

Optimal parameters for delineating agricultural parcels from satellite images based on supervised Bayesian optimization

Gideon Okpoti Tetteh^{a,*}, Alexander Gocht^a, Christopher Conrad^b

^a Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany

^b Martin-Luther-University Halle-Wittenberg, Institute of Geosciences and Geography, 06099 Halle, Germany

ARTICLE INFO

Keywords:

Agricultural parcel delineation
Multiresolution segmentation
Bayesian optimization
Land parcel identification system
Empirical segmentation evaluation

ABSTRACT

Accurate spatial information of agricultural parcels is fundamental to any system used in monitoring greenhouse gas emissions, biodiversity developments, and nutrient loading in agriculture. The inefficiency of the traditional methods used in obtaining this information is increasingly paving the way for Remote Sensing (RS). The Multiresolution Segmentation (MRS) algorithm is a well-known method for segmenting objects from images. The quality of segmentation depends on the *a priori* knowledge of which scale, shape and compactness values to use. With each parameter taking a varied range of input values, this research developed an automated approach for identifying the optimal parameter set without testing all possible combinations. At the core of our approach is Bayesian optimization, which is a sequential model-based optimization (SMBO) method for maximizing or minimizing an objective function. We maximized the Jaccard index, which is a measure that indicates the similarity between segmented agricultural objects and their corresponding reference parcels. As the optimal parameter combination varies between different agricultural landscapes, they were determined at a grid resolution of 10 km. Mono-temporal Sentinel-2 images covering Lower Saxony in Germany were tiled to these grids and the optimal parameters were subsequently identified for each tiled grid. The optimal parameter combinations identified over the grids varied considerably, which indicated that a single parameter combination would have failed to achieve optimal segmentation. We found that the quality of segmentation correlated with the size of agricultural parcels. Under-segmentation was largely minimized but in areas with a predominant agricultural land-use, it was unavoidable. In agricultural parcels composed of heterogeneous pixels, over-segmentation was prevalent. Our approach outperformed other segmentation optimization methods existing in the literature.

1. Introduction

The increasing world population places enormous pressure on agricultural lands due to the growing demand for food. To meet this demand, natural habitats are being converted to farmlands, while existing farmlands are being intensively utilized (Dudley and Alexander, 2017). These conversions often lead to the destruction of biodiversity, high nutrient surpluses, and greenhouse gas emissions (Dudley and Alexander, 2017). The challenge then is to increase food production through sustainable agricultural management practices that leave minimal impact on the environment (Foley et al., 2011).

The foundation of any effective agricultural management scheme is accurate spatial information of all agricultural parcels. The most prominent agricultural parcel information system within the European Union (EU) is the Land Parcel Identification System (LPIS), which is a

spatial record of agricultural parcels declared by farmers (Taşdemir and Wirnhardt, 2012). It was established as part of the Common Agricultural Policy (CAP) framework to ensure that subsidies are correctly paid to farmers (Schmedtmann and Campagnolo, 2015). There have been suggestions by land managers to use LPIS as the foundation for developing sustainable agricultural schemes (Zielinski et al., 2008). Unfortunately, the LPIS has some drawbacks, which limits its use in an effective agricultural management scheme. Firstly, the LPIS does not record all agricultural parcels especially those with large grassland shares used for purposes like nature conservation and horse farming. This makes it difficult to monitor those parcels. Secondly, even though the LPIS is increasingly becoming available as open data in some EU countries, there is still restricted access in many countries including Germany. Thirdly, LPIS comes with a time lag, which makes in-season monitoring of parcels infeasible. In-season monitoring is critical to

* Corresponding author.

E-mail address: gideon.tetteh@thuenen.de (G.O. Tetteh).

<https://doi.org/10.1016/j.compag.2020.105696>

Received 9 April 2020; Received in revised form 30 July 2020; Accepted 3 August 2020

0168-1699/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

understanding any emerging threats to biodiversity on agricultural lands in near real-time so that timely action can be taken to deal with those threats. Additionally, using LPIS requires a lot of pre-processing effort due to different technical implementations between different regions and countries within the EU. These drawbacks inhibit various agricultural bodies including public institutions from making use of the LPIS as source data for developing real-world agronomic and environmental applications.

An automated system for generating a spatial database of all agricultural parcels could provide a solution to these drawbacks. With satellite data becoming more accessible, Remote Sensing (RS) presents the best means of obtaining cost-effective, accurate and up-to-date information of agricultural parcels due to its ability to obtain information over large areas with a high repetition rate (Atzberger, 2013). The launch and subsequent provision of free data by the Sentinel-2 satellites provide us a unique opportunity to obtain spatial information on agricultural parcels at a high spatial resolution over large geographical areas. The use of Sentinel-2 data for obtaining information on agricultural parcels is an active area of research as depicted by these studies (Belgiu and Csillik, 2018; Immitzer et al., 2016; Nasrallah et al., 2018; Watkins and Van Niekerk, 2019). The main limitation of those studies is that their respective methodologies were applied to small test areas and not geographic regions. One commonality amongst them is the use of object-based image analysis (OBIA), where spectrally similar pixels are grouped into homogenous objects through image segmentation and then the land-use type of each object is determined through object classification. A successful image segmentation process is the most critical step in this OBIA paradigm (Baatz and Schäpe, 2000; Benz et al., 2004).

Various algorithms have been developed to segment agricultural parcels from satellite images. The first category of algorithms is based on edge detection. They involve the extraction of edges and the subsequent linking of those edges to form boundaries. Turker and Kok (2013) and Ji (1996) applied edge-based methods to extract agricultural fields from SPOT and Landsat images respectively. The edge-based methods generate incomplete objects and often fail in landscapes where agricultural fields are small and boundaries are indistinct (Persello et al., 2019). Due to this limitation, region-based methods are more favored. Their underlying principle is that neighboring pixels with similar spectral values are merged into objects. Image clustering is one of the most simple and popular region-based methods. The Simple Linear Iterative Clustering (SLIC) algorithm developed by Achanta et al. (2012) was combined with supervised classification to segment agricultural fields from a WorldView-2 image by García-Pedrero et al. (2017). Nasrallah et al. (2018) used the mean shift clustering (Fukunaga and Hostetler, 1975) algorithm to delineate wheat fields from Sentinel-2 images. The main problem with clustering-based methods is that they often create very big objects, which do not follow the natural boundaries of image features. To deal with this problem, the third category of algorithms, a hybrid approach, is sometimes used. This approach involves the use of an edge extraction algorithm to produce an edge map, which is then given as an input to a region-growing algorithm. Using this approach, Li and Xiao (2007) and Yan and Roy (2014) respectively segmented crop fields from SPOT and Landsat images. The last category of algorithms, which is gaining traction in the RS world, is the use of Deep Neural Networks (DNN). A Fully Convolutional Network (FCN) called SegNet was employed by Persello et al. (2019) to identify initial agricultural boundaries from WorldView-3 images, which were later post-processed through a watershed transform and combinatorial grouping to obtain complete agricultural fields. Rieke (2017) adopted the Fully Convolutional Instance-aware Semantic Segmentation (FCIS) architecture of Li et al. (2016) for the segmentation of agricultural fields from a Sentinel-2 image. The computationally complex nature of DNNs puts them at a disadvantage for use in RS because they take a lot of time for model training and optimization (Kamilaris and Prenafeta-Boldú, 2018).

Therefore, they are mostly applied to small test areas as was done by Persello et al. (2019) and Rieke (2017).

Even though there are many segmentation algorithms of choice, the Multiresolution Segmentation (MRS) algorithm proposed by Baatz and Schäpe (2000) and implemented in eCognition Developer (Trimble Germany GmbH, 2019) is the most widely used segmentation algorithm as evidenced by Marpu et al. (2010) and Neubert et al. (2008). Many researchers (Belgiu and Csillik, 2018; Conrad et al., 2010; Lebourgeois et al., 2017; Peña-Barragán et al., 2011; Vogels et al., 2019) have applied the MRS algorithm for the delineation of agricultural parcels. MRS is a bottom-up region merging algorithm that starts with one-pixel objects and then subsequently merges neighboring objects into bigger objects where the change in the combined spectral and spatial heterogeneity is minimal (Benz et al., 2004). In the implementation in eCognition Developer, the three main parameters that control the output of the MRS algorithm are scale, shape, and compactness. Each of these parameters takes a varied range of input values, thereby yielding an infinite number of parameter combinations. Therefore, determining the optimal parameter combination is critical to achieving optimal segmentation results.

This research aims to develop an efficient approach to identify the optimal parameters needed to segment agricultural parcels using the MRS algorithm. The traditional approach to parameter optimization is the grid search method. Given any domain space of parameters, this method evaluates all possible parameter combinations using a given model and then returns the combination with the highest or lowest evaluation score as the optimal. As the number of parameters and elements in each parameter space increases, the computational time exponentially increases. This limitation was dealt with through the random search (Bergstra and Bengio, 2012) method. Using a smaller number of model evaluations within a shorter time frame, the random search method outperformed grid search (Bergstra and Bengio, 2012). However, these two methods are very inefficient in the process of identifying the optimal parameter because they do not consider the results of previous model evaluations before sampling new combinations, thereby wasting time on needless model evaluations. Overcoming these limitations requires the use of sequential model-based optimization (SMBO) (Bergstra et al., 2011). SMBO intuitively makes an informed prediction of which new combinations to test based on results from the previous model evaluations.

SMBO is a succinct formalism of Bayesian optimization (Dewancker et al., 2016). Bayesian optimization is used in globally optimizing black-box functions (Mockus, 2012) with unknown derivatives that take a long time to evaluate (Frazier, 2018). It outperforms the grid and random search methods (Bergstra et al., 2011; Snoek et al., 2012). Instead of directly solving a computationally expensive objective function, Bayesian optimization first constructs a surrogate model with prior information of the objective function. The surrogate model is initiated with some samples drawn from the domain space to obtain posterior information of the objective function. A new sample is automatically identified by maximizing an acquisition function over the posterior surrogate model. This new sample is evaluated with the objective function, and then the posterior surrogate model is updated with the result. This process is repeated until the maximum number of iterations or time allocation given by a user is reached (Dewancker et al., 2016; Shahriari et al., 2016). The sample from all the tests that minimized or maximized the surrogate model is returned as the optimal. The usage of Bayesian optimization for parameter optimization has become an active research area (Eggenberger et al., 2013). It has been used in solving optimization problems (Brochu et al., 2010; Shahriari et al., 2016) in various areas such as robotics (Lizotte, 2007), environmental monitoring (Marchant and Ramos, 2012), sensor networks (Osborne et al., 2010), and machine learning (Snoek et al., 2012; Thornton et al., 2013). To the best of our knowledge, there has not been any research geared towards the use of Bayesian optimization to optimize the parameters needed for segmenting satellite images, hence our

research is novel.

To apply Bayesian optimization, the objective function must take a parameter combination from any domain space as an input and then return an evaluation score. We derived this score by making use of empirical segmentation evaluation (Zhang, 1996), where the similarity between segmented agricultural parcels and their corresponding reference objects was numerically assessed through the Jaccard index. The combination with the highest Jaccard index is returned by the Bayesian optimization method as the optimal. We tested our approach on mono-temporal Sentinel-2 images covering Lower Saxony, which is a federal state in Germany. The result achieved was compared with other optimization methods based on the MRS algorithm.

The rest of the paper is structured as follows: the study area and the data are first described. Afterward, we explain the overall methodology including data preparation, development of our optimization approach, and its application to images in Lower Saxony. The results achieved are then discussed. We finish the paper by drawing some conclusions and pointing out further research directions.

2. Study area and data

We selected Lower Saxony (Fig. 1) as the study area because the Ministry of Food, Agriculture and Consumer Protection of Lower Saxony permitted us to use the LPIS as reference data. The coordinates of the map in Fig. 1 and all other maps in this paper are in UTM Zone 32 N (EPSG:32632). Most of its landmass is located in the temperate climate zone of Europe. The southeastern part is located in the continental climate zone. Apart from the southeastern part, where one can

locate the Harz mountain range, the terrain is relatively flat, making it suitable for farming. Various agricultural land-use types cover about 62% of its total land area of about 4,770,041 ha. The LPIS data is made up of 907,564 agricultural parcels. Based on this data, the most dominant agricultural land-use types, in order of percentage coverage, are grasslands (40%), summer cereals (23%), winter cereals (17%), potatoes (3%), winter rapeseed (3%), and sugar beet (2%). The size of the agricultural parcels ranges from as low as 0.1 ha to as high as 155 ha. The average parcel size is 3 ha. Even though the agricultural landscape is composed of heterogeneous parcel sizes, the minimum parcel size is large enough to be detected by the Sentinel-2 satellite. To ensure that the optimal segmentation parameters are representative of the wide range of parcel sizes, a square tile grid system made up of 10 km × 10 km tiles covering Lower Saxony was created. The total number of tiles came up to 562. Neighboring tiles have an overlap of 1 km. These tiles served as the basic unit for which the optimal MRS parameter combination had to be determined.

The Sentinel-2 images provided by the European Space Agency (ESA) were used for this research. Sentinel-2 is an optical satellite with thirteen spectral bands in the visible, near-infrared, and short-wave infrared regions of the electromagnetic spectrum. The spatial resolution ranges from 10 m to 60 m. The Level 1C images with a maximum cloud cover of 20% in May of 2018 were downloaded from the data repository of ESA. In May, winter crops are nearly at peak growth, while summer crops are just about shooting up. This makes it easier to differentiate and segment agricultural parcels, hence the choice of images in May. Fourteen Sentinel-2 images were downloaded to cover every part of Lower Saxony. For each tile, the first image that is cloud-free and non-

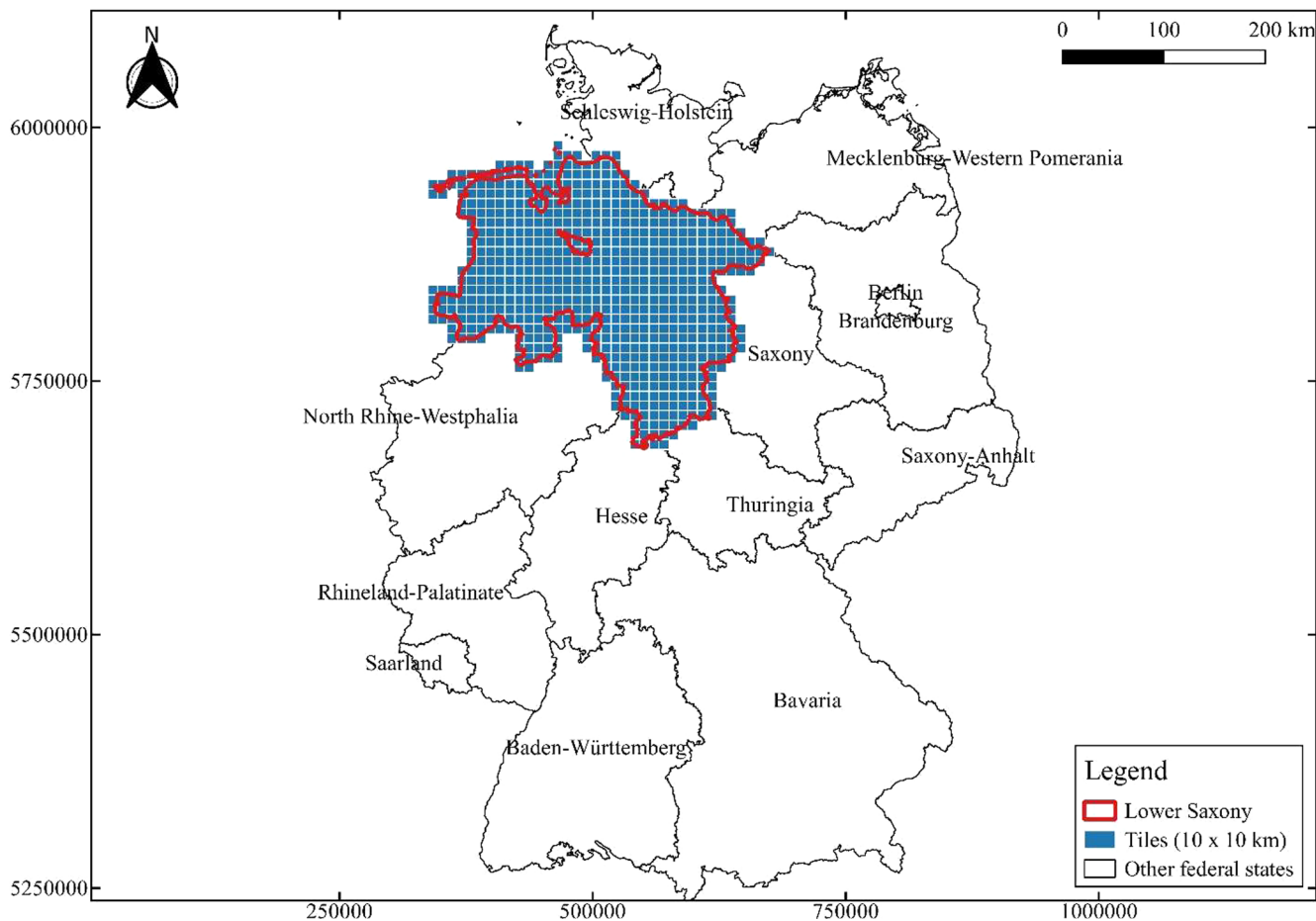


Fig. 1. The study area is Lower Saxony, Germany. This federal state was divided by a tile grid system made up of 10 km × 10 km tiles (blue polygons) numbering 562. For each tile, a Sentinel-2 image was extracted and used as input for segmentation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

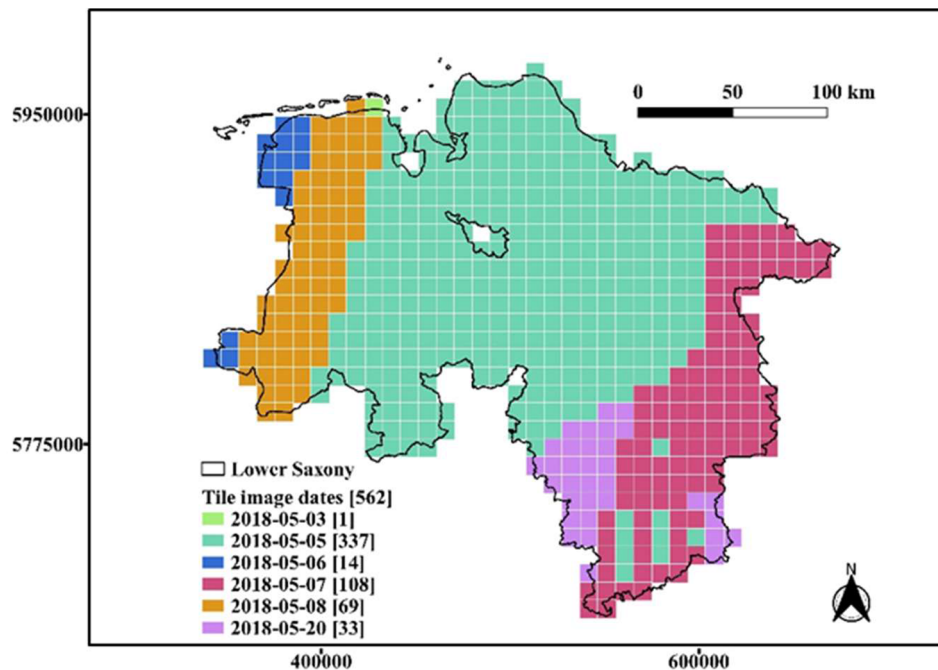


Fig. 2. A total of six cloud-free Sentinel-2 images covering Lower Saxony were used. The total number of tiles for each image acquisition date is shown as well. The majority of the tiles were captured by images from 5 May to 8 May 2018. The 5 May image covered most tiles.

defective was identified via visual inspection. The final image acquisition dates used in this research came up to six (Fig. 2).

Three auxiliary datasets were used to mask out non-agricultural areas from the Sentinel-2 images. The first dataset was the German Official Topographic Cartographic Information System (ATKIS) provided by the German Federal Agency for Cartography and Geodesy. ATKIS is a geographic database that captures the topography of Germany. The agricultural and non-agricultural vector layers covering Lower Saxony were extracted from this database. Out of the total agricultural land-cover area of about 2,936,292 ha, 65% are farmlands, 33% are grasslands, and the rest is composed of fruit plantations, tree nurseries, horticultural lands, and orchard meadows. The second one was the hydrological network dataset provided by the German Federal Institute of Hydrology, which contains all watercourses in Germany. Finally, linear features representing roads in Lower Saxony were downloaded from Open Street Map (OSM). The total length of roads is about 268,529 km. According to the OSM feature classification, there are 22 different types of roads in Lower Saxony. Majority of them are tracks (38%), followed by residential roads (16%), minor roads that link villages and hamlets (9%), service roads (7%), cycleways (6%), tertiary roads (6%), paths (5%), and the others.

3. Methodology

This section describes the development of our optimization approach and its application to image tiles in Lower Saxony, Germany. Our approach is a framework for identifying the optimal parameters needed for the segmentation of agricultural parcels and the actual delineation of those parcels from satellite images. Fig. 3 shows the general workflow.

3.1. Data preparation

Using the Sen2Cor (Main-Knorn et al., 2017) plugin in the Sentinel Application Platform (SNAP) of ESA, the Sentinel-2 Level 1C images were atmospherically and terrain corrected to obtain Bottom-Of-Atmosphere (BOA) Level 2A images. For each Level 2A image, only the visible (red, green, blue) and near-infrared bands were used. These four

bands have a spatial resolution of 10 m, unlike the other bands that have a lower spatial resolution (≥ 20 m). A higher spatial resolution leads to a higher segmentation quality (Mesner and Oštir, 2014). We stacked the four bands together into an image. Therefore, this image, which is henceforth named S2-VNIR, has a spatial resolution of 10 m.

The LPIS and ATKIS datasets sometimes contain sliver polygons. Those polygons were deleted based on their perimeter-to-area ratio. This deletion was more significant in the LPIS as the total number of parcels reduced to 853,892. The motorway line features in the OSM were buffered by 10 m, while the other line features like tracks and residential roads were buffered by 5 m to obtain polygon features. A buffer distance of 10 m was applied to all the watercourse line features. All the buffer distances were empirically determined by overlaying the line features on different images and testing various buffer distances such that the area of any resultant polygon was large enough to contain at least one pixel. Out of the various buffer distances we tested, the aforementioned buffer distances we used in this study were identified as optimal because they resulted in polygons with minimal encroachment on the boundaries of neighboring agricultural parcels. A no-data mask layer was created by merging the non-agricultural vector layer with the OSM and watercourse polygons. All pixels in each S2-VNIR data that intersected the no-data mask layer were masked out. Finally, each S2-VNIR data was clipped to the tile grid it spatially covered. These clipped S2-VNIR datasets were used for further processing. Fig. 4a shows one S2-VNIR image containing both agricultural and non-agricultural areas, while Fig. 4b shows the same image with all non-agricultural areas removed using the no-data mask layer.

3.2. Optimization design and application

At the heart of our optimization approach are image segmentation, supervised evaluation of segmentation quality, and Bayesian optimization.

3.2.1. Image segmentation

The Multiresolution Segmentation (MRS) algorithm as implemented in eCognition Developer 9.5.0 was used for image segmentation. MRS is a pair-wise merging process that starts with single-pixel objects well

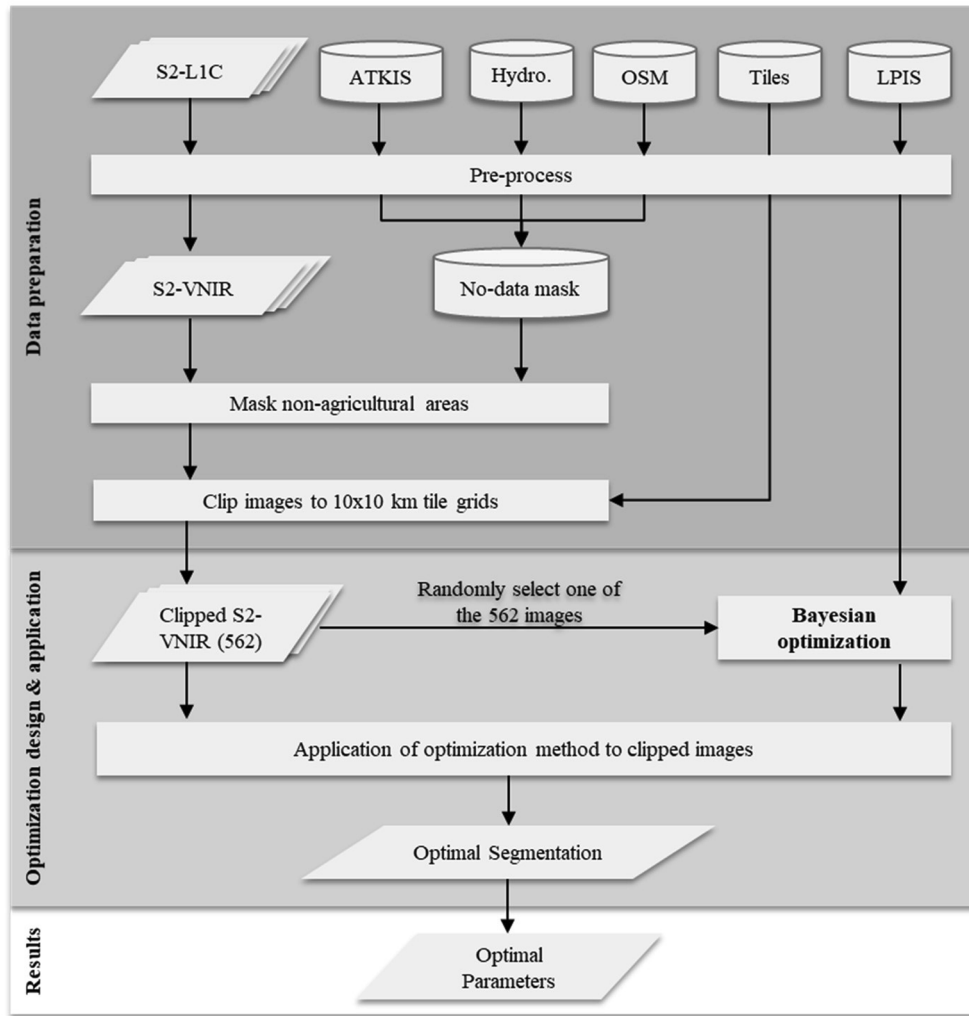


Fig. 3. The general workflow that was followed to determine the optimal parameter combinations for Lower Saxony.

distributed over an image. For each pixel object, a neighboring pixel object is found such that the change in heterogeneity between them is minimal. The heterogeneity of each object is computed as a function of the color and shape of that object. Where the change in heterogeneity is

minimal, the two objects are merged into a bigger object. Each object is handled once per loop cycle. This merging process stops as soon as the number of pixels in any object exceeds a user-given threshold value.

The three parameters that influence the segmentation outcome are

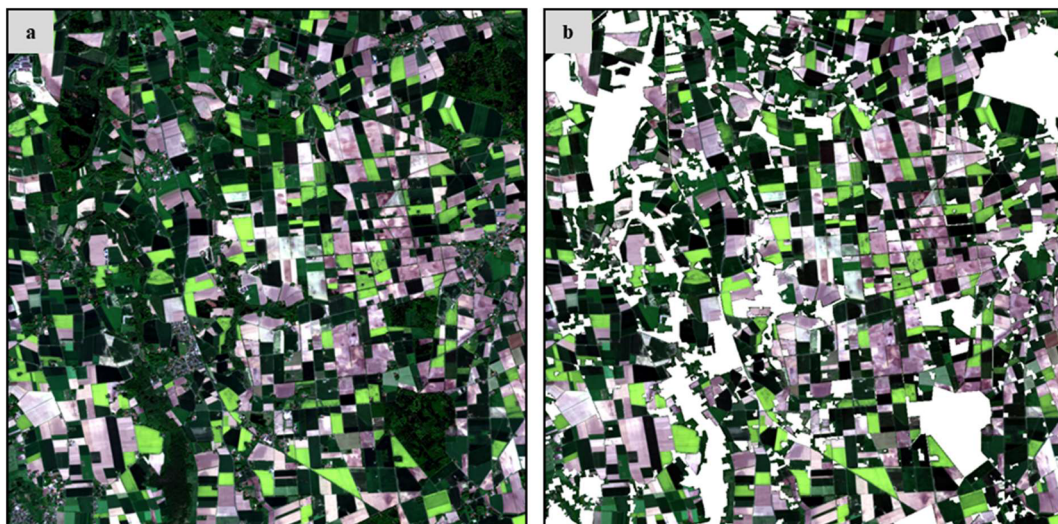


Fig. 4. (a) A non-masked S2-VNIR image. (b) The same image with non-agricultural areas, watercourses, and streets removed.

scale, shape, and compactness. Scale refers to the minimum object size, which is used as the stopping criteria for the algorithm. It is not bounded but mathematically, it cannot be lesser than one or greater than the size of the input image. The shape indicates the weight to put on the form of objects during the segmentation process as compared to color (spectral) information. Any change made to shape inversely affects color. The sum of the shape and color weights is equal to 1. Color is always required during the segmentation process, hence shape ranges from 0 to 0.9. The compactness weight defines the influence of the squareness of objects as opposed to their smoothness. The compactness and smoothness weights also add up to 1. Compactness ranges from 0 to 1. For a more detailed mathematical explanation of the MRS algorithm, readers are referred to the relevant literature (Baatz and Schäpe, 2000; Benz et al., 2004; Trimble Germany GmbH, 2019).

3.2.2. Evaluation of segmentation quality

Empirical segmentation evaluation (Zhang, 1996) was adopted for this research. It involves the computation of the geometric discrepancy or similarity between the LPIS and each segmentation layer. For each segmentation layer, the first step is to identify the segment that corresponds to a reference parcel in the LPIS. This was done using the two-sided overlap criteria (Clinton et al., 2010). A segment was considered to be a corresponding segment if the area of the intersection between that segment and a reference parcel was either more than half of the area of the segment or the reference parcel. A modification was made such that if a segment has more than one reference parcel with the same land-use type, those parcels are merged as a single reference parcel for that segment (Fig. 5). This was done to minimize under-segmentation.

The similarity between the reference parcel and the corresponding segment was computed via the Jaccard index (Jaccard, 1901), which is popularly known as Intersection over Union (IoU). It is a statistic widely used in computer vision tasks to measure the accuracy at which objects in an image or a video are detected by an algorithm. Its mathematical formulation is shown in Eq. (1);

$$IoU(Y) = \frac{Area(X \cap Y)}{Area(X \cup Y)} \quad (1)$$

where X is the reference parcel, Y is the corresponding segment, $X \cap Y$ is the spatial intersection between the two objects and $X \cup Y$ represents the spatial union of the two objects. It is bounded between 0 (no spatial similarity) and 1 (complete spatial match). The overall segmentation quality (OSQ) of each segmentation layer was finally computed as a weighted average of IoU over all segments using Eq. (2);

$$OSQ = \frac{\sum_{i=1}^n Area(Y_i) * IoU(Y_i)}{\sum_{i=1}^n Area(Y_i)} \quad (2)$$

where Y represents a segment and n is the total number of segments in

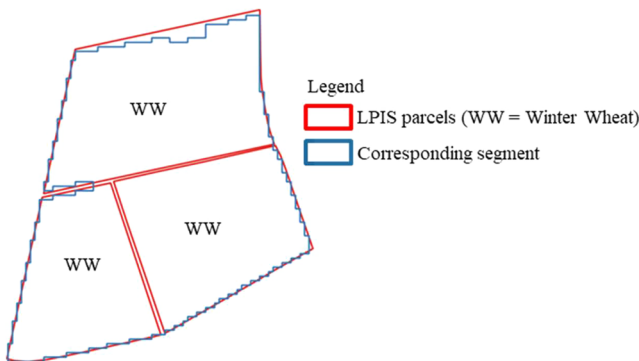


Fig. 5. The merger of reference parcels based on their land-use type. The segment corresponds to three reference parcels based on the two-sided overlap criteria. The three parcels were merged into one because winter wheat is grown on all of them.

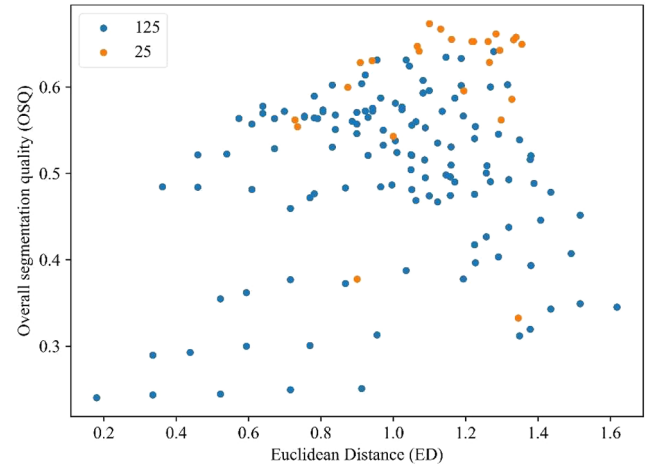


Fig. 6. The scatterplot showing the non-inverted OSQ computed for each segmentation layer of the test image. Each dot represents a data point of scale, shape, and compactness with its corresponding non-inverted OSQ. The blue dots (125) represent the initial parameter combinations and the red dots (25) represent the actual Bayesian iterations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

each segmentation layer. The segments along the spatial boundary of each segmentation layer were eliminated from the computation of the OSQ since they are artifacts created as a result of clipping the S2-VNIR images to the tiles.

3.2.3. Bayesian optimization

The essential components of our Bayesian optimization approach are:

- Domain space: this refers to the parameter space of each MRS parameter, of which the Bayesian optimization routine has to identify the optimal parameter combination. Scale ranged from 20 to 200, shape from 0 to 0.9, and compactness from 0 to 1.
- An objective function to minimize: our objective function, $f(x)$, takes a parameter combination, x , from the domain space, generates a segmentation vector layer, uploads the vector layer into a PostgreSQL database, computes the OSQ, and then returns an inverted OSQ as $1 - OSQ$.
- A surrogate model: it is a predictive probability model that captures the prior probability distribution, $p(y)$, of the objective function and is iteratively updated to capture the objective function's posterior probability distribution, $p(y|x)$, where y is the inverted OSQ. The surrogate model is a realization of the Bayes' rule (Eq. (3));

$$p(y|x) = \frac{p(x|y) * p(y)}{p(x)} \quad (3)$$

where $p(x|y)$ is a likelihood distribution and is a marginal probability. To build the posterior probability distribution, we need to define two things: the prior distribution function and the initial parameter combinations with their corresponding inverted OSQ. Two of the most used prior distribution functions are Gaussian Process (GP) (Rasmussen and Williams, 2006) and Random Forest (RF) (Breiman, 2001). However, GP has become a standard prior (Brochu et al., 2010; Dewancker et al., 2016) in Bayesian optimization. GP is parametrized by a mean function, μ , and covariance or kernel function, k . For convenience, μ is set as a zero function, leaving the user with the more interesting k , which defines the quality of the surrogate model (Brochu et al., 2010). The default choice of k for GP regression is the automatic relevance determination (ARD) squared exponential kernel (Brochu et al., 2010; Snoek et al., 2012). However, Snoek et al. (2012) recommended the

Table 1

The impact of different initial parameter combinations on the results of Bayesian optimization. The number of initial combinations for TS1 was 64 and TS2 was 27. The number of Bayesian iterations for TS1 was 86 and for TS2 123.

Test	Scale		Shape		Comp.		Optimal Parameters			OSQ	Time
	Range	Interval	Range	Interval	Range	Interval	Scale	Shape	Comp.		
TS1	20–200	60	0.0–0.9	0.3	0.0–0.9	0.3	60	0.884	0.919	66.82%	36 min
TS2	30–190	80	0.0–0.8	0.4	0.0–0.8	0.4	56	0.9	0.677	67.54%	49 min

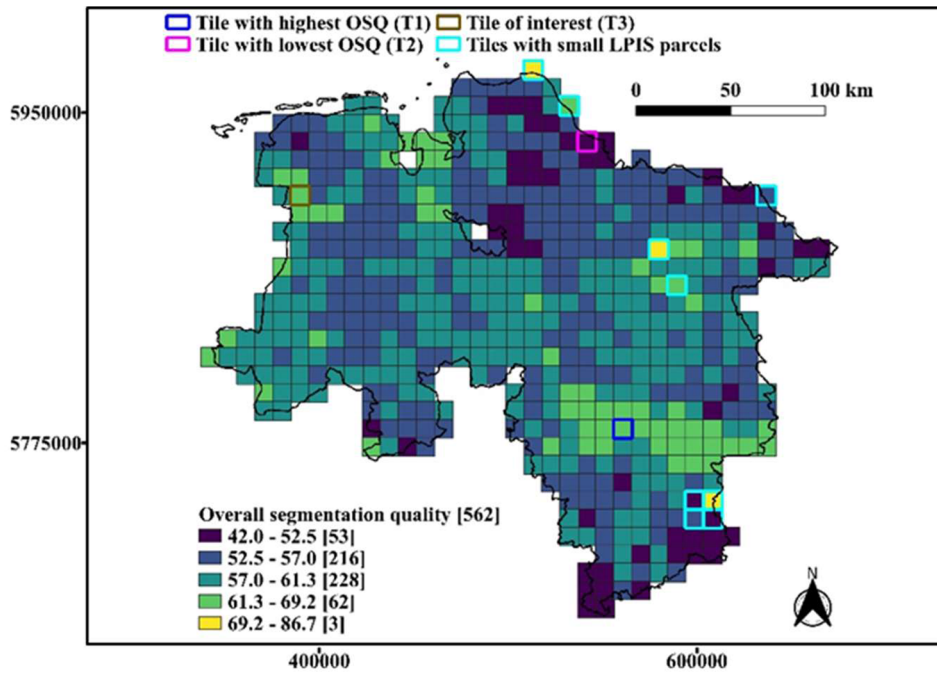


Fig. 7. The highest OSQ computed for each tile in Lower Saxony.

use of the ARD Matérn kernel (Stein, 1999) as captured by Eq. (4) because the squared exponential kernel is unrealistically smooth for practical optimization problems;

$$k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j) \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j) \right) \quad (4)$$

where ν and l are non-negative parameters, $d(x_i, x_j)$ is the distance between two parameter combinations x_i and x_j , Γ is the gamma function, and K_ν is the modified Bessel function (Rasmussen and Williams, 2006). We tested the first and second-order Matérn kernels recommended by Rasmussen and Williams (2006) for machine learning, and the first order proved superior, so we kept that. For first-order, ν is 1.5 and for the second-order, it is 2.5. The next step is to initialize the GP prior model with actual data. This is usually done by randomly sampling a user-given number of parameter combinations from the domain space and then the inverted OSQs are computed with $f(x)$. This randomness would prevent reproducibility, so we opted for systematic sampling. To obtain the initial samples, D , we always sampled 125 parameter combinations covering the low, middle and high ends of each parameter range. The values for scale are [40, 80, 120, 160, 200], and for both shape and compactness [0.1, 0.3, 0.5, 0.7, 0.9]. We used two parallel processes for segmenting and calculating the inverted OSQ of each parameter combination in D .

- (d) An acquisition function: it is used to propose new x combinations in the domain space to evaluate with $f(x)$ by making use of the GP posterior probability distribution, $p(y|x)$. Even though there are many acquisition functions, expected improvement (EI) (Jones et al., 1998) is the most commonly used (Frazier, 2018). The

possible improvement on the current optimal parameter combination, x^o , at any new parameter combination, x , is given by Eq. (5);

$$I(x) = \max \{ f(x^o) - f(x), 0 \} \quad (5)$$

where $f(x^o)$ is the inverted OSQ value at x^o . Given that $f(x)$ is computationally expensive to evaluate, the approach to identifying which x to evaluate next is rather to compute the expected value of $I(x)$ using the GP posterior, $p(y|x)$, which is faster to compute. Computing this expected value involves several partial integrations of $I(x)$ to obtain a closed form (Jones et al., 1998) as shown by Eq. (6);

$$EI(x) = (y^o - \hat{y}) \Phi \left(\frac{y^o - \hat{y}}{s} \right) + s \phi \left(\frac{y^o - \hat{y}}{s} \right) \quad (6)$$

where y^o is the inverted OSQ value at x^o , \hat{y} is the GP posterior $p(y|x)$, ϕ and Φ are the standard normal density and distribution functions, and s is the standard error of the GP posterior at x . Normally, many parameter combinations are randomly sampled from the given domain space, and the combination with the highest expected improvement is selected as a candidate and passed to $f(x)$ as the next point to evaluate. We used 10,000 random parameter combinations from our domain space as defined in (a). Alternatively, the x candidate can be identified using the Limited Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) (Liu and Nocedal, 1989) algorithm. From the 10,000 random combinations, we identified the first five combinations with the highest expected improvement. Using these combinations as initialization points, L-BFGS is iteratively able to find the local minima of EI, of which the optimal is then selected as the best candidate. We chose L-BFGS simply because it is more intuitive and ensures some level of reproducibility. This x candidate is then given to $f(x)$ to compute the inverted OSQ. The x

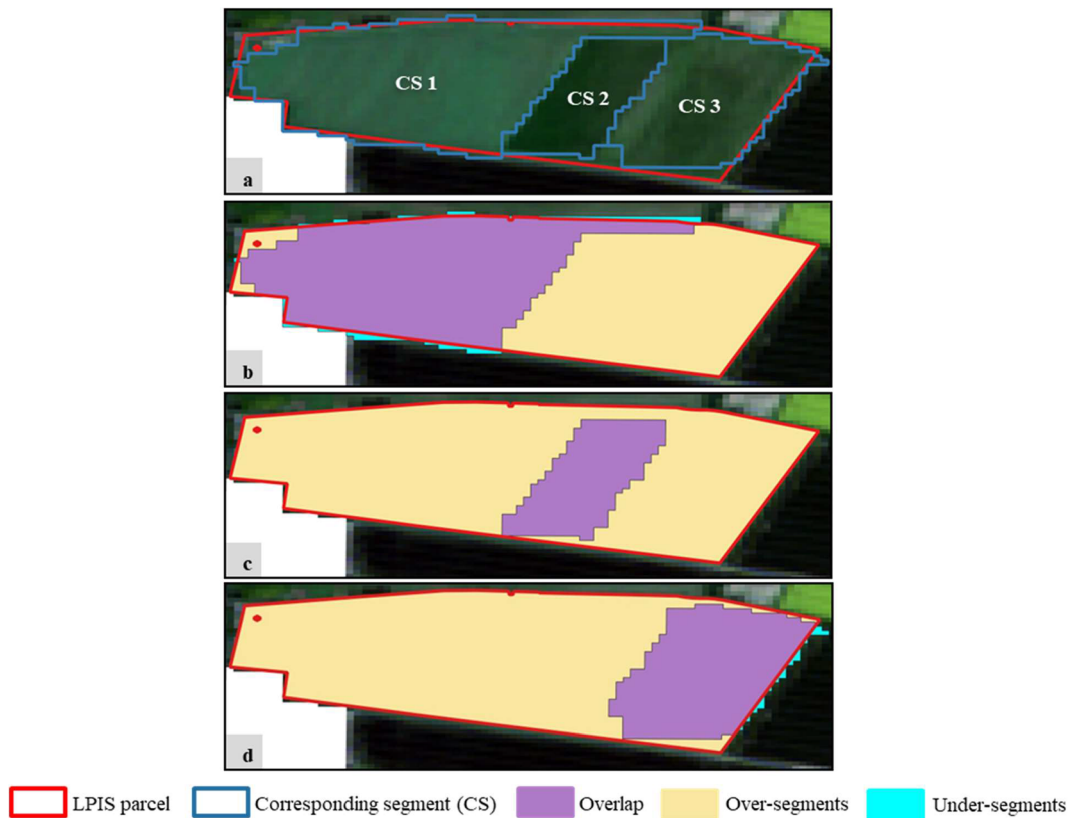


Fig. 8. Some of the contributing factors preventing the segmentation quality from reaching 100%. (a) A parcel in LPIS with three corresponding segments (CSs) overlaid on a Sentinel-2 image. The overlapping area, over-segments, and under-segments between each CS and the LPIS parcel are shown in (b), (c) and (d) respectively.

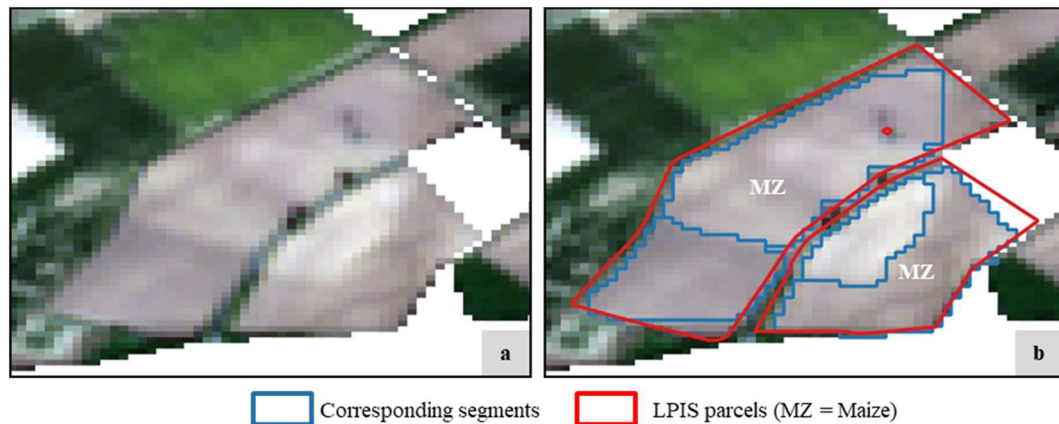


Fig. 9. The prevalent instances of over-segmentation. (a) A sentinel 2 image. (b) LPIS parcels and their corresponding segments overlaid on the Sentinel-2 image. Two corresponding segments were respectively created for each reference parcel due to the heterogeneity of the soil pixels in both parcels.

candidate and its inverted OSQ are then appended to D . The process is repeated from (c) to (d) with the current $p(y|x)$ replacing $p(y)$ to continuously update the posterior probability distribution. The repetition is done until a user-given number of iterations is completed. We allowed 150 function calls to $f(x)$ including the 125 initial samples, meaning the actual number of Bayesian iterations was 25.

For the complete mathematical foundation of Bayesian optimization, readers are referred to Brochu et al. (2010), Frazier (2018) and Shahriari et al. (2016). For automation purposes, we used eCognition Server 9.5.0 and its command-line interface (CLI) for segmentation. We limited parallel execution of the MRS segmentation to two because we have two eCognition Server licenses. The Python programming language was used to chain everything together. We used the Bayesian

optimization implementation of *Scikit-optimize* in Python. It is important to mention that, the computed OSQ was inverted because *Scikit-optimize* is programmed for function minimization.

We randomly selected one of the 562 tiled images to test the effectiveness of our Bayesian optimization approach in identifying the optimal MRS parameters. To visualize the parameter combinations sampled by the Bayesian optimization routine alongside the corresponding non-inverted OSQs in two-dimensions, we first conflated each parameter combination of scale, shape, and compactness into a single value using Euclidean Distance (ED). Each scale value was normalized between zero and one before being used in the calculation of the ED. We then plotted each ED against its corresponding non-inverted OSQ (Fig. 6). Our Bayesian optimization approach was very efficient as it

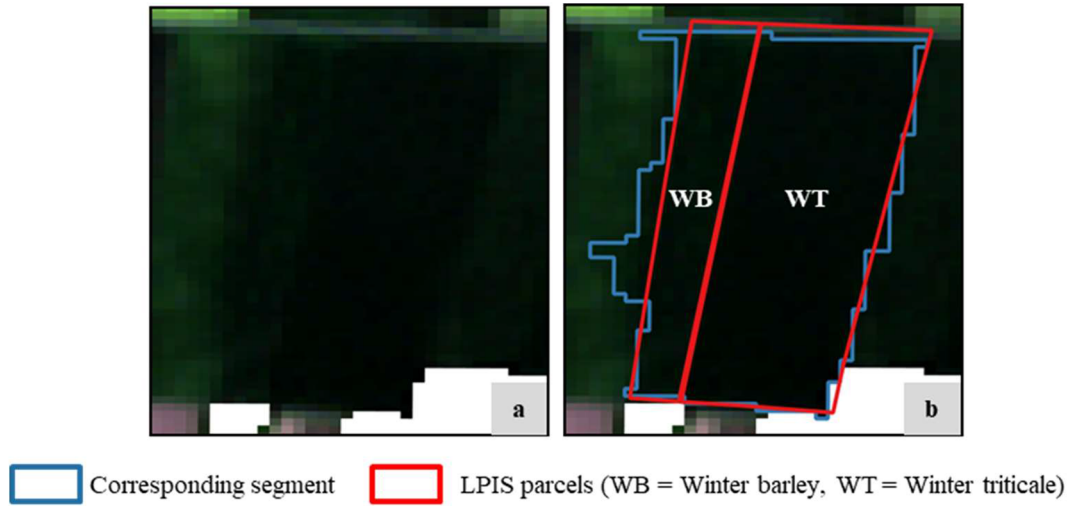


Fig. 10. Under-segmentation caused by adjacent parcels with similar spectral behavior. (a) A Sentinel-2 image. (b) Two LPIS parcels and their corresponding segment overlaid on the Sentinel-2 image. One corresponding segment was created due to the spectral similarity of winter barley and winter triticale.

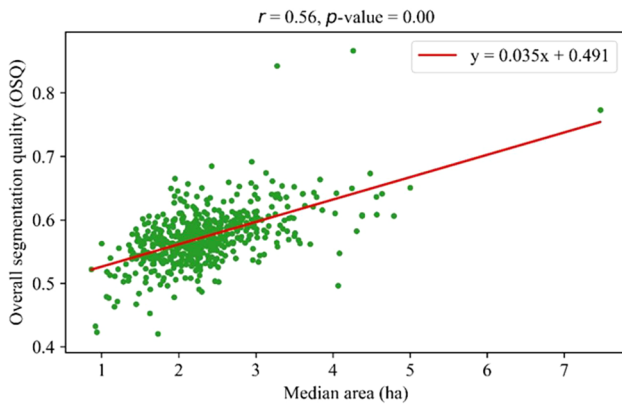


Fig. 11. Correlation between the highest OSQ and the median area of agricultural parcels for all tiles.

mostly exploited parameter combinations that yielded high OSQs. Out of the 25 combinations it sampled, 16 of them yielded OSQs above 60%. Using a total execution time of 16 min, the highest OSQ of 67.33% (ED = 1.1) was achieved with the parameter combination of 55 for scale, 0.9 for shape, and 0.602 for compactness.

Additional information discernable from Fig. 6 is that the distribution of the initial parameter combinations can play a role in identifying the optimal combination. Therefore, before applying our approach to the 562 tiles, we also tested it to see how the variation of the initial parameter combinations could affect the OSQ. Table 1 shows two tests labeled TS1 and TS2 that were done using different initial parameter combinations. It also captures the optimal parameter combination, OSQ, and execution time per test. The number of initial combinations for TS1 was 64 and for TS2 27. For all tests, the number of iterations was kept at 150, meaning 86 and 123 Bayesian iterations were run for TS1 and TS2 respectively.

From Table 1, the differences in OSQ between each test and the approach we adopted, which yielded an OSQ of 67.33%, was very marginal. Therefore, we concluded that the initial parameter combinations do not significantly affect the Bayesian optimization results as long as they are well distributed over the domain space and the number of Bayesian iterations is increased accordingly. Due to the increased number of Bayesian iterations, the execution time per test drastically increased. It is imperative to mention here that the actual optimal parameter combination needed for segmentation based on the MRS algorithm is unknown. This is true especially for shape and

compactness because they take floating-point numbers as input unlike scale, which accepts only integers, hence more deterministic. Additionally, different methods under different time constraints will most likely yield different results, an example of which is shown in Table 1. Therefore, a method that can approximate this unknown optimal combination in a time-efficient manner is the goal of any segmentation optimization approach. Our approach of using 125 initial combinations was more viable than the other two tests (TS1 and TS2) in terms of approximating the optimal MRS parameter combination within a shorter execution time. Therefore, we applied our approach to the 562 tiles in Lower Saxony to approximate the optimal segmentation parameters and delineate agricultural parcels. The execution was completed in seven days.

3.3. Other optimization methods

Two existing segmentation optimization methods based on the MRS algorithm in eCognition Developer were compared with our approach. The first one is the segmentation accuracy assessment (SAA) method (Anders et al., 2011). The SAA, just like our approach, is a supervised method. It creates segments at different scale levels, computes the discrepancy between reference objects and their corresponding segments at each level, and then identifies the level with the least discrepancy as optimal. To calculate the discrepancy measure for any segmentation layer, it first generates a frequency distribution from the spectral values of pixels that fall within each reference object and its corresponding segment, respectively. The two frequency distributions are then normalized with the respective number of pixels in each distribution. The segmentation error between the reference object and its corresponding segment is then calculated as the sum of absolute error between the two normalized frequency distributions. The discrepancy measure is finally computed as the average sum of absolute error over all segments (Eq. (7));

$$AAE = \frac{1}{n} \sum_{i=1}^n \left(\sum \left| \frac{H_x}{a_x} - \frac{H_y}{b_y} \right| \right)_i \quad (7)$$

where n is the total number of segments in the segmentation layer, H_x is the frequency distribution of pixels in the reference object, H_y is the frequency distribution of pixels in the corresponding segment, and a_x and b_y are the number of pixels within the reference object and the corresponding segment, respectively. The segmentation layer with the lowest AAE value is the optimal. The default values of 0.1 for shape and 0.5 for compactness were used. The scale ranged from 20 to 200 with

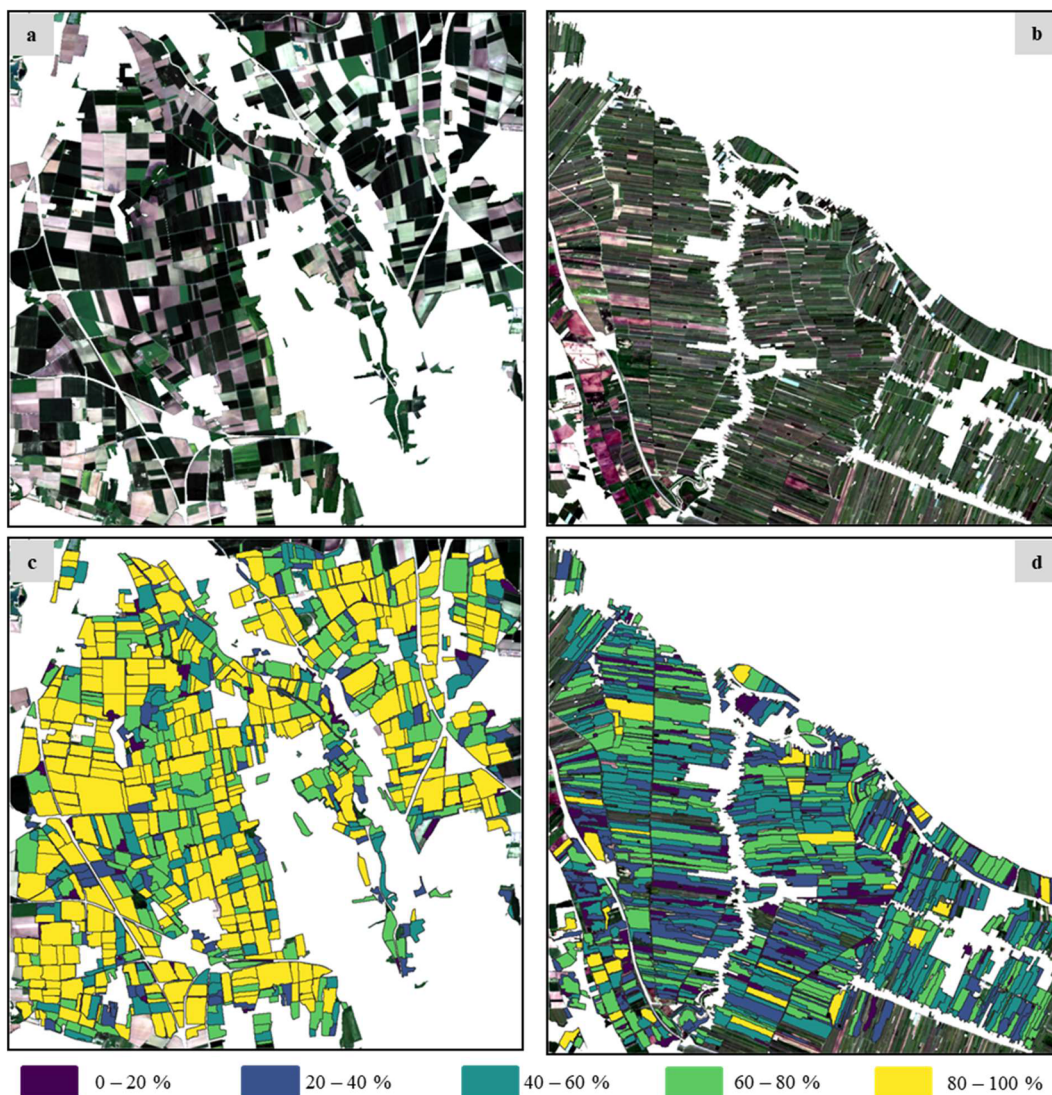


Fig. 12. Segmentation evaluation at the segment level for T1 (highest OSQ) and T2 (lowest OSQ). (a) and (b) show the Sentinel-2 images used for segmenting the T1 and T2 tiles respectively. The created segments have been colored according to the IoU computed for each of them and subsequently overlaid on each image respectively at (c) and (d). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

increments of 5.

The second optimization method is the estimation of scale parameter (ESP-2) tool (Drăguț et al., 2014), which is an improvement of the original version (Drăguț et al., 2010). Unlike the SAA and our approach, the ESP-2 tool is an unsupervised method, which is purely driven by the image content and does not use any reference data. Additionally, it is fully automated and applicable to images with multiple bands. Due to those characteristics, it is very popular in the world of MRS. It is underpinned on the concept of local variance (LV) (Woodcock and Strahler, 1987). It creates segments in a stepwise manner using incremental scale values. For each segmentation layer, the standard deviation of pixels in each segment is computed for each image band. The LV per band is calculated as the average standard deviation over all segments. The LV is finally averaged over all bands to obtain one LV per segmentation layer (Eq. (8));

$$ALV = \frac{1}{b} \sum_{j=1}^b \left(\frac{1}{n} \sum_{i=1}^n \sigma_i \right)_j \quad (8)$$

where b is the total number of bands in an image, n is the total number of segments in a segmentation layer, and σ_i is the standard deviation of pixels per segment. ALV is a measure indicating the level of

homogeneity within a segmentation layer. When the ALV of the current scale level is equal to or lower than the previous ALV , the iteration stops, and the segments created at the previous scale level are maintained. Here again, the default values of 0.1 for shape and 0.5 for compactness were kept and the scale was automatically determined by the ESP-2 tool.

4. Results and discussion

4.1. Analysis of segmentation quality

The highest OSQ identified for each tile is shown in Fig. 7. Most of the values ranged from 42.0% to 69.2%. The three tiles with values above 69.2% are highly dominated by non-agricultural land-use such that only a few reference parcels were used for segmentation evaluation. The lower the number of reference parcels, the higher the probability of obtaining high OSQ values. This relationship was also observed by Novelli et al., (2017), who emphasized the importance of using a high number of reference objects in supervised segmentation evaluation after establishing a positive correlation between segmentation accuracy and the amount of reference data used for evaluation.

It is important to state that a 100% OSQ is not achievable for a

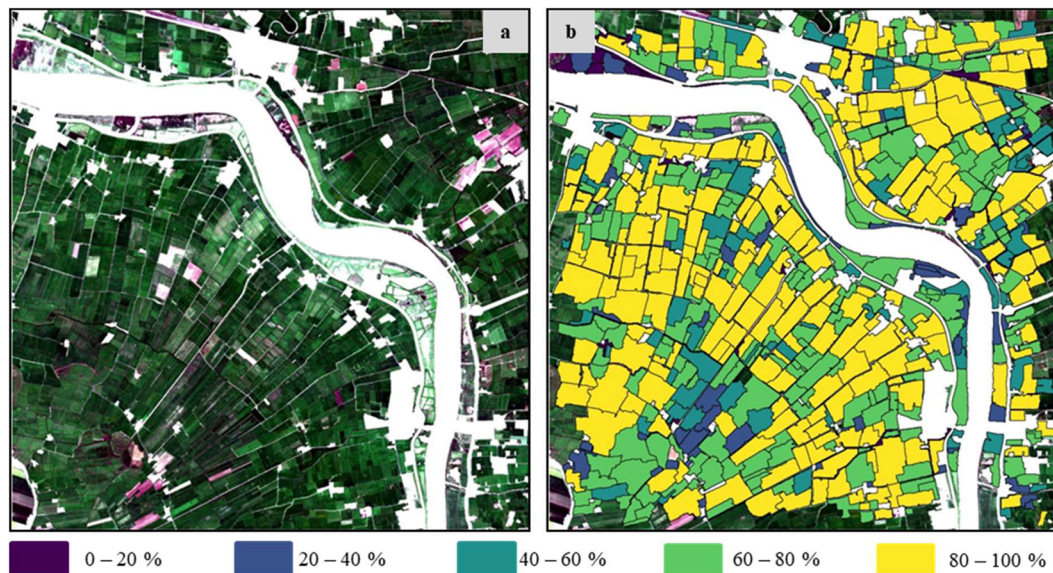


Fig. 13. Visual inspection of tile T3 with a relatively high OSQ of 68.46%. (a) The Sentinel-2 image of T3. (b) The evaluated corresponding segments colored by their respective IoU and draped over the Sentinel-2 image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

couple of reasons. The LPIS parcels are polygons with straight edges representing a single land-use digitized from orthophotos. Conversely, the segments automatically extracted from the Sentinel-2 images have pixelated edges. Additionally, the orthophotos have a higher spatial resolution (≤ 1 m) than the Sentinel-2 images. These differences culminate in the segments and the LPIS parcels being misaligned especially at the borders (Fig. 8a). Simplifying the segments did not eliminate this problem.

Over-segmentation was the main contributing factor that negatively affected the OSQ. It mostly occurred due to the heterogeneity of pixels within a parcel. In Fig. 8a, the LPIS indicates that there is one parcel on which three types of clover are grown. Due to the different growth stages, the MRS algorithm created three different segments (CS1, CS2, and CS3) as captured by Fig. 8a. This led to over-segmentation (Fig. 8b–d). Even though this type of over-segmentation is acceptable within the context of biodiversity monitoring and structural change analysis of agricultural parcels, it reduces the OSQ. The IoU computed respectively for CS1, CS2, and CS3 are 49.27%, 16.48% and 26.59%. Using those three segments, the OSQ amounted to 37.39%. When all the segments are first merged into a single polygon before computing the OSQ, the OSQ increased to 90.51%. This underscores the negative impact of over-segmentation on the OSQ. Even though this instance of over-segmentation led to a lower OSQ, it is entirely acceptable given that the segmentation algorithm correctly delineated the different parcels present in that area as visible from the satellite image. The inability of LPIS to correctly capture the different agricultural parcels present in that area was the negative driving force behind the low OSQ. Therefore, the low OSQ can largely be attributed to the error in LPIS and not the segmentation.

Fig. 9 captures another instance of over-segmentation. In Fig. 9b, both LPIS parcels are used to grow maize. When the Sentinel-2 image (Fig. 9a) was taken on 5 May 2018, the parcels were bare consisting of soil patches with different colors, which led to the creation of small fragments within those parcels (Fig. 9b). Such instances of over-segmentation were more prevalent. As over-segmentation increases, the OSQ decreases. In general, the best possible way to deal with over-segmentation will be to merge neighboring segments with the same land-use type after classifying the segments.

Under-segmentation, which mostly occurred when the same crop types are grown on adjacent parcels, was highly minimized due to the

modification made at the segmentation evaluation stage. In adjacent parcels with different crop types but similar spectral properties, under-segmentation was unavoidable. In Fig. 10b, even though the LPIS indicates that two distinct parcels are present, one big segment was created because winter barley and winter triticale have similar spectral properties as shown in Fig. 10a.

Another factor that influenced the segmentation quality was the size of agricultural parcels. Fig. 11 shows the linear relationship between the OSQ and the median area of agricultural parcels per tile. The Pearson correlation coefficient (r) of 0.56 indicated that at tiles with larger agricultural parcels, the OSQ was higher as opposed to tiles with smaller agricultural parcels. Tiles with similar median areas have similar OSQ values, thereby naturally clustering together as visible in Fig. 7.

To further demonstrate the impact of the area of agricultural parcels on the segmentation quality, two tiles with contrasting agricultural parcel structures were selected for analysis at the segment level. Fig. 12a captures the Sentinel-2 image of tile T1 with the highest OSQ (69.17%), while Fig. 12b shows that of tile T2 with the lowest OSQ of 42.04%. The median area of agricultural parcels in T1 and T2 is 4.12 ha and 1.73 ha respectively. The segments created for each tile are shown in Fig. 12c and d respectively. Each segment is colored by the geometric match, here the IoU, between that segment and its corresponding LPIS parcel. In Fig. 12a, the agricultural parcels are big and compact. Different agricultural land-use types like sugar beets and winter wheat exist there. This made it easier to delineate the parcels, which led to most parcels having high IoU values. Fig. 12b, on the other hand, shows that the parcels are small and elongated. Almost all of them are used to grow pome fruits with virtually no boundaries between them discernable from the Sentinel-2 image. This led to the creation of segments way bigger than the LPIS parcels, which led to most of the segments having very low IoU values. This consequently led to a low OSQ for that tile.

The fidelity of the OSQ was checked by visually inspecting tile T3, which has a relatively high OSQ of 68.46%. The Sentinel-2 image and the generated segments are shown in Fig. 13. About 86% of this tile is made up of pasture lands. With our segmentation evaluation process mostly focused on minimizing under-segmentation, it culminated in the creation of big segments, which were not representative of real-world agricultural parcels. Therefore, the OSQ should not be used alone but supported with visual inspection to make full deductions on

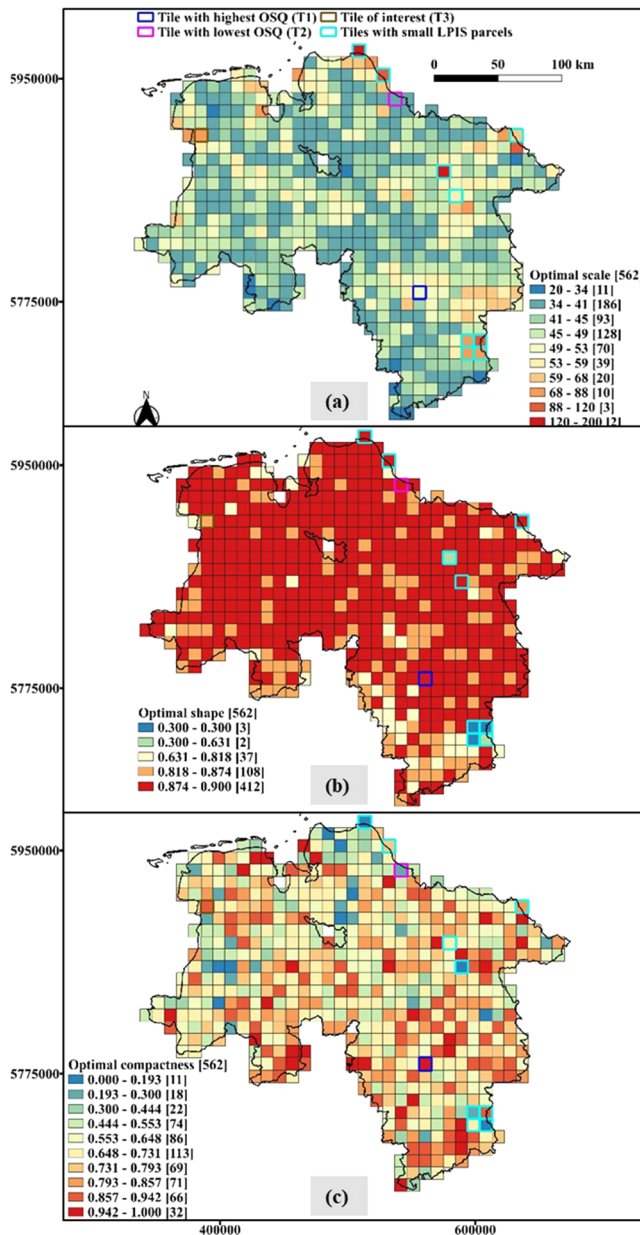


Fig. 14. The optimal parameter values identified for each tile in Lower Saxony.

segmentation quality.

4.2. Optimal parameter combination per tile

The scale, shape, and compactness values that resulted in the highest OSQ for each tile are presented here. Fig. 14a shows the optimal scale values. Most of the scale values were from 34 to 53, with the modal value being 40. Scale values above 77 correspond to the tiles with a very small number of LPIS parcels for evaluation. The optimal shape value identified for each tile is depicted in Fig. 14b. The shape values did not show a lot of variability over its possible range. Most of the values were above 0.8 and often reached the maximum of 0.9, which signifies the dominance of the shape of the agricultural parcels as compared to their spectral information during the segmentation evaluation process. Fig. 14c shows the optimal compactness values for the tiles. Unlike shape, the compactness values were so variable that visible clusters were not established. Even though the values stretched over the possible compactness range, most of them were above 0.5, with 0.7 being very dominant. This stands to reason given that the optimal shape

Table 2

The optimal parameters and corresponding OSQ values obtained by the two optimization methods and our approach.

Tile	Method	Shape	Compactness	Scale	OSQ
T1	SAA	0.1	0.5	85	55.65%
	ESP-2	0.1	0.5	73	50.81%
	Our approach	0.9	0.966	51	69.17%
T2	SAA	0.1	0.5	45	30.24%
	ESP-2	0.1	0.5	147	33.23%
	Our approach	0.9	0.3	40	42.04%
T3	SAA	0.1	0.5	65	48.59%
	ESP-2	0.1	0.5	83	52.52%
	Our approach	0.842	0.906	77	68.46%

values were relatively high. Additionally, in Germany most agricultural farms have square or rectangular shapes, hence very compact.

4.3. Comparison with other optimization methods

The SAA and ESP-2 methods were applied to the T1, T2, and T3 images. Table 2 captures the optimal parameters and corresponding OSQ values obtained by the two optimization methods and ours. As a reminder, the other two methods only optimized scale, while shape and compactness were kept at their default. At all tiles, SAA and ESP-2 obtained different scale and OSQ values, which is contrary to the results of Belgiu and Drăguț (2014). Belgiu and Drăguț (2014) evaluated the SAA and ESP-2 methods for optimal extraction of buildings from very high-resolution satellite images. In the test areas with big buildings, both methods achieved very similar results. In general, buildings are very compact and have very homogeneous surfaces making it easier to delineate them compared to agricultural parcels. This could have contributed to the differences in the results alongside the different satellite images used.

At all three tiles, the OSQs of SAA and ESP-2 were significantly lower than our approach (Table 2). The optimal shape and compactness values identified by our approach gave a better indication of the structural composition of parcels in each tile. Further, the visual assessment of the IoU computed at the segment level showed that there were more segments with higher qualities based on our optimization approach (Fig. 15a–c) than the SAA (Fig. 15d–f) and the ESP-2 (Fig. 15g–i) methods.

The outcome of the two methods using the optimal shape and compactness values identified with our optimization approach is shown in Table 3. This time around, their OSQs improved significantly and got closer to those of our approach. This underscores the importance of determining the optimal values not only for scale as is done by the SAA and ESP-2 but also for the other MRS parameters. This is what differentiates our approach from the other two, making ours demonstrably more accurate. At T2, the segmentation challenge remained. Both methods, as well as our approach, performed poorly due to the overwhelming presence of small and elongated agricultural parcels. This is more of a data issue than the segmentation optimization method. The likely solution to this problem is the use of an image with a higher spatial resolution than Sentinel-2 such that distinct boundaries between the agricultural parcels can be identified, thereby making it easier to delineate the parcels while minimizing under-segmentation.

5. Conclusions

Accurate and up-to-date information on agricultural parcels is pivotal to any agricultural management system. The most prominent spatial database of agricultural parcels within the European Union (EU) called the Land Parcel Identification System (LPIS) suffers certain drawbacks such as the inadequate coverage of all agricultural parcels, restricted access to the data, the time lag that comes with the data, and

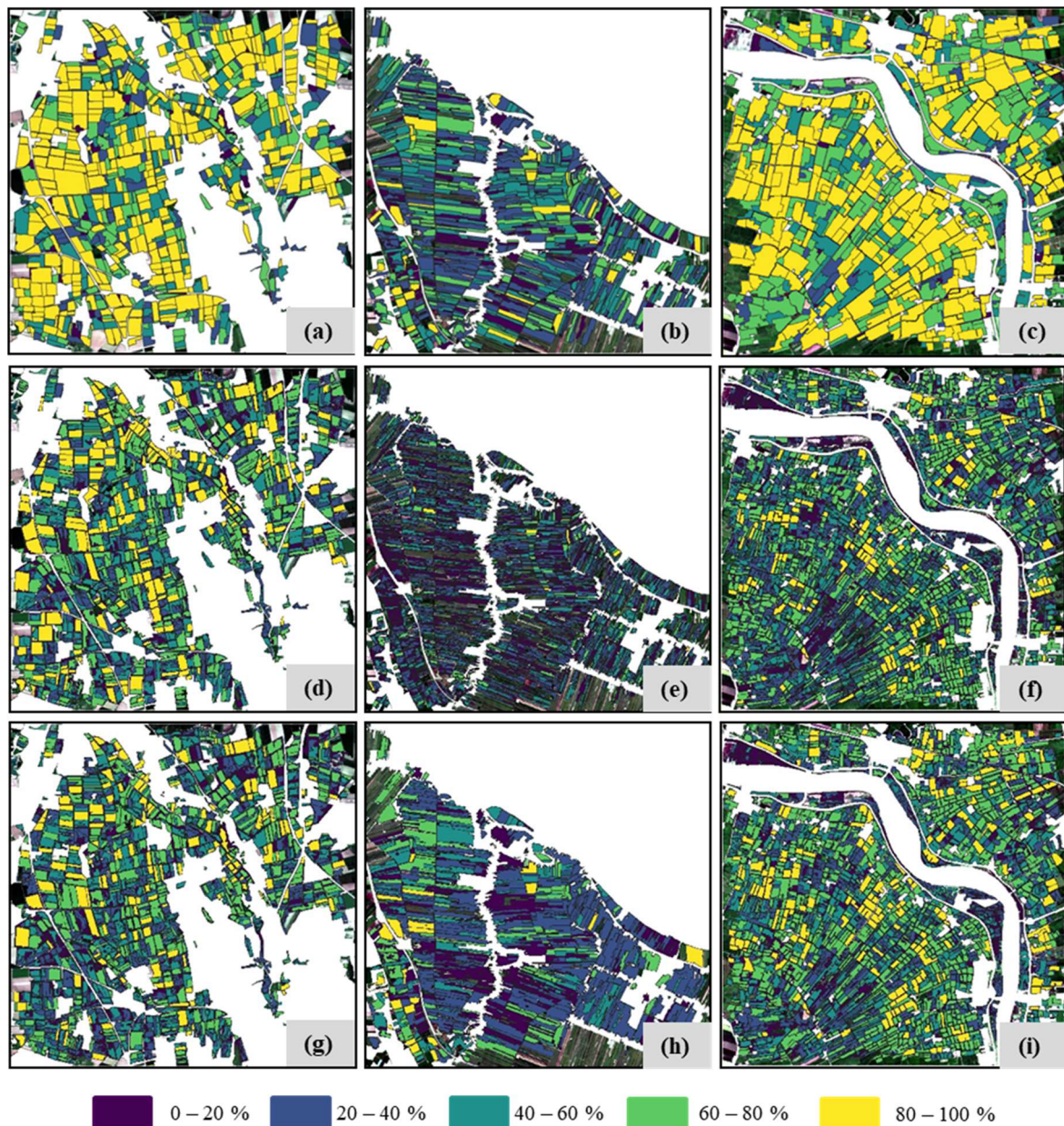


Fig. 15. Segmentation evaluation at the segment level for the different optimization approaches at T1 (a, d, g), T2 (b, e, h), and T3 (c, f, i). Each segment is symbolized by its IoU and draped over their respective images. (a)–(c) are the segments based on our optimization approach, (d)–(f) are based on the SAA method, and (g)–(i) are those of the ESP-2 method.

Table 3

The optimal scale and corresponding OSQ values obtained by the other optimization methods using the shape and compactness values identified with our approach.

Tile	Shape	Compactness	Scale		OSQ	
			SAA	ESP-2	SAA	ESP-2
T1	0.9	0.966	35	45	63.91%	68.75%
T2	0.9	0.3	30	54	42.00%	39.30%
T3	0.842	0.906	35	64	58.50%	67.48%

the different implementation methods used by the different EU countries to generate the data leading to different sources of error. To deal with those drawbacks, a supervised and automated Bayesian optimization framework was developed to identify the optimal parameters of the Multiresolution Segmentation (MRS) algorithm for segmenting

agricultural parcels in the federal state of Lower Saxony in Germany based on mono-temporal Sentinel-2 images.

To determine the optimal parameters, an area-weighted Jaccard index was used as a proxy for segmentation quality. The parameter combination with the highest weighted Jaccard index was adjudged the optimal for each 10 km × 10 km tile grid in Lower Saxony. The established optimal parameters were variable especially the compactness and scale, which indicated that a single parameter combination could not have guaranteed optimal segmentation for Lower Saxony. This reinforces the significance of our approach to determine the optimal parameters for different parts of Lower Saxony using tile grids. Given that the sizes of agricultural parcels in Germany do not drastically change from one year to another, the MRS parameters established for one year can potentially be used to segment agricultural parcels from images acquired within the same time window (month) from another year. We came to this preliminary conclusion after doing two tests. In

the first test, we used the MRS parameters established for tiles in May of 2018 to segment cloud-free images of those same tiles in May of 2019 and then used the LPIS data of 2019 to calculate the overall segmentation quality (OSQ) per tile. For the second test, we applied our Bayesian optimization approach to optimally segment those cloud-free images in May of 2019 using the LPIS data of 2019 as a reference and subsequently identified the OSQ of the optimal parameter combination per tile. The average difference in OSQ between those two tests over all the tiles was below 2%. Therefore, in the absence of LPIS for a particular year, the optimal MRS parameters established from a previous year can be used to segment images acquired within the same month of the current year. The developed approach can also be used to segment images taken at different times of the year to do in-season monitoring of the structural changes on agricultural parcels. Our approach outperformed the scale optimization method of the SAA and ESP-2 in all test areas. Those methods only focus on scale optimization, neglecting the other MRS parameters like shape and compactness. Given that different agricultural landscapes may have different structural compositions, our approach proved the importance of optimizing all three MRS parameters to achieve optimal segmentation. Our approach is independent of the input data, hence can be applied to any satellite image and reference data to optimize segmentation.

The research showed that the structural composition of agricultural parcels in a particular area influences the segmentation quality. The bigger the sizes of agricultural parcels are, the higher the segmentation quality. Over-segmentation was another factor that influenced the segmentation quality. It showed up when crops on a parcel are at substantially different stages of growth or when pixels within a parcel are very heterogeneous, thereby leading to the creation of small objects within a parcel. Under-segmentation, on the other hand, was largely dealt with in this research by merging LPIS parcels of the same land-use type during the segmentation evaluation process. This research also revealed that discrepancy measures alone do not give a complete picture of segmentation quality. Therefore, they should not be used in isolation but supported by visual inspection to make final decisions.

It is imperative to finally mention here that as we used LPIS as reference data to optimize the segmentation process, we did not achieve better geometric results than the LPIS as we saw in the best obtained OSQ being 69.17% at T1. LPIS is generated based on very high-resolution orthoimages with the spatial resolution being at least 1 m. We used Sentinel-2, which has a lower spatial resolution. With very high-resolution orthoimages like those used to create the LPIS, our Bayesian optimization approach can potentially be used to generate segments with similar geometric accuracy as the LPIS.

6. Future outlook

Going into the future, we will focus on the tiles with low segmentation quality to develop new methods of improving the segmentation quality. In this research, we applied our optimization approach to mono-temporal images. Therefore, multi-temporal Sentinel-2 images would be tested. Other auxiliary datasets like a Digital Elevation Model (DEM) and soil map would be used to augment the satellite images during the segmentation process to check if they can improve the segmentation quality. Instead of segmenting all agricultural areas in an image at once, segmentation would be done by separating arable and grassland areas. Initial tests based on mono-temporal images show promising results when this separation is done. In areas dominated by a single land-use, the merger of reference parcels with the same land-use would be turned off during the segmentation evaluation stage. This can potentially lead to the creation of smaller segments. Finally, another area of possible improvement could be the creation of tile grids based on similar structural compositions of agricultural parcels.

CRediT authorship contribution statement

Gideon Okpoti Tetteh: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft. **Alexander Gocht:** Conceptualization, Supervision, Writing - review & editing. **Christopher Conrad:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgments

The authors thank the Ministry of Food, Agriculture and Consumer Protection of Lower Saxony for providing the LPIS reference layer, the German Federal Agency for Cartography and Geodesy (BKG) for providing the ATKIS dataset, and the European Space Agency (ESA) for the Sentinel-2 images. The authors are grateful to the Thünen Institute of Farm Economics for providing the computing resources and working environment for this research. We are also grateful to Stefan Erasmi for proofreading the article.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compag.2020.105696>.

References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 2274–2282. <https://doi.org/10.1109/TPAMI.2012.120>.
- Anders, N.S., Seijmonsbergen, A.C., Bouten, W., 2011. Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping. *Remote Sens. Environ.* 115, 2976–2985. <https://doi.org/10.1016/j.rse.2011.05.007>.
- Atzberger, C., 2013. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. *Remote Sens.* 5, 949–981. <https://doi.org/10.3390/rs5020949>.
- Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informations-Verarbeitung XII*. Wichmann Verlag, Karlsruhe, pp. 12–23.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 204, 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>.
- Belgiu, M., Drăguț, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS J. Photogramm. Remote Sens.* 96, 67–75. <https://doi.org/10.1016/j.isprsjprs.2014.07.002>.
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* 58, 239–258. <https://doi.org/10.1016/j.isprsjprs.2003.10.002>.
- Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13, 280–305.
- Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), *Advances in Neural Information Processing Systems 24*. Curran Associates Inc, pp. 2546–2554.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brochu, E., Cora, V.M., de Freitas, N., 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *ArXiv10122599 Cs*.
- Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. *Photogramm. Eng. Remote Sens.* 76, 289–299. <https://doi.org/10.14358/PERS.76.3.289>.
- Conrad, C., Fritsch, S., Zeidler, J., Rucker, G., Dech, S., 2010. Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sens.* 2, 1035–1056. <https://doi.org/10.3390/rs2041035>.
- Dewancker, I., McCourt, M., Clark, S., 2016. Bayesian Optimization Primer [WWW Document]. *Bayesian Optim. Primer*. URL <https://app.sigopt.com/static/pdf/SigOpt>.

- Bayesian Optimization Primer.pdf (accessed 3.4.20).
- Drăguț, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* 88, 119–127. <https://doi.org/10.1016/j.isprsjprs.2013.11.018>.
- Drăguț, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inf. Sci.* 24, 859–871. <https://doi.org/10.1080/13658810903174803>.
- Dudley, N., Alexander, S., 2017. Agriculture and biodiversity: a review. *Biodiversity* 18, 45–49. <https://doi.org/10.1080/14888386.2017.1351892>.
- Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H.H., Leyton-Brown, K., 2013. Towards an Empirical Foundation for Assessing Bayesian Optimization of Hyperparameters [WWW Document]. *Empir. Found. Assess. Bayesian Optim. Hyperparameters*. URL https://ml.informatik.uni-freiburg.de/papers/13-BayesOpt_EmpiricalFoundation.pdf (accessed 3.22.20).
- Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. *Nature* 478, 337–342. <https://doi.org/10.1038/nature10452>.
- Frazier, P.I., 2018. A Tutorial on Bayesian Optimization. *ArXiv180702811 Cs Math Stat.*
- Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory* 21, 32–40. <https://doi.org/10.1109/TIT.1975.1055330>.
- García-Pedrero, A., Gonzalo-Martín, C., Lillo-Saavedra, M., 2017. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *Int. J. Remote Sens.* 38, 1809–1819. <https://doi.org/10.1080/01431161.2016.1278312>.
- Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sens.* 8, 166. <https://doi.org/10.3390/rs8030166>.
- Jaccard, P., 1901. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz.
- Ji, C.Y., 1996. Delineating agricultural field boundaries from TM imagery using dyadic wavelet transforms. *ISPRS J. Photogramm. Remote Sens.* 51, 268–283. [https://doi.org/10.1016/0924-2716\(95\)00017-8](https://doi.org/10.1016/0924-2716(95)00017-8).
- Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* 13, 455–492. <https://doi.org/10.1023/A:1008306431147>.
- Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. *Comput. Electron. Agric.* 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>.
- Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégué, A., 2017. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated sentinel-2 time series, VHRS and DEM). *Remote Sens.* 9, 259. <https://doi.org/10.3390/rs9030259>.
- Li, P., Xiao, X., 2007. Multispectral image segmentation by a multichannel watershed-based approach. *Int. J. Remote Sens.* 28, 4429–4452. <https://doi.org/10.1080/01431160601034910>.
- Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y., 2016. Fully Convolutional Instance-aware Semantic Segmentation. *ArXiv161107709 Cs*.
- Liu, D.C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. *Math. Program.* 45, 503–528. <https://doi.org/10.1007/BF01589116>.
- Lizotte, D., 2007. *Automatic gait optimization with Gaussian process regression*. *Proc. IJCAI* 944–949.
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2. In: Bruzzone, L., Bovolo, F., Benediktsson, J.A. (Eds.), *Image and Signal Processing for Remote Sensing XXIII*. Presented at the Image and Signal Processing for Remote Sensing, SPIE, Warsaw, Poland, p. 3. <https://doi.org/10.1117/12.2278218>.
- Marchant, R., Ramos, F., 2012. Bayesian optimisation for Intelligent Environmental Monitoring. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Presented at the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012). IEEE, Vilamoura-Algarve, Portugal, pp. 2242–2249. <https://doi.org/10.1109/IROS.2012.6385653>.
- Marpu, P.R., Neubert, M., Herold, H., Niemeyer, I., 2010. Enhanced evaluation of image segmentation results. *J. Spat. Sci.* 55, 55–68. <https://doi.org/10.1080/14498596.2010.487850>.
- Mesner, N., Oštir, K., 2014. Investigating the impact of spatial and spectral resolution of satellite images on segmentation quality. *J. Appl. Remote Sens.* 8, 083696. <https://doi.org/10.1117/1.JRS.8.083696>.
- Mockus, J., 2012. *Bayesian Approach to Global Optimization: Theory and Applications*. Springer Science & Business Media.
- Nasrallah, A., Baghdadi, N., Mhaweji, M., Faour, G., Darwish, T., Belhouchette, H., Darwich, S., 2018. A novel approach for mapping wheat areas using high resolution sentinel-2 images. *Sensors* 18, 2089. <https://doi.org/10.3390/s18072089>.
- Neubert, M., Herold, H., Meinel, G., 2008. Assessing image segmentation quality – concepts, methods and application. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), *Object-based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg, pp. 769–784. https://doi.org/10.1007/978-3-540-77058-9_42.
- Novelli, A., Aguilar, M., Aguilar, F., Nemmaoui, A., Tarantino, E., 2017. AssesSeg—a command line tool to quantify image segmentation quality: a test carried out in Southern Spain from satellite imagery. *Remote Sens.* 9, 40. <https://doi.org/10.3390/rs9010040>.
- Osborne, M.A., Garnett, R., Roberts, S.J., 2010. Active Data Selection for Sensor Networks with Faults and Change-points. In: 2010 24th IEEE International Conference on Advanced Information Networking and Applications. Presented at the 2010 24th IEEE International Conference on Advanced Information Networking and Applications. IEEE, Perth, Australia, pp. 533–540. <https://doi.org/10.1109/AINA.2010.36>.
- Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E., Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sens. Environ.* 115, 1301–1316. <https://doi.org/10.1016/j.rse.2011.01.009>.
- Persello, C., Tolpekin, V.A., Bergado, J.R., de By, R.A., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* 231, 111253. <https://doi.org/10.1016/j.rse.2019.111253>.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning*. MIT Press, Cambridge, Mass.
- Rieke, C., 2017. *Deep Learning for Instance Segmentation of Agricultural Fields* (Master's thesis dissertation). Friedrich-Schiller-University Jena, Thuringia, Germany.
- Schmedtmann, J., Campagnolo, M., 2015. Reliable crop identification with satellite imagery in the context of common agriculture policy subsidy control. *Remote Sens.* 7, 9325–9346. <https://doi.org/10.3390/rs70709325>.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N., 2016. Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* 104, 148–175. <https://doi.org/10.1109/JPROC.2015.2494218>.
- Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian Optimization of machine learning algorithms. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 2951–2959.
- Stein, M.L., 1999. *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4612-1494-6>.
- Taşdemir, K., Wirthardt, C., 2012. Neural network-based clustering for agriculture management. *EURASIP J. Adv. Signal Process.* 2012. <https://doi.org/10.1186/1687-6180-2012-200>.
- Thornton, C., Hutter, F., Hoos, H.H., Leyton-Brown, K., 2013. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '13*. Presented at the the 19th ACM SIGKDD international conference. ACM Press, Chicago, Illinois, USA, p. 847. 10.1145/2487575.2487629.
- Trimble Germany GmbH, 2019. *eCognition Developer 9.5.0 Reference Book*. Trimble Germany GmbH, Germany.
- Turker, M., Kok, E.H., 2013. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS J. Photogramm. Remote Sens.* 79, 106–121. <https://doi.org/10.1016/j.isprsjprs.2013.02.009>.
- Vogels, M.F.A., de Jong, S.M., Sterk, G., Douma, H., Addink, E.A., 2019. Spatio-temporal patterns of smallholder irrigated agriculture in the horn of Africa using GEOBIA and sentinel-2 imagery. *Remote Sens.* 11, 143. <https://doi.org/10.3390/rs11020143>.
- Watkins, B., Van Niekerk, A., 2019. Automating field boundary delineation with multi-temporal Sentinel-2 imagery. *Comput. Electron. Agric.* 167, 105078. <https://doi.org/10.1016/j.compag.2019.105078>.
- Woodcock, C.E., Strahler, A.H., 1987. The factor of scale in remote sensing. *Remote Sens. Environ.* 21, 311–332. [https://doi.org/10.1016/0034-4257\(87\)90015-0](https://doi.org/10.1016/0034-4257(87)90015-0).
- Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sens. Environ.* 144, 42–64. <https://doi.org/10.1016/j.rse.2014.01.006>.
- Zhang, Y.-J., 1996. A survey on evaluation methods for image segmentation. *Pattern Recognit.* 29, 1335–1346.
- Zielinski, R., Grandgirard, D., European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, 2008. *Land Parcel Identification System (LPI) anomalies' sampling and spatial pattern towards convergence of ecological methodologies and GIS technologies*. Publications Office, Luxembourg.