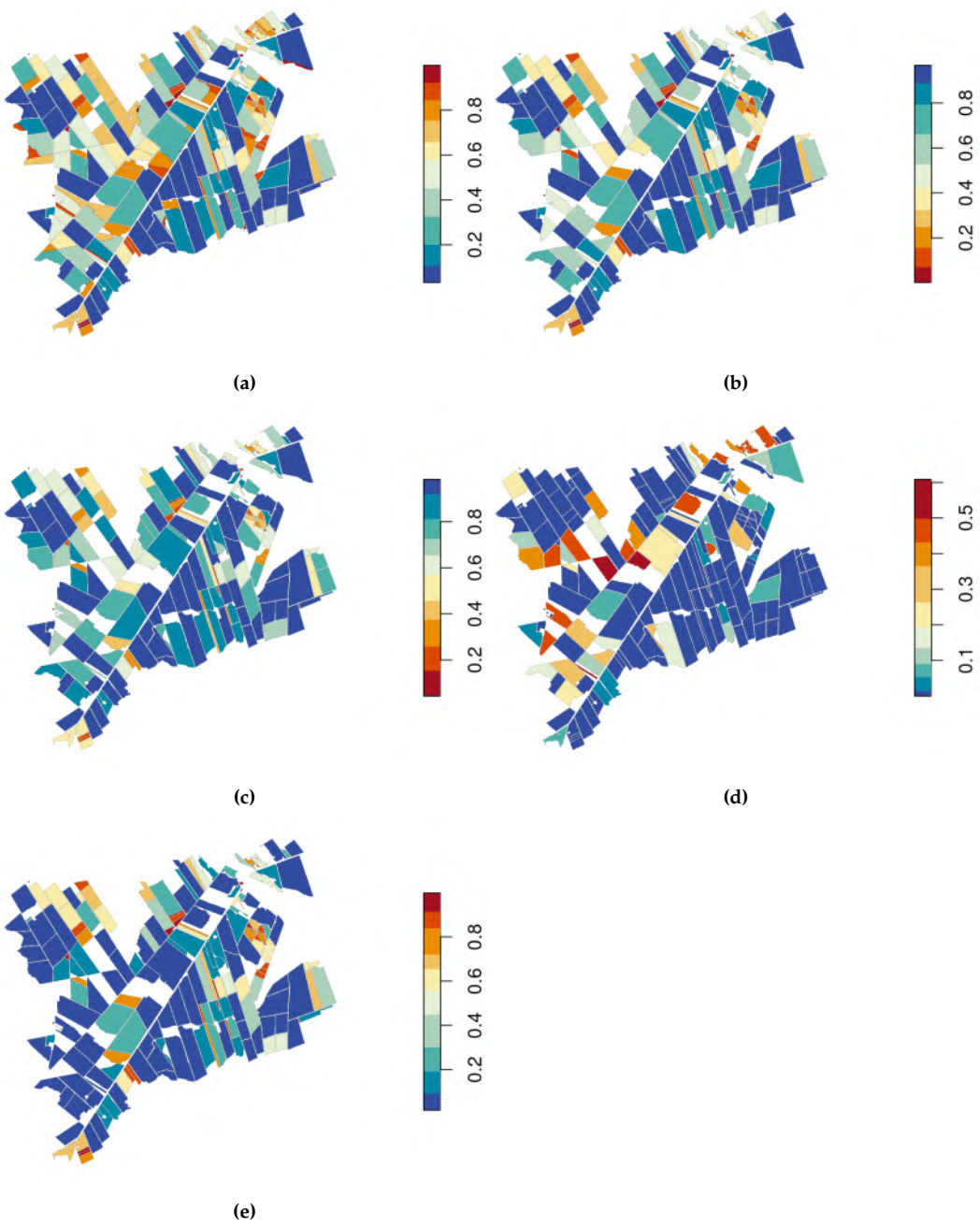

Figure 6: Spatial distribution of the metrics: (a) Quality Rate, (b) Intersection over Union, (c) Match, (d) Oversegmentation, and (e) Undersegmentation.

Summary

The segmetric package provides 28 metrics that can be used to evaluate different segmentation purposes. The package also offers innovative visualization options to assist qualitative spatial assessment, allowing diagnostics of the quality, issues, and potential biases of the segmentation. Plotting the segmented objects along their reference polygons and spatially visualizing the metrics help users to evaluate and improve segmentation procedures, select segmentation parameters, and decide on adequate validation metrics depending on their objectives.

To the extent of our knowledge, segmetric is the first available package in R that provides several supervised metrics based on reference polygons. segmetric also enables users to implement new metrics. In the future, we plan to add more supervised metrics and other ways to visualize metrics, and to use parallel processing to speed up computations.

Acknowledgments

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No 677140 MIDLAND).

Bibliography

A. M. Azevedo. *ExpImage: Tool For Analysis of Images in Experiments*, 2022. URL https://cran.r-project.org/web/packages/ExpImage. R package version 0.6.0. [p1]

M. Baatz and A. Schape. Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation. In J. Strobl, T. Blaschke, and G. Griesbner, editors, *Proceedings of Angewandte Geographische Informationsverarbeitung, XII*, pages 12–23, Salzburg, 2000. Herbert Wichmann Verlag. [p6]

A. Carleer, O. Debeir, and E. Wolff. Assessment of Very High Spatial Resolution Satellite Image Segmentations. *Photogrammetric Engineering & Remote Sensing*, 71(11):1285–1294, 2005. ISSN 0099-1112. URL https://doi.org/10.14358/PERS.71.11.1285. [p3, 7]

N. Clinton, A. Holt, J. Scarborough, L. Yan, and P. Gong. Accuracy Assessment Measures for Object-based Image Segmentation Goodness. *Photogrammetric Engineering & Remote Sensing*, 76(3):289–299, 2010. ISSN 0099-1112. URL https://doi.org/10.14358/PERS.76.3.289. [p1, 2, 3, 7, 8]

G. Costa, R. Feitosa, T. Cazes, and B. Feijó. *Genetic adaptation of segmentation parameters*, pages 679–695. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-77058-9. URL https://doi.org/10.1007/978-3-540-77058-9_37. [p3]

H. Costa, G. M. Foody, and D. S. Boyd. Supervised methods of image segmentation accuracy assessment in land cover mapping. *Remote Sensing of Environment*, 205:338–351, 2018. ISSN 0034-4257. URL https://doi.org/10.1016/j.rse.2017.11.024. [p1, 2, 8]

L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945. doi: https://doi.org/10.2307/1932409. [p3]

D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, dec 1973. ISSN 0317-7173. URL https://doi.org/10.3138/FM57-6770-U75U-7727. [p6]

R. Feitosa, R. Ferreira, C. Almeida, F. Camargo, and G. Costa. Similarity metrics for genetic adaptation of segmentation parameters. In *3rd International Conference on Geographic Object-Based Image Analysis (GEOBIA 2010)*, volume 29, Ghent, 2010. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. [p3]

M. D. Hossain and D. Chen. Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:115–134, 2019. ISSN 0924-2716. URL https://doi.org/https://doi.org/10.1016/j.isprsjprs.2019.02.009. [p1]

ISO 19125-1. Geographic information - Simple feature access - Part 1: Common architecture. Technical report, International Standard Organization, 2004. URL https://www.iso.org/standard/40114.html. [p4]

P. Jaccard. The Distribution of the Flora in the Alpine Zone. *New Phytologist*, 11(2):37–50, feb 1912. ISSN 0028-646X. URL https://doi.org/10.1111/j.1469-8137.1912.tb05611.x. [p3, 5]

L. Janssen and M. Molenaar. Terrain objects, their dynamics and their monitoring by the integration of GIS and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):749–758, 1995. URL https://doi.org/10.1109/36.387590. [p3]

S. Jozdani and D. Chen. On the versatility of popular and recently proposed supervised evaluation metrics for segmentation quality of remotely sensed images: An experimental case study of building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160(November 2019):275–290, feb 2020. ISSN 09242716. URL https://doi.org/10.1016/j.isprsjprs.2020.01.002. [p1, 2, 8]

I. Kotaridis and M. Lazaridou. Remote sensing image segmentation advances: A meta-analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:309–322, mar 2021. ISSN 09242716. URL https://doi.org/10.1016/j.isprsjprs.2021.01.020. [p1]

M. D. Levine and A. M. Nazif. *An experimental rule based for testing low level segmentation strategies*. Academic Press, New York, 1982. [p3]

A. Lucieer and A. Stein. Existential uncertainty of spatial objects segmented from satellite sensor imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2518–2521, 2002. URL https://doi.org/10.1109/TGRS.2002.805072. [p3]

M. Moller, L. Lymburner, and M. Volk. The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 9(3): 311–321, 2007. ISSN 0303-2434. URL https://doi.org/doi.org/10.1016/j.jag.2006.10.002. [p3]

L. Mouselimis. *SuperpixelImageSegmentation: Image Segmentation using Superpixels, Affinity Propagation and Kmeans Clustering*, 2022a. URL https://CRAN.R-project.org/package=SuperpixelImageSegmentation. R package version 1.0.5. [p1]

L. Mouselimis. *OpenImageR: An Image Processing Toolkit*, 2022b. URL https://CRAN.R-project.org/package=OpenImageR. R package version 1.2.7. [p1]

J. Niedballa, J. Axtner, T. F. Döbert, A. Tilker, A. Nguyen, S. T. Wong, C. Fiderer, M. Heurich, and A. Wilting. imageseg: An r package for deep learning-based image segmentation. *Methods in Ecology and Evolution*, 13(11):2363–2371, 2022. doi: https://doi.org/10.1111/2041-210X.13984. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13984. [p1]

OGC. Simple Feature Access-Part 1: Common Architecture. Technical report, Open Geospatial Consortium, 2011. URL http://www.opengeospatial.org/standards/sfa. [p4]

L. V. Oldoni, I. D. Sanches, M. C. A. Picoli, R. M. Covre, and J. G. Fronza. LEM+ dataset: For agricultural remote sensing applications. *Data in Brief*, 33:106553, 2020. ISSN 2352-3409. URL https://doi.org/10.1016/j.dib.2020.106553. [p6, 7]

E. Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1): 439–446, 2018. URL https://doi.org/10.32614/RJ-2018-009. [p4]

C. Persello and L. Bruzzone. A novel protocol for accuracy assessment in classification of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3):1232–1244, 2010. URL https://doi.org/10.1109/TGRS.2009.2029570. [p3]

Planet Team. Planet Application Program Interface: In Space for Life on Earth, 2017. URL https://api.planet.com. [p6]

A. Räsänen, A. Rusanen, M. Kuitunen, and A. Lensu. What makes segmentation good? A case study in boreal forest habitat mapping. *International Journal of Remote Sensing*, 34(23):8603–8627, 2013. URL https://doi.org/10.1080/01431161.2013.845318. [p8]

H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019. URL https://doi.org/10.1109/CVPR.2019.00075. [p3, 5]

F. Van Coillie, L. Verbeke, and R. De Wulf. *Semi-automated forest stand delineation using wavelet based segmentation of very high resolution optical imagery*, pages 237–256. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-77058-9. URL https://doi.org/10.1007/978-3-540-77058-9_13. [p3]

C. van Rijsbergen. *Information Retrieval*. Butterworths, 2nd edition, 1979. [p3, 7]

U. Weidner. Contribution to the assessment of segmentation quality for remote sensing applications. In *XXI International Society for Photogrammetry and Remote Sensing Congress (XXI ISPRS 2008)*, pages 479–484, Beijing, 2008. International Society for Photogrammetry and Remote Sensing. [p3, 7]

J. Wijffels. *image.Otsu: Otsu's Image Segmentation Method*, 2020. URL https://CRAN.R-project.org/package=image.Otsu. R package version 0.1. [p1]

J. Yang, P. Li, and Y. He. A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 94:13–24, 2014. ISSN 0924-2716. URL https://doi.org/10.1016/j.isprsjprs.2014.04.008. [p3]

J. Yang, Y. He, J. Caspersen, and T. Jones. A discrepancy measure for segmentation evaluation from the perspective of object recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101:186–192, 2015. ISSN 0924-2716. URL https://doi.org/10.1016/j.isprsjprs.2014.12.015. [p8]

J. Yang, Y. He, J. P. Caspersen, and T. A. Jones. Delineating Individual Tree Crowns in an Uneven-Aged, Mixed Broadleaf Forest Using Multispectral Watershed Segmentation and Multiscale Fitting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4):1390–1401, 2017. doi: https://doi.org/10.1109/JSTARS.2016.2638822. [p3]

Q. Zhan, M. Molenaar, K. Tempfli, and W. Shi. Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing*, 26(14):2953–2974, 2005. doi: https://doi.org/10.1080/01431160500057764. [p3]

H. Zhang, J. E. Fritts, and S. A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260–280, 2008. doi: https://doi.org/10.1016/j.cviu.2007.08.003. [p1]

X. Zhang, X. Feng, P. Xiao, G. He, and L. Zhu. Segmentation quality evaluation using region-based precision and recall measures for remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102:73–84, 2015. ISSN 0924-2716. URL https://doi.org/10.1016/j.isprsjprs.2015.01.009. [p3, 7]

Rolf Simoes
National Institute for Space Research (INPE)
Avenida dos Astronautas, 1758, 12227-010, Sao Jose dos Campos
Brazil
ORCiD: https://orcid.org/0000-0003-0953-4132
rolf.simoes@inpe.br

Alber Sanchez
National Institute for Space Research (INPE)
Avenida dos Astronautas, 1758, 12227-010, Sao Jose dos Campos
Brazil
ORCiD: https://orcid.org/0000-0001-7966-2880
alber.ipia@inpe.br

Michelle C. A. Picoli
Earth and Life Institute, UCLouvain
Place Louis Pasteur 3, 1348, Louvain-la-Neuve
Belgium
ORCiD: https://orcid.org/0000-0001-9855-2046
michelle.picoli@uclouvain.be

Patrick Meyfroidt
Earth and Life Institute, UCLouvain
Place Louis Pasteur 3, 1348, Louvain-la-Neuve
Belgium
ORCiD: https://orcid.org/0000-0002-1047-9794
patrick.meyfroidt@uclouvain.be