

Improving hyperspectral image segmentation by applying inverse noise weighting and outlier removal for optimal scale selection

Phuong D. Dao^{a,b,*}, Kiran Mantripragada^c, Yuhong He^{a,*}, Faisal Z. Qureshi^c

^a Department of Geography, Geomatics and Environment, University of Toronto, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada

^b School of the Environment, University of Toronto, 33 Willcocks Street, Toronto, ON M5S 3E8, Canada

^c Faculty of Science, Ontario Tech University, 2000 Simcoe Street North, Oshawa, ON L1G 0C5, Canada



ARTICLE INFO

Keywords:

Inverse noise weighting
Outlier detection
Optimal scale selection
Image segmentation
Object-Based Image Analysis
Hyperspectral image classification

ABSTRACT

Optimal scale selection for image segmentation is an essential component of the Object-Based Image Analysis (OBIA) and interpretation. An optimal segmentation scale is a scale at which image objects, overall, best represent real-world ground objects and features across the entire image. At this scale, the intra-object variance is ideally lowest and the inter-object spatial autocorrelation is ideally highest, and a change in the scale could cause an abrupt change in these measures. Unsupervised parameter optimization methods typically use global measures of spatial and spectral properties calculated from all image objects in all bands as the target criteria to determine the optimal segmentation scale. However, no studies consider the effect of noise in image spectral bands on the segmentation assessment and scale selection. Furthermore, these global measures could be affected by outliers or extreme values from a small number of objects. These issues may lead to incorrect assessment and selection of optimal scales and cause the uncertainties in subsequent segmentation and classification results. These issues become more pronounced when segmenting hyperspectral data with large spectral variability across the spectrum. In this study, we propose an enhanced method that 1) incorporates the band's inverse noise weighting in the segmentation and 2) detects and removes outliers before determining segmentation scale parameters. The proposed method is evaluated on three well-established segmentation approaches – k-means, mean-shift, and watershed. The generated segments are validated by comparing them with reference polygons using normalized over-segmentation (OS), under-segmentation (US), and the Euclidean Distance (ED) indices. The results demonstrate that this proposed scale selection method produces more accurate and reliable segmentation results. The approach can be applied to other segmentation selection criteria and are useful for automatic multi-parameter tuning and optimal scale parameter selections in OBIA methods in remote sensing.

1. Introduction

1.1. Image segmentation in Object-Based Image Analysis

The OBIA approaches that outperform traditional pixel-based schemes have received growing attention in recent years (Myint et al., 2011; Yu et al., 2006). OBIA leverages objects present in an image for subsequent tasks, such as image classification. The key advantage of OBIA over pixel-based processes is that the objects carry richer spectral and spatial information than pixels (Blaschke, 2010; Dao et al., 2019b). Image segmentation is often the first step to create objects in an OBIA pipeline. A large number of image segmentation techniques has been

explored over past decades and they can be divided into four categories (Hossain and Chen, 2019): 1) edge-based methods (Martin et al., 2004; Vincent and Soille, 1991; Z. Wang et al., 2005); 2) region-based methods (Baatz and Schäpe, 2010; Bellens et al., 2008; Epshtein et al., 2010; Karl and Maurer, 2010; Zhang et al., 2013); 3) hybrid approaches (Kruse et al., 1993; Yin et al., 2015); and 4) more recent Machine Learning (ML) techniques that rely on Support Vector Machines (SVMs) (Mitra et al., 2004) or Neural Networks (Kurnaz et al., 2005; Long et al., 2015).

With the increased availability of high spatial and spectral resolution imagery from a variety of close-range, airborne, and spaceborne imaging platforms, there is a growing demand for improved image segmentation approaches in the remote sensing community in general, and for

* Corresponding authors at: Department of Geography, Geomatics and Environment, University of Toronto, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada (P.D. Dao and Y. He).

E-mail addresses: phuong.dao@mail.utoronto.ca (P.D. Dao), yuhong.he@utoronto.ca (Y. He).

hyperspectral images in particular. Deep learning (DL), a subfield of ML, has become one of the most powerful and widely used segmentation methods (Gao et al., 2018; Mou et al., 2017; Nalepa et al., 2019; Zhong et al., 2017). Despite recent advances in DL-based segmentation techniques, however, hyperspectral image segmentation remains an open problem, in part due to the large variations in sizes, shapes, and spectral properties of objects. On the one hand, supervised DL segmentation algorithms (i.e. semantic segmentation) require large and representative ground-truth data at a high cost, and this is especially true when segmenting high-resolution hyperspectral (HrHS) images with a large number of spectral bands (Nalepa et al., 2019). On the other hand, unsupervised DL algorithms such as fully convolution–deconvolution network (Mou et al., 2018) for spectral-spatial feature learning, deep clustering with convolutional autoencoders (X. Guo et al., 2017), and 3-D convolutional autoencoders (Nalepa et al., 2019) have been proposed to segment images without known labelled data. However, the high computational expense of these unsupervised DL algorithms limits its applicability in hyperspectral image segmentation scale selection that requires empirical evaluation of segmentation results across large ranges of scales. Therefore, there is a need to develop robust algorithms that meet technical requirements while assuring computational expense when processing these large hyperspectral datasets.

1.2. Optimal segmentation scale selection

Image segmentation often requires a careful selection of scales to achieve satisfactory performance. A number of techniques have been proposed for task-specific optimal scale selection. A common scale selection approach is to visually inspect segmentation results and tweak the parameters to achieve acceptable results (B. Johnson and Xie, 2011). This approach is tedious, time-consuming, and relies upon the knowledge, interpretation ability, and experience of the individual tasked with image segmentation (Y. Wang et al., 2019). Equally importantly, this approach quickly becomes infeasible as the amount of data grows. Another approach is to ask the users to assess the performance of the classification task that comes after image segmentation (Dronova et al., 2012; Smith, 2010). This approach may not yield an optimal scale for image segmentation methods since the performance of the classification task depends not only on image segmentation results but also on the classifier itself. Similar to the visual inspection, this method also needs a user in the loop, which restricts its use to situations involving a small number of images. Quantitative analysis methods that do not require the involvement of users are widely used in the community. These methods use statistical measures extracted from segmentation results to optimize the image segmentation scale (Y. Wang et al., 2019). These methods leverage both within segment features and segment-segment relationships to pick the correct scale. In this work, we are interested in quantitative methods for segmentation scale selection that do not need user input.

Supervised quantitative methods (Belgiu and Drăguț, 2014; Clinton et al., 2010; Ghosh and Joshi, 2014; Su and Zhang, 2017; Winter, 2000; Witharana and Civco, 2014) use ground truth data to learn to select the optimal segmentation scale. The need for labelled data has major resource implications; therefore, these methods are not applicable in many situations. Unsupervised parameter tuning methods that do not need labelled data (Dao and Liou, 2015; Dao et al., 2015; Drăguț et al., 2010; Y. Wang et al., 2019; Yang et al., 2014), have been wide-used within the remote sensing community. These methods attempt to capture intra-segment homogeneity and inter-segment heterogeneity to select good parameter values (Y. Wang et al., 2019).

A number of indices and criteria have been proposed to evaluate the quality of unsupervised parameterization methods. Most of these methods use 1) intra-segment homogeneity and 2) inter-segment heterogeneity of segmented images as segmentation assessment criteria. For example, an area-weighted variance and Global Moran's I was used to identify optimal scale for image segmentation (Kim et al., 2008), and

evaluating the segmentation quality and parameterization (Böck et al., 2017; Espindola et al., 2006; B. Johnson and Xie, 2011). A multi-band spectral angle (i.e. a measure of spectral homogeneity of segments) was used as an indicator in the unsupervised selection of optimal scales for segmentation of QuickBird and Worldview-2 images (Yang et al., 2014). Object's Local Variance (LV) and Rate of Change (RoC) graphs, built upon the fundamental idea by Woodcock and Strahler (1987), have recently been used for the detection and automation of the selection of optimal scales (Dao et al., 2019b; Drăguț et al., 2014; Drăguț et al., 2010; Hay et al., 2005; Kim et al., 2008; Ming et al., 2015). Object's Local Variance (LV) and Rate of Change (RoC) graphs used the highest peak on the original graph of the RoC of the LV (Dao et al., 2019b; Drăguț et al., 2010), representing the abrupt changes in intra-segment homogeneity, as the optimal segmentation scale. The method has quickly become one of the most widely used scale selection methods in remote sensing in the past few years. Despite the promising segmentation results they have achieved, these algorithms rely upon the global statistics of all the averaged values of LV and RoC calculated from all segments in all bands, which makes these algorithms susceptible to outliers with extremely high LV values (Leys et al., 2013). In other words, the quantitative evolution of segmentation results at different scales is not accurately evaluated as information from the LV and RoC graphs is disrupted.

Furthermore, previous studies mostly used all original spectral bands of an image as input layers in the segmentation process and assumed an equal contribution of each band in segmentation algorithms. This could lead to biases when segmenting a remote sensing image given its tradition of collecting images at multiple spectral bands across the spectrum, including NIR and SWIR bands with a low signal-to-noise ratio, which should be weighted less in the segmentation models. However, to the best of our knowledge, no studies have attempted to tackle this problem in the literature.

Multispectral sensors that record electromagnetic energy at a few bands with wide spectral ranges produce images at lower noise levels. Hence, the impact of noise and outliers on scale selection and segmentation in multispectral images is usually less profound than in hyperspectral images. Most previous studies on optimal scale selection were conducted on gray-scale images (Espindola et al., 2006; Ming et al., 2015), colour images (Drăguț et al., 2010), multispectral images (Drăguț et al., 2014; Drăguț et al., 2010; Y. Wang et al., 2019; Yang et al., 2015; Yang et al., 2014), digital elevation model (DEM) data, or Lidar data (Drăguț et al., 2010).

In contrast, airborne HrHS images that offer various potential applications for different land cover types at different spectral regions (Bondur, 2014; Lu et al., 2020; Lu et al., 2019; Tarabalka et al., 2010) require specialized methods and algorithms for data processing and image analysis (Camps-Valls et al., 2013). Narrow spectral bands are able to delineate ground features and monitor their subtle changes over time, especially to detect and monitor vegetation health under early environmental stress (Dao et al., 2019a). However, the use of such high-dimensional data in a given analysis may induce noise (Liu et al., 2019b) and outliers that result in an inaccurate calculation of local and global measures such as the mean LV and RoC for the selection of optimal scales for image segmentation. Feature selection and feature extraction can be used to, to some extent, reduce dimension and noise and speed up the process. Feature selection is a process of selecting a few bands (B. Guo et al., 2006) from hundreds of bands. However, selected bands in different spectral regions still have noise at different levels, and they should not be considered equally in the scale selection and segmentation. Similarly, feature extractions such as Principal Component Analysis (PCA) (Rodarmel and Shan, 2002), autoencoders (Theis et al., 2017), flexible unsupervised feature extraction (FUFE) (Liu et al., 2019a), non-greedy L1-norm based 2-D linear discriminant analysis (L1-LDA) (Y. Liu et al., 2017) are the transformation of original features to a new feature space and a few components are then selected. However, extracted components usually contain various information and noises,

such as the first component of PCA contains more information and less noise compared to the subsequent components (Rodarmel and Shan, 2002), and they also should not be considered equally. Hence, it is essential to synthesize the input data, either they are original features, selected features, or extracted features, to reduce the effect of noise before segmentation and selection of optimal scales for hyperspectral images. However, optimal scale selection for segmentation with noise normalization and outlier removal in HrHS images remains an unexplored area.

1.3. Proposed research design

In this study, we propose a new multi-band optimal segmentation scale selection for hyperspectral images by applying inverse noise weighting and outlier removal. The measure we use in this study is the noise-normalized non-outlier rate of change (RoC) of the coefficient of variation (NN-nRoC). Specifically, hyperspectral imagery is segmented at different scales, and the Isolation Forest (IF) algorithm is applied to detect and remove outliers before calculating and constructing NN-nCV and NN-nRoC graphs. The proposed method is tested on airborne HrHS images of three study sites of various land cover types. The performance of this approach is evaluated on three widely used segmentation algorithms, including k-means, kernelized mean-shift, and compact watershed. These algorithms are well-established, stable, and computationally efficient that are suitable for segmenting hyperspectral data. In term of potential contribution to the literature, this is the first study that 1)

applies the noise inverse weighting to adjust the confounding effect of high-noise bands in image segmentation and scale selection; 2) removes the outliers in the calculation and use of local statistics of segments to select optimal scales for image segmentation; 3) determines the optimal scale to segment hyperspectral data; and 4) uses the coefficient of variation and its rate of change in optimal scale estimation produces more reliable segmentation results.

2. Data and methods

2.1. Overview

The step-by-step process to determine the optimal scale parameters and to assess the quality of image segmentation is provided in Fig. 1. The collected images were first orthorectified and registered to the base map and then radiometrically calibrated to obtain surface reflectance using the empirical line calibration method with the combination of ground reflectance reference collected using the *FieldSpec 3* spectroradiometer (Malvern Panalytical, Malvern, United Kingdom). The image pre-processing steps, including the radiometric calibration, orthorectification, and registration, were conducted using ENVI 5.3, and the details about these steps are described in Section 2.2. The images were normalized using inverse noise weighting and then segmented at different scales with k-means, mean-shift, and watershed approaches, followed by computing the CV and RoC values, and detecting and removing outliers for all segmentation scales. The NN-nCV and NN-

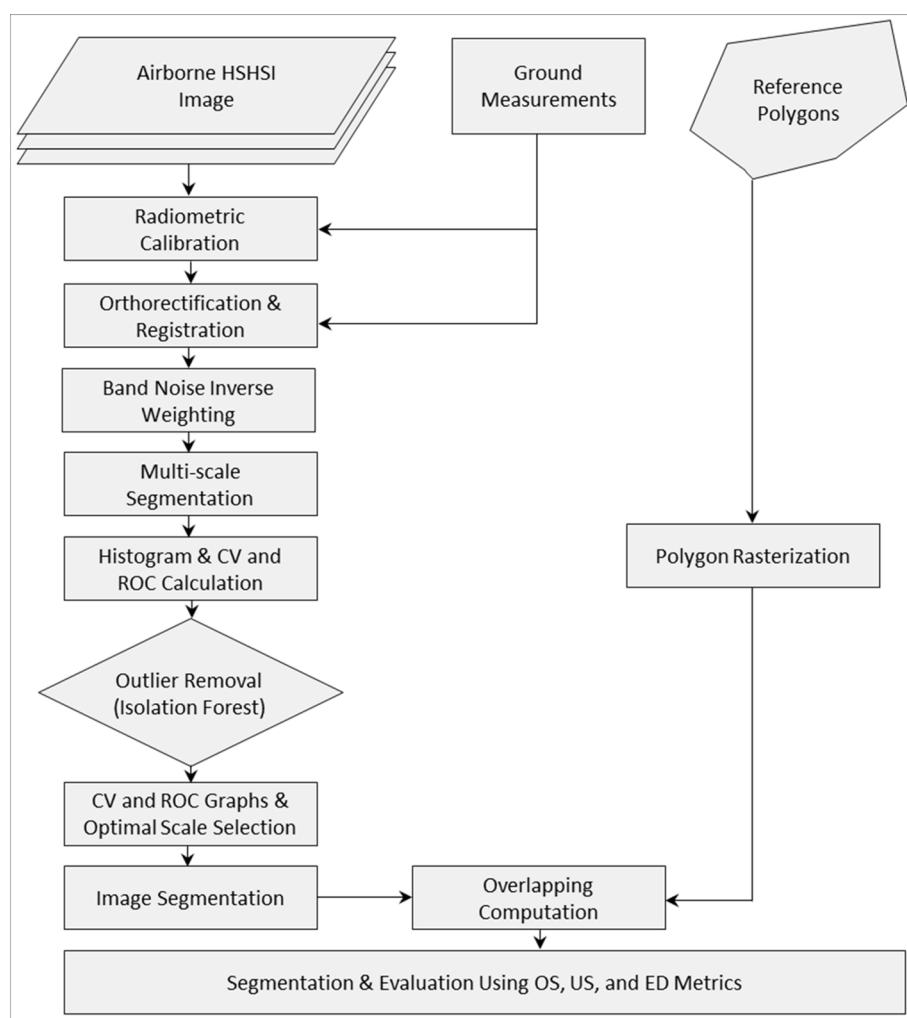


Fig. 1. The workflow for determining the optimal parameters and scales for image segmentation.

nRoC were calculated, and their graphs were then constructed for optimal scale selection. The selected optimal segmentation using NN-nRoC (NN-nRoC method) was evaluated using reference polygons, and compared to the results derived from the original images using RoC (RoC method). All these steps were conducted in Python 3 programming language using Scikit-image (Van der Walt et al., 2014), Scikit-learn (Comaniciu and Meer, 2002), and FAISS (J. Johnson et al., 2019) packages for image segmentation.

2.2. Study area and data

In this study, we used three hyperspectral datasets for the segmentation experiments. The first dataset is a suburban image captured around the Bolton area, which is an urban-rural transition town in Ontario, Canada. The area covers between $43^{\circ}52'32''$ and $43^{\circ}53'04''$ in latitude and $-79^{\circ}44'15''$ and $-79^{\circ}43'34''$ in longitude with the size of $190\text{ m} \times 215\text{ m}$. The site consists of various land cover types, including rooftops, asphalt roads, swimming pools, ponds, grassland, shrubs, urban forest, and shadows with both linear and non-linear shapes that make it an ideal scene for evaluating the performance of image segmentation. The second dataset is an urban image, also captured around the Bolton area. This image covers between $43^{\circ}45'30''$ and $43^{\circ}45'43''$ in latitude and $-79^{\circ}50'06''$ and $-79^{\circ}49'51''$ in longitude with the size of $165\text{ m} \times 360\text{ m}$. The area contains rooftops, under-construction houses, roads, lawn, shadows, which are appropriate for testing the performance of the approach on linear features. The third dataset is a natural forest located at an experimental biological site of the University of Toronto in King City, Canada, between $44^{\circ}01'58''$ and $44^{\circ}02'04''$ in latitude and $-79^{\circ}32'06''$ and $-79^{\circ}31'55''$ in longitude with the size of $255\text{ m} \times 203\text{ m}$. The image is ideal for evaluating the performance of the algorithm for segmenting tree canopies for species classification for forest inventory. The hyperspectral images of the study area and reference polygons for

segmentation validation are illustrated in Fig. 2.

The hyperspectral images were captured on 20 August 2017, around 10:30 am under a clear sky condition using the Mirco-Hyperspec VNIR push-broom sensor from Headwall Photonics Inc, USA. The images were projected to the North American Datum 1983 (NAD83) coordinate system, zone 17, with a spatial resolution of 30 cm. The original images of 325 bands ranging from 400 nm to 1000 nm were resampled to 301 bands with an interval of 2 nm. Since the signal-to-noise ratio is high in the NIR region from 902 nm to 1000 nm, bands in this range were removed. As a result, we used 251 bands from 400 nm to 900 nm for our purposes. To reduce the impact of noise on the image segmentation, the Savitzky-Golay filter (Savitzky and Golay, 1964) with a 3rd-order polynomial kernel function was used to smooth the spectral curves. The images were first orthorectified and registered to reduce the geometric distortion, and then atmospherically corrected using the empirical line calibration method proposed by (Dao et al., 2019a), to obtain surface reflectance. In addition, reference polygons over different land cover types were manually digitized to provide ground truth data that were used to evaluate the performance of any segmentation algorithm.

2.3. Band noise estimation and weighting

The bands in the NIR region usually exhibit a higher degree of noise, due to the effect of atmospheric water vapour, than that is seen in the visible region. While it is possible to disregard NIR bands entirely, these bands do provide important information that can be used to delineate different land cover types. Consequently, it is preferable to include these bands during the segmentation process, while treating the information in these bands carefully to minimize the adverse effects of poor signal-to-noise ratio (Liu et al., 2019b). In this study, we applied a band inverse noise weighting procedure to normalize spectral bands before inputting them into the segmentation. First, the noise level of each band was

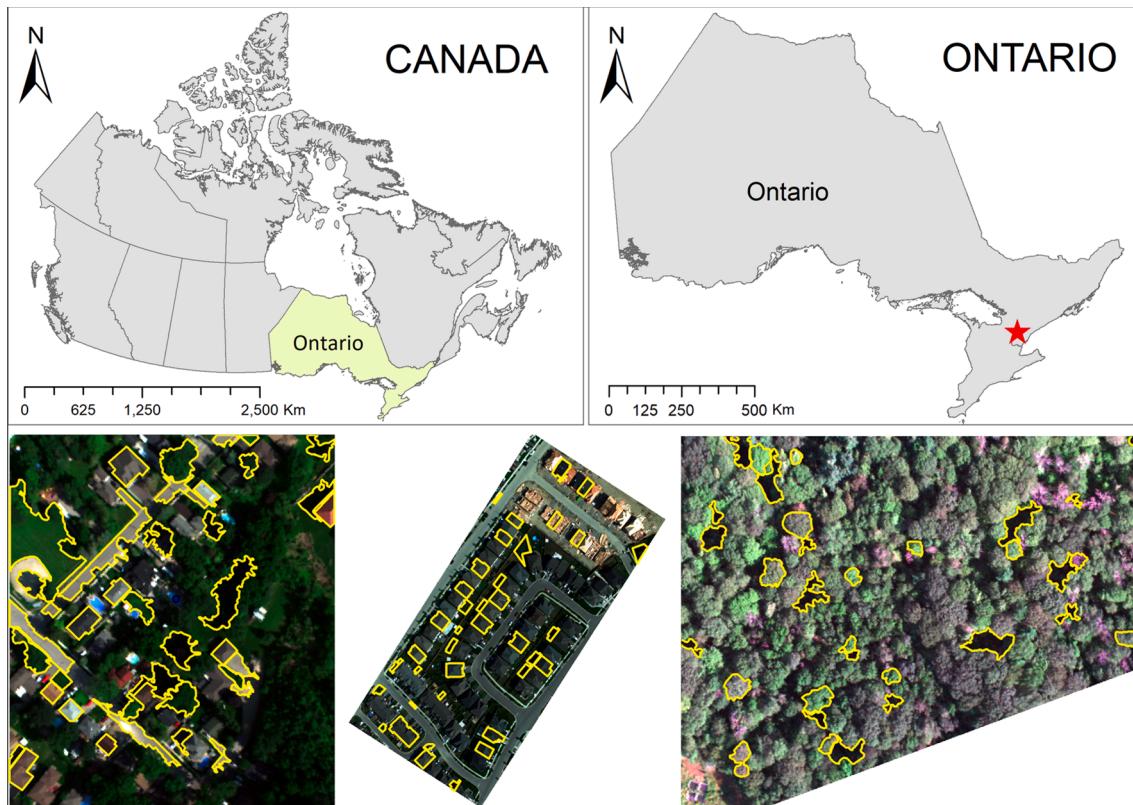


Fig. 2. Hyperspectral images (left bottom – the suburban image, middle bottom – the urban image, and right bottom – the forest image) and yellow-boundaries indicate manually digitized reference polygons used for segmentation validation in this experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

computed using a wavelet-based estimator that estimates the noise standard deviation with an assumption of Gaussian distribution of noise (Donoho and Johnstone, 1994). Second, the spectral value of each pixel in each band was multiplied with its corresponding noise inverse. The calculated noise level and weight for the three images are illustrated in Fig. 3. Noise varies in different spectral bands, with bands in the green and NIR spectral regions being noisier so assigned lower weights, while bands in the blue, red, and red-edge regions experiencing less noise then applied with higher weights. The curves depicted in this figure are random pixels sampled from each of the HrHS datasets with and without Savitzky-Golay filtering, then the band weights were used for all pixels.

2.4. Optimal scale selection for segmentation

An image can be segmented at various levels, and the target levels of interest depend on the spatial and spectral resolutions, the complexity of the scene, and the target application of the segmented images. The selection of a suitable segmentation scale is thus essential not only for creating meaningful image objects and improving the classification performance but also for avoiding unexpected over- or under-segmentation.

The optimal scale selection procedure proposed by (Drăguț et al., 2010), based on the LV graphs and RoC, is one of the most widely used scale selection procedures in remote sensing image segmentation in recent years. Optimal scale parameters are defined using the highest peak point of dramatic change in LV (in the form of absolute standard deviation) and RoC graphs (Dao and Liou, 2015). This method is easily implemented and reproducible that has been confirmed by an automatic scale selection tool that was created by Drăguț et al. (2014) using this procedure. However, the standard deviation is only an absolute measure of clustering (Sorensen, 2000) of individual segments; therefore it cannot be used for the comparison of homogeneity among segments, especially the homogeneity of neighbouring segments which is needed during region merging and region splitting steps in the segmentation process. Furthermore, the use of LV and RoC graphs without removing outliers may result in biases in the assessment of segmentation quality, leading to an inaccurate estimation of segmentation scales. To avoid these issues, we propose an improved method based on the non-outlier statistics and the curves of the coefficient of variation (CV) (also called the relative standard deviation) (Abdi, 2010; Weber et al., 2004) and its RoC. This method is expected to be a more reliable measure of dispersion for the selection of optimal scales for HrHS images.

In this study, the optimal scale selection for HrHS image segmentation proceeds as follows: 1) segmenting the HrHS image at different scales and calculating CV; 2) detecting and removing outliers using IF algorithm; 3) calculating and constructing NN-nCV and NN-nRoC graphs; 4) inspecting NN-nRoC graphs and selecting an optimal scale.

The CV of a segment in a one-band image is calculated as follows:

$$CV_p = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_p)^2}}{\mu_p}, \quad (1)$$

where CV_p is the CV of segment p , N is the number of pixels within this segment, x_i is the gray value of i^{th} pixel within the segment p , and μ_p is the mean value of all the pixels within the segment p . The averaged CV of a segment across all hyperspectral bands is expressed in the following equation.

$$CV_B = \frac{\sum_{b=1}^B CV_{p,b}}{B}, \quad (2)$$

where B is the number of spectral bands, and $CV_{p,b}$ is the coefficient of variation of the segment p in band b . The final averaged CV of all segments (P) in the image is calculated as follows:

$$CV_{avg} = \frac{\sum_{p=1}^P CV_{B,p}}{P} \quad (3)$$

After calculating CV_{avg} , segments with extreme values are removed from the calculation of the mean CV and the construction of CV and RoC graphs at all segmentation scales using the IF algorithm (F. T. Liu et al., 2012). The method is less computationally expensive and requires smaller memory. The method recursively builds an ensemble of isolation trees for a given dataset. Due to the susceptibility to isolation, abnormal observations are more likely to be isolated at a shorter distance to the root of an isolation tree compared to normal observations. IF randomly selects a feature and a splitting value between the minima and maxima. The splitting is processed if the observation is greater or smaller than the splitting value. The splitting process is stopped if the depth limit is reached. The algorithm returns an anomaly score for each observation.

$$S(x, \psi) = 2^{\frac{-E(h(x))}{c(\psi)}}, \quad (4)$$

where, ψ is the subsampling size, $h(x)$ is the path length of observation x , $c(\psi)$ is the average of $h(x)$ given ψ , and $E(h(x))$ is the average of $h(x)$ from a set of isolation trees. A threshold score of 0.5 is usually used in literature. A score greater than 0.5 indicates anomalies, a score smaller than 0.5 indicates normal observations, while a score equal to 0.5 indicates there are no distinct anomalies in the dataset. In this study, a score of 0.5 was used as the threshold for outlier detection. After removing the outliers, the NN-nRoC is calculated from the final NN-nCV using the following equation

$$NN - nRoC = \left(\frac{NN - nCV_n - NN - nCV_{n-1}}{NN - nCV_{n-1}} \right), \quad (5)$$

Where $NN - nCV_n$ and $NN - nCV_{n-1}$ are the averaged NN-nCV of all segments at scale n and $n-1$, respectively.

The final NN-nCV and NN-nRoC graphs were constructed for each of

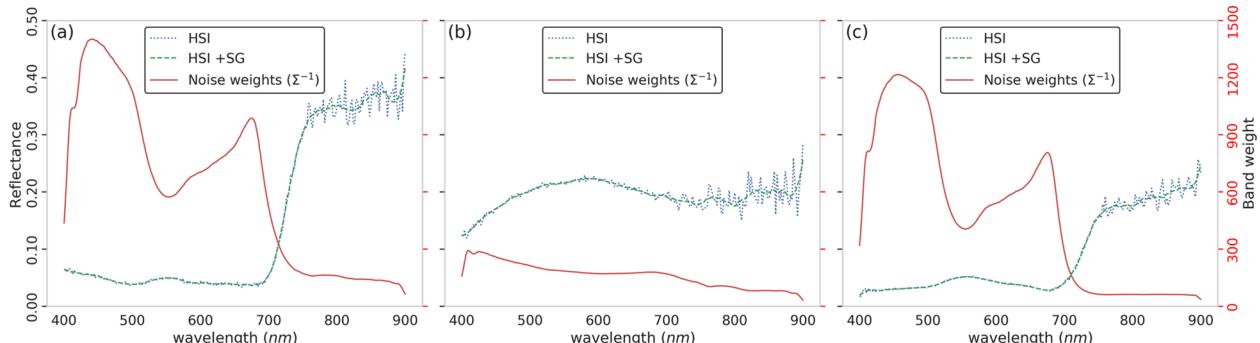


Fig. 3. The computation of noise and weights for each spectral band in three tested images. (a) for the suburban image, (b) for the urban image, and (c) for the forest image. HSI is the pixel's original spectral curve, HIS + SG is the smoothed spectral curve.

the three segmentation methods that we applied, and the change on the NN-nRoC graphs was used for optimal segmentation scale selection. Specifically, the peaks on the NN-nRoC graph where the NN-nCV changes abruptly indicate dramatic changes in intra-segment homogeneity and the segmentation scale. In this study, the first highest peak (Dao and Liou, 2015; Dao et al., 2019b; Drăguț et al., 2010) on the NN-nRoC graph was selected as the optimal scale.

2.4.1. K-means Segmentation

The k-means segmentation was implemented using the FAISS library developed by Facebook AI Research (J. Johnson et al., 2019). Despite being a simple algorithm, k-means clustering usually outperforms other segmentation methods while also offering high computing and memory efficiency. Given the number of pixels in an image $X \subset R^d$, k-means aims to find the number (k) of clusters C that minimizes the sum of squared distance between each pixel to its closest distance in the feature space (Arthur and Vassilvitskii, 2006; Kanungo et al., 2002). The minimal sum of squared distance between points and their closest centers is expressed in the following equation.

$$d_{min} = \sum_{x_i \in X} \min_{\mu_j \in C} (\|x_i - \mu_j\|^2), \quad (6)$$

where x_i is the gray value of the pixel i^{th} in the image, μ_j is the mean pixel value of the cluster j^{th} . The final clustering is obtained through a four-step iterative refinement technique that is inclined to find the minimum distance (Arthur and Vassilvitskii, 2006). The first step is to choose the initial centroids $C = \{c_1, c_2, \dots, c_k\}$ by choosing the number of clusters (k samples) from the dataset X . The second step is to assign each sample (image pixel) to its nearest centroid c_i in the feature space. The third step is to create new centroids using the mean value of all the samples assigned to each of the previous centroids. In the final step, the difference between the old and new centroids is computed, and the second and third steps are repeated until the difference is smaller than a pre-defined threshold or until the centroids do not move significantly.

The convergence of k-means may be resulted from a local minimum, depending on the initialization of centroids. To mitigate the sensitivity to the initial starting conditions, we usually conduct repeated computations with different initializations of centroids. For the empirical experiment of k-means, the number of centroids to generate is usually tuned to obtain different segmentation scales. After an investigation in the source-code of the FAISS library, we observed the initialization of centroids is based on the k-means++ algorithm (Arthur and Vassilvitskii, 2006), which reduces the chances of converging into local minima. Hence, the parameter k was used for tuning and the evaluation of segmentation results for the NN-nCV-based selection of optimal segmentation scale. The method is computationally efficient and can also work with GPUs, which is more suitable and very applicable to hyperspectral data due to the high-dimensionality on the feature vector.

2.4.2. Mean-shift segmentation

The mean-shift algorithm is an adaptive clustering method that was first introduced by (Fukunaga and Hostetler, 1975) and has been widely used in many applications, including image segmentation in remote sensing. The technique incorporates a non-parametric feature-space estimation of density to locate the maxima of density function (Ming et al., 2015). Given N data points x_i in a d-dimensional space R^d , the estimation of kernel density at a location x can be expressed by the following equation.

$$\hat{f}_{h,K}(x) = \frac{c_{k,d}}{Nh^d} \sum_{i=1}^N k\left(\left\|\frac{x-x_i}{h}\right\|^2\right), \quad (7)$$

where > 0 is the bandwidth parameter, and $K(x)$ is the radially symmetric kernel which can be expressed in the following equation

$$K(x) = c_{k,d}k(\|x\|^2) > 0, \|x\| \leq 1 \quad (8)$$

where $c_{k,d}$ is the normalization constant, and $k(x)$ is the profile of the kernel. For the density gradient estimation, assuming that the derivative of the kernel profile $k(x)$ exists for all locations x , using profile $g(x) = -k'(x)$, the kernel $G(x)$ can be defined as $G(x) = c_{g,d}g(\|x\|^2)$. The mean-shift vector is then calculated as follows.

$$m_G(x) = C \frac{\widehat{\nabla} f_K(x)}{\widehat{f}_G(x)} \quad (9)$$

where C is a positive constant at a location x , the mean-shift vector calculated with the kernel G is typically proportional to the normalized density gradient computed with kernel K . The final mean-shift vector is expressed as follows.

$$m_{h,G}(x) = \frac{\sum_{i=1}^N x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^N g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x, \quad (10)$$

As stated by Comaniciu and Meer (2002), this vector always points toward the direction of maximum density increase. The mean-shift algorithm is obtained by iterative computation of mean-shift vector $m_{h,G}(x)$ and the translation of the kernel $G(x)$ by $m_{h,G}(x)$. The iterative process is guaranteed to converge at nearby points where the gradient of the estimate is zero, which makes the mean-shift algorithm an adaptive gradient ascent approach (Comaniciu and Meer, 2002; Ming et al., 2015). The sequence of successive locations for the kernel G is expressed in the following equation.

$$y_{j+1} = \frac{\sum_{i=1}^N x_i g\left(\left\|\frac{y_j-x_i}{h}\right\|^2\right)}{\sum_{i=1}^N g\left(\left\|\frac{y_j-x_i}{h}\right\|^2\right)} j = 1, 2, \dots \quad (11)$$

When the single bandwidth parameter of the kernel is replaced by a multi-variate kernel, the formula (8) is re-written as follows.

$$K_{h_s, h_r}(x) = \frac{c}{h_s^p h_r^p} k\left(\left\|\frac{x_s}{h_s}\right\|^2\right) k\left(\left\|\frac{x_r}{h_r}\right\|^2\right) \quad (12)$$

where spatial bandwidth h_s is defined as the spatial window size in the segmentation, h_r is the spectral bandwidth, which represents the spectral distance between classes in the spectral domain, and p is the number of spectral bands in the image. Another parameter that can be incorporated in the segmentation process is the minimum spatial size (the number of pixel M) of patches that makes up a class. Accordingly, the parameters that determine the scale of mean-shift segmentation are the spatial size (h_s), the spectral size or compactness (h_r), and the minimum spectral size of patches in a class, and number of seeds. In this study, we used a k parameter that defines the number of seeds for the initialization and therefore the bandwidth of the density estimator. This parameter was tuned for the optimal scale selection. The mean-shift segmentation was implemented using the Scikit-learn library in Python 3.

2.4.3. Compact watershed segmentation

Since introduced in 1979 by Beucher and Lantuéjoul (1979), the watershed algorithm has been a powerful technique for image segmentation and is widely applied in the areas of computer vision, pattern recognition, and image processing. The algorithm views a gray-scale image as a topographic landscape that is flooded to find boundary lines for separating an image into a number of catchment basins of one minimum in the image (Tarabalka et al., 2010). Each distinct basin forms a different image segment. Compact watershed is an extension of the traditional seeded watershed segmentation algorithm with the incorporation of compactness constraint that helps turn the traditional

watershed segmentation from over-segmentation into superpixel and produces more regularly shaped (Neubert and Protzel, 2014). The process of multiband compact watershed segmentation includes: 1) per-band feature extraction through the calculation of gradients, 2) combining images of gradients, 3) compact watershed image segmentation, and 4) regions merging.

Instead of using a colour image as the input, watershed uses a grayscale gradient image that defines transitions between regions with high values at the segment borders and minima in homogeneous areas. In this study, we used the Sobel operator (Chen et al., 2016; Gupta and Mazumdar, 2013) to compute the 2-D spatial gradient image of each band in the HrHS images. As mentioned earlier, since the watershed approach is only applied to gray-scale images, the gradient images of all bands need to be combined before using the actual watershed algorithm. There are three common methods to combine gradient images: 1) vectorial gradient, 2) multidimensional gradient, and 3) segmentation map combination (Tarabalka et al., 2010). In this study, we used the multidimensional gradient images and combined them to test our method. Each band-gradient in an N -band hyperspectral image $\rho_E(x_\lambda)$, $\lambda = 1, 2, \dots, N$, were computed using the following equation.

$$\nabla_E^+(x) = \sum_{\lambda=1}^N w_\lambda \rho_E(x_\lambda), \quad (13)$$

where w_λ is the weight of the gradient of the band λ , computed as the inverse of the noise of band λ . The combining operation results in a single two-dimensional gradient image. The compact watershed segmentation (Neubert and Protzel, 2014) was then conducted at different scales by changing the number of markers while fixing all other parameters as default – compactness = 10^{-5} and gradient was calculated using Sobel filter in Scikit-image with a kernel size of 3×3 pixels. This algorithm was implemented using the Scikit-image library in Python 3.

2.5. Segmentation evaluation

Image segmentation result from an optimal scale is often evaluated through visual inspection by comparing it to the manually digitized reference polygons over different land cover types. However, qualitative visual inspection is labour-intensive and subjective, and the results vary when conducted by different technicians. In contrast, quantitative methods that measure the arithmetic and geographic discrepancy between segments (Yang et al., 2015) and reference polygons are more effective and can be repeated and automated easily in the process of segmentation quality assessment. Among those, the ED index (Yang et al., 2014), computed from OS and US, shows to be a more robust measure as it embeds both US and OS metrics. The ED index is calculated as follows.

$$OS_{i,j} = \left(1 - \frac{area(r_i \cap s_j)}{area(r_i)} \right), r_i \in R, s_j \in S \quad (14)$$

$$US_{i,j} = \left(1 - \frac{area(r_i \cap s_j)}{area(s_j)} \right), r_i \in R, s_j \in S \quad (15)$$

$$ED_{i,j} = \sqrt{\frac{US_{i,j}^2 + OS_{i,j}^2}{2}} \quad (16)$$

where r_i is the i^{th} polygon in the set R of reference polygons, and s_i is the corresponding segment from the set S of evaluated segments. Both OS and US are normalized between 0 and 1, ED, therefore, indicates both the geometric and arithmetic relationships. The smaller index values indicate good segmentation results and vice versa.

To evaluate the image segmentation results, we manually digitized 50 polygons with different shapes and sizes over different land cover types, such as rooftops, vegetation, and shadows. The US, OS, and ED

indices were evaluated over these land cover types to confirm the robustness of our approach for different datasets and applications.

For some of the scales, the segmentation algorithms can produce tiny objects (tens of pixels), which in turn affects the evaluation of the metrics. These small objects are generated due to the high variability of hyperspectral data and can lead to misinterpretation of the results since we are evaluating a small number of cover types. Therefore, we applied a threshold where only the pairs of segments-polygons with 10% overlap and segments equal to or greater than 9 pixels were selected for evaluation.

3. Results

3.1. Outlier removal and optimal scale selection

The Isolation Forest outlier detection results are illustrated in Fig. 4. It was computed from the parameters of the coefficient of variation function and calculated for all scales. For the demonstration purpose, we randomly selected only one scale for each algorithm – k-means with $k = 10$, mean-shift with seeds = 12, and watershed with markers = 1500. Also, since the coefficient of variation is the ratio of the standard deviation to the mean segment spectral value, the mean was used for the x-axis, and the standard deviation was used as the y-axis for visualization of the outlier detection results. Outlier segments with scores greater than 0.5 exists in all methods and images. The higher the segmentation scale, the more details the images are segmented, which leads to the creation of more outliers. This pattern confirms that it is common to have segments with extreme values in segmented images over different land cover types due to the high spectral variability in some regions in the images. This issue is particularly more threatening in hyperspectral image segmentation. The outliers with extreme values affect the graphs of the mean coefficient of variation and rate of change that lead to the inaccurate segmentation scale selection.

The segments assigned as outliers are shown in Fig. 5. As seen, outlier segments present in all segmentation methods and images. These segments were usually in areas with high degrees of spectral variability. Even though the outliers can exist in a small number, they may affect the scale selection results by its extreme CV values. The higher the segmentation scales, the more outlier segments are generated.

The forest dataset presented a larger number of outliers due to the high spectral variability between gaps and tree canopies and within canopies with multiple overstory and understory vegetation layers. However, the noise normalization step helped reduce the number of outlier segments for all datasets.

For the selection of the optimal segmentation scale, we used the NN-nRoC curves and also evaluated the influence of noise and outliers on the CV and RoC linegraphs across ranges of scales. Figs. 6, 7, 8 shows the boxplots of CV and NN-nCV at different scales in the first column and CV, NN-nCV, RoC, and NN-nRoC curves in the second column, where rows are for the suburban, urban, and forest images respectively. For the suburban image, the CV boxplots of all algorithms show a high number of outliers exist in segmented images across segmentation scales, ranging from 0 to 100 from k-means and mean-shift and from 0 to 10,000 in the watershed. From our experiments, the number of outliers increased with the scale parameters, the size and the complexity of the study area, the image resolution, and the spectral richness in all algorithms.

After detecting and removing the extreme values, the NN-nCV boxplots in these figures show the substantial influence of outliers on the calculation of the segments' coefficient of variation. The average curves (right-hand column), also show dramatic changes at different scales compared to the corresponding CVs. Despite the overall tendencies of the curves, the peaks, which are indicators of optimal segmentation scales, on the RoC and NN-nRoC graphs are different in both magnitude and position. This demonstrates the great influence of the band's noise and the extreme CV values of outlier segments on the curves. If the

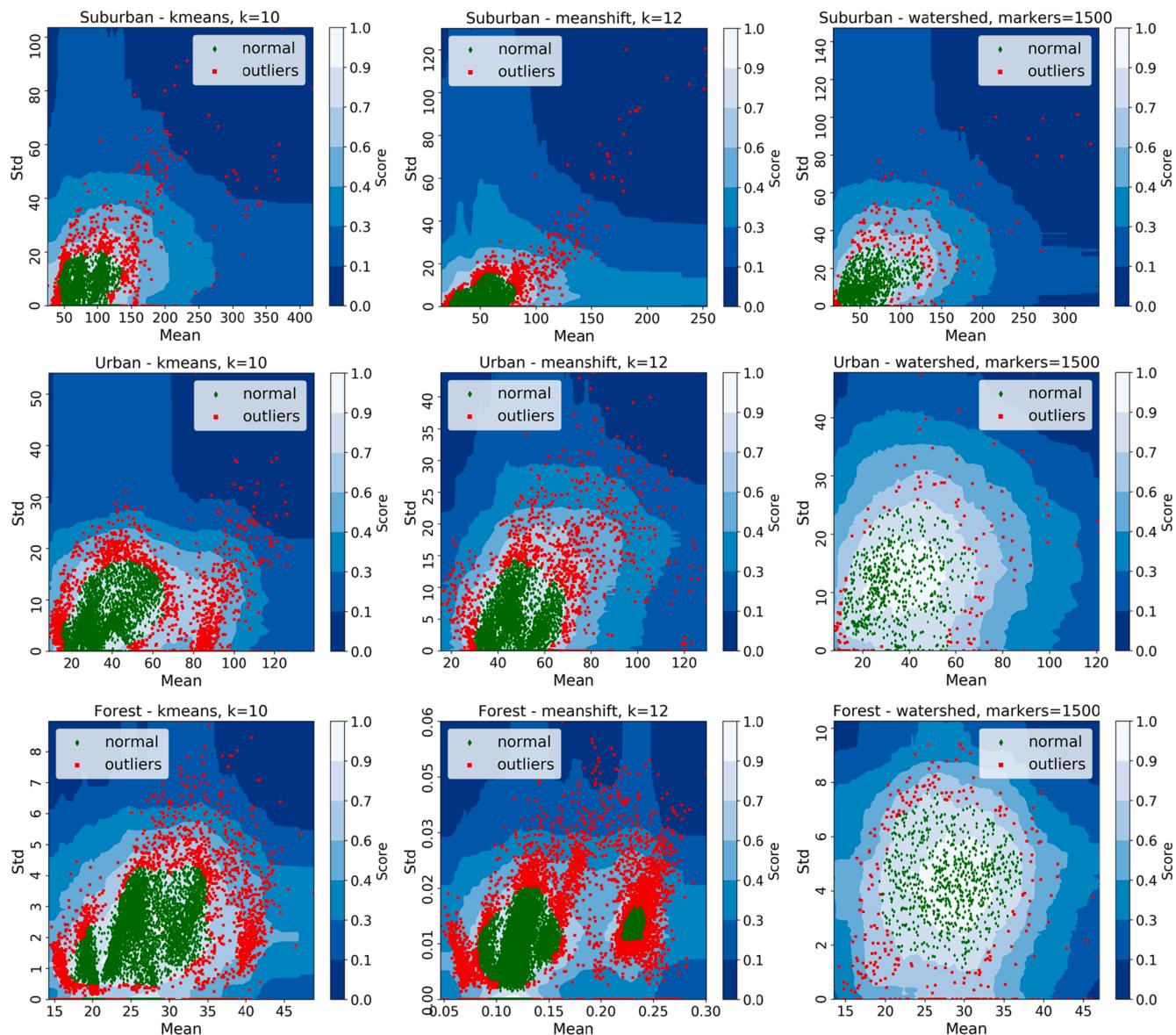


Fig. 4. Outlier detection results for the suburban (top figures), urban (middle figures), forest (bottom figures). The anomaly score is based on the parameters of a gaussian distribution $[\mu, \sigma]$, for all segments.

optimal scales are selected based on the original RoC graphs, the selected scales may not produce accurate segmentation results for the majority of the ground features in the images. Therefore, these curves should not be used for the scale selection as they misrepresent the actual quality of the segmentation process. In contrast, the NN-nRoC graphs represent the change in the coefficients of variation of the majority of segments in the segmented images across scales. The use of these curves for the optimal scale selection produced reliable and accurate segmentation results. Similar trends and patterns of these metrics, observed in the boxplots and line graphs of all suburban, urban, and forest images and all segmentation algorithms in Figs. 6, 7, and 8, indicate that noise and outliers are inevitable and influence the scale selection in images of different land cover types and different segmentation methods.

The optimal segmentation scales for each image were then selected at the first highest peaks on the NN-nRoC graphs, and the selected scales are shown in Table 1. The optimal scales, which were selected based on the RoC graphs, are also included in the table for comparison. It was clear that the selected optimal scales were different in most of the algorithms and images. This indicates the impacts of noise and outliers in the scale selection and the scale selection using the original image

statistics in previous studies may yield unreliable scales, which leads to inaccurate segmentation results. It should be noted that there was one exception in the k-means algorithm for the suburban image, in which the selected scales were not different ($k = 11$). This can be explained by the distinction of these peaks on the graphs that were not shifted under the impacts of noise and outliers. However, further exploration is required to examine whether or not the segmentation results are different in this particular case, and if positive, it would confirm the impacts of noise on the image segmentation, which should be removed before the segmentation step.

3.2. Image segmentation results

Final image segmentation results using NN-nRoC-based optimal scale selection, in Table 1, are illustrated in Figs. 9, 10, 11. The top figures are the results from RoC-based method, and the bottom figures are the results from the NN-nRoC method. Yellow polygons are manually digitized reference polygons for validation. For the suburban image (Fig. 9), both RoC and NN-nRoC methods worked well in all three algorithms that successfully delineated the boundaries between various

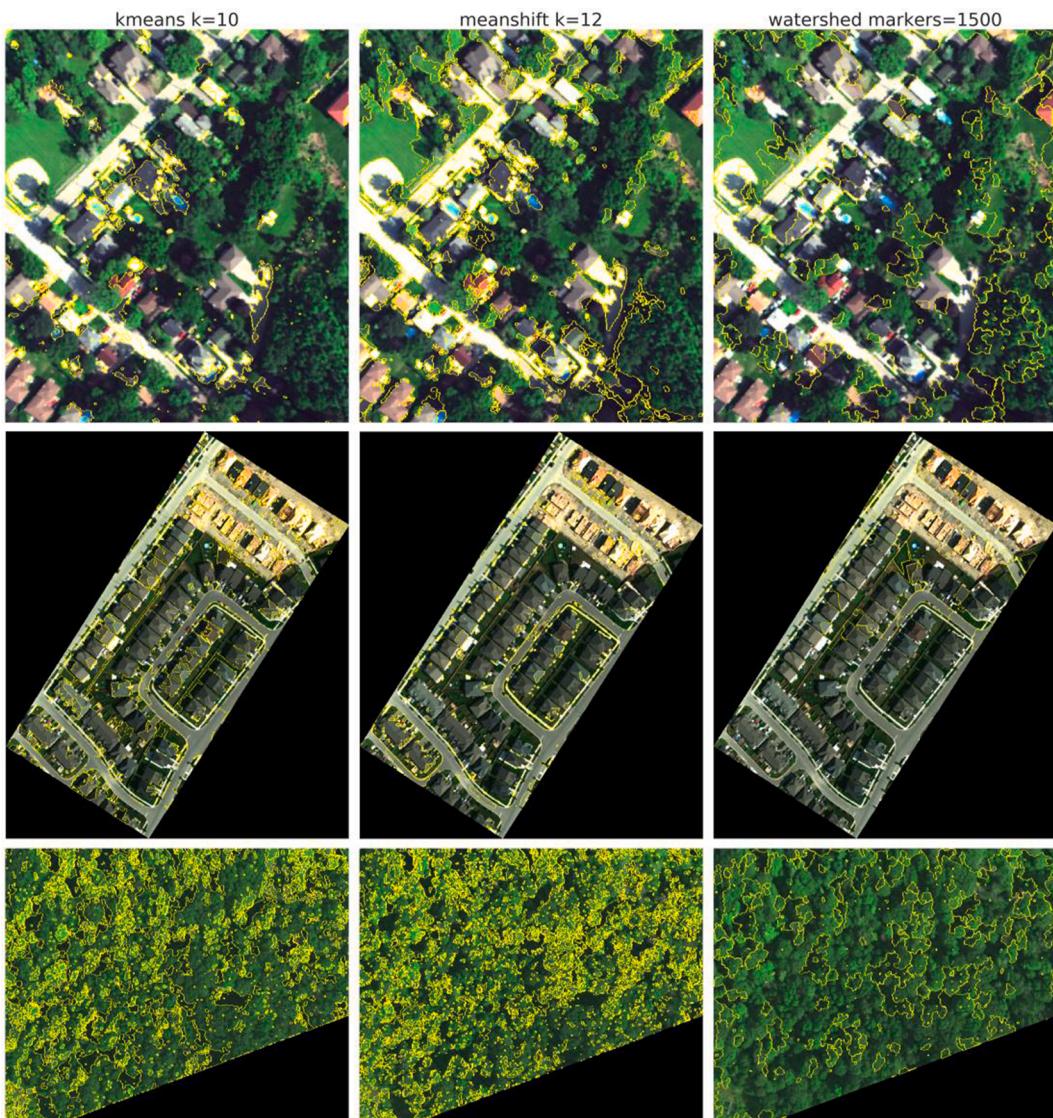


Fig. 5. Segments classified as outliers from the IF anomaly detection method. From left to right: k-means, mean-shift, and watershed segmentation method; from top to bottom: suburban, urban, and forest datasets.

land cover types. Some of the rooftops with very similar colours to the shadows in the true-colour image were also accurately distinguished in the segmented image. In this image, mean-shift seemed to perform the best since most of the ground features were accurately delineated with the least over- and under-segmentation in comparison with the other two algorithms. Watershed performed well with most features separated, and only slight miss-segmentation was observed in some places. K-means performed intermediately and successfully separated various land cover types with only slight over-segmentation is found in some areas. Between the two methods, NN-nRoC performed better than RoC methods in all three algorithms. In the RoC method, the image was over-segmented in k-means and mean-shift, especially in vegetated regions, while under-segmentation was observed in watershed segmentation results. More interestingly, the NN-nRoC method outperformed that the RoC method, even when the segmentation scales were the same ($k = 11$). This finding demonstrates the impact of noise in image segmentation results even it did not significantly influence the scale selection in this particular case. This finding also indicates the improvement of our proposed segmentation method.

For the urban image (Fig. 10), overall, mean-shift performed the best, followed by k-means and watershed. The mean-shift algorithm successfully delineated rooftops from lawns and shadows, while some

shadows were segmented to roof in k-means. However, over-segmentation was observed in spectrally varied areas, which made the segmentation results noisy, especially in the areas with under-construction buildings and bare soil on the north side of the image. Watershed seemed to perform well in these areas with less over-segmentation. Overall, the results from our proposed NN-nRoC method were better than the results from the RoC method in all three segmentation algorithms. Particularly in the watershed, the NN-nRoC outperformed the RoC method, and the image was segmented at a higher level of detail even the scale parameter was smaller (markers = 1300 vs. 1500). It should be noted that in the watershed, the higher the markers, the more details of the segmentation. This finding also indicates the improvement of our proposed method.

For the forest image (Fig. 11), the performance of these segmentation algorithms was not comparable to their performance in the suburban and urban images. The k-means and mean-shift algorithms performed well in separating tree canopies from the shadows with minimal over- and under-segmentation. The watershed performance was less efficient, especially in segmenting gaps and canopies with irregular shapes. From visual interpretation, it is clear that the NN-nRoC performed better than the RoC in all three segmentation algorithms. In the RoC method, over-segmentation in k-means and mean-shift, and under-segmentation in the

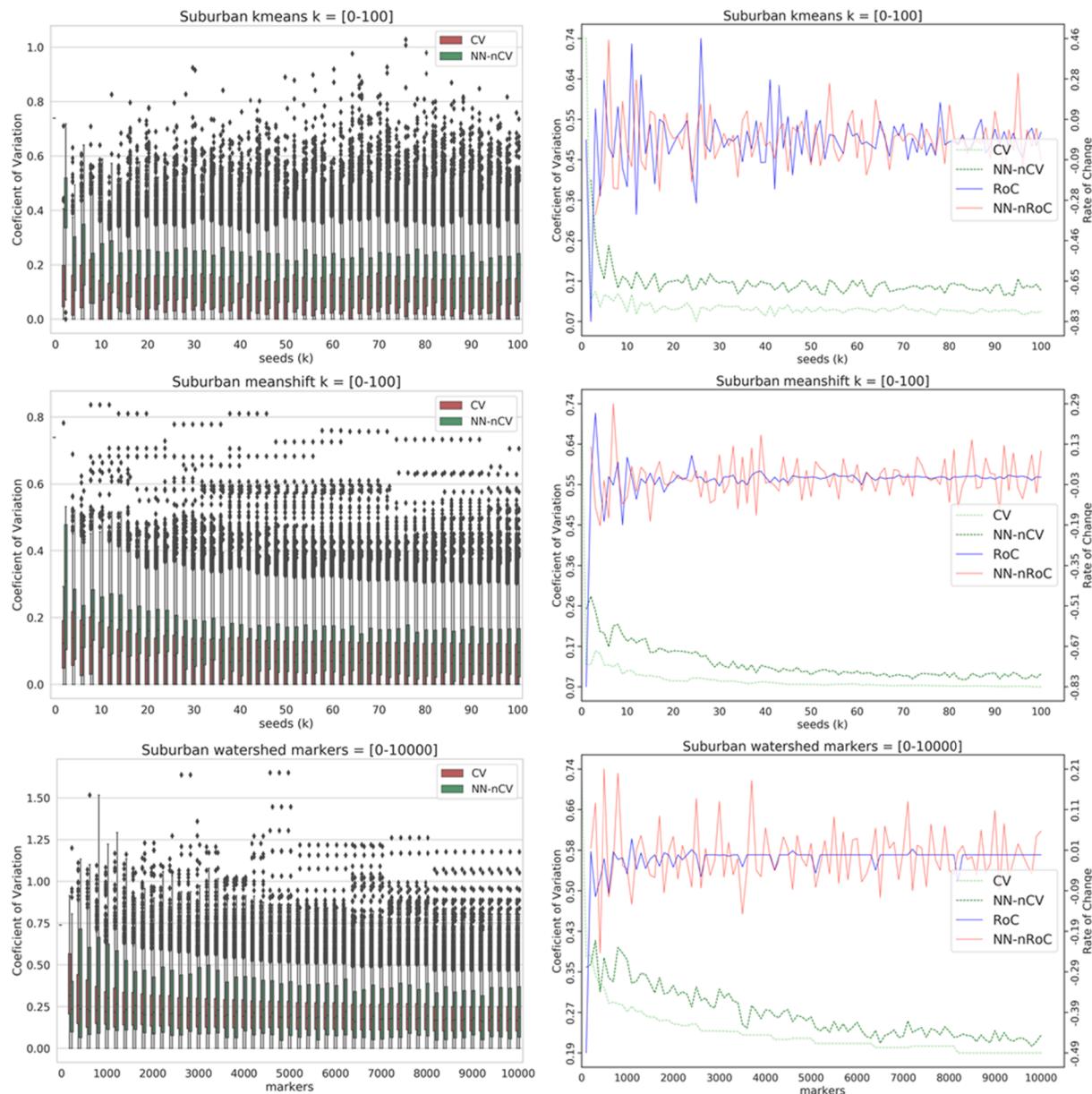


Fig. 6. The statistics of CV, NN-nCV, RoC, and NN-nRoC across large segmentation scales in all three segmentation algorithms for the suburban image. The left figures are the combined boxplots of CV and NN-nCV and the right figures are the combined CV, NN-nCV, RoC, and NN-nRoC of k-means ($k = 0\text{--}100$), mean-shift (seeds = 0–100), and watershed (markers = 0–10,000), respectively.

watershed were much more severe compared to the results from the NN-nRoC method.

3.3. Segmentation quantitative evaluation

For quantitative evaluation, the US, OS, and ED indices were calculated using the optimal scales in Table 1 to evaluate the fitness between image segments and reference polygons over different land cover types. The results from NN-nRoC were also compared to that produced from the RoC method in Figs. 12, 13, 14. The top figures are the results from RoC, while the bottom figures are the results of the NN-nRoC method. For US, OS, and ED, the smaller values indicate better performance of the algorithm. It should be noted that US and ED should be considered the most important measures in the evaluation, but over-segmentation since over-segmentation does not affect the subsequent image classification results.

The validation and comparison for the suburban image are shown in

Fig. 12. K-means and watershed of the NN-nRoC method with lower USs outperformed the results of the RoC method over different land cover types. The distinction was clear in the watershed, in which the over-segmentation was much higher, and the under-segmentation was lower in the NN-nRoC method compared to the results of the RoC method. This indicates the image was under-segmented more greatly in RoC method. The RoC approach performed slightly better in the mean-shift algorithm with lower US values. However, the EDs were higher than those in the NN-nRoC approach since the OSs were much higher. This indicates that the images were over-segmented in the RoC approach. Among the four ground features in this image, the segmentation of asphalt and rooftop seemed to be the worst in all methods. Overall, the NN-nRoC method performed better than the RoC method with lower USs, OSs, and EDs in most cases. This quantitative assessment was also in agreement with the earlier visual interpretation.

The validation and comparison for the urban image are shown in Fig. 13. It was clear that the NN-nRoC approach with much lower USs,

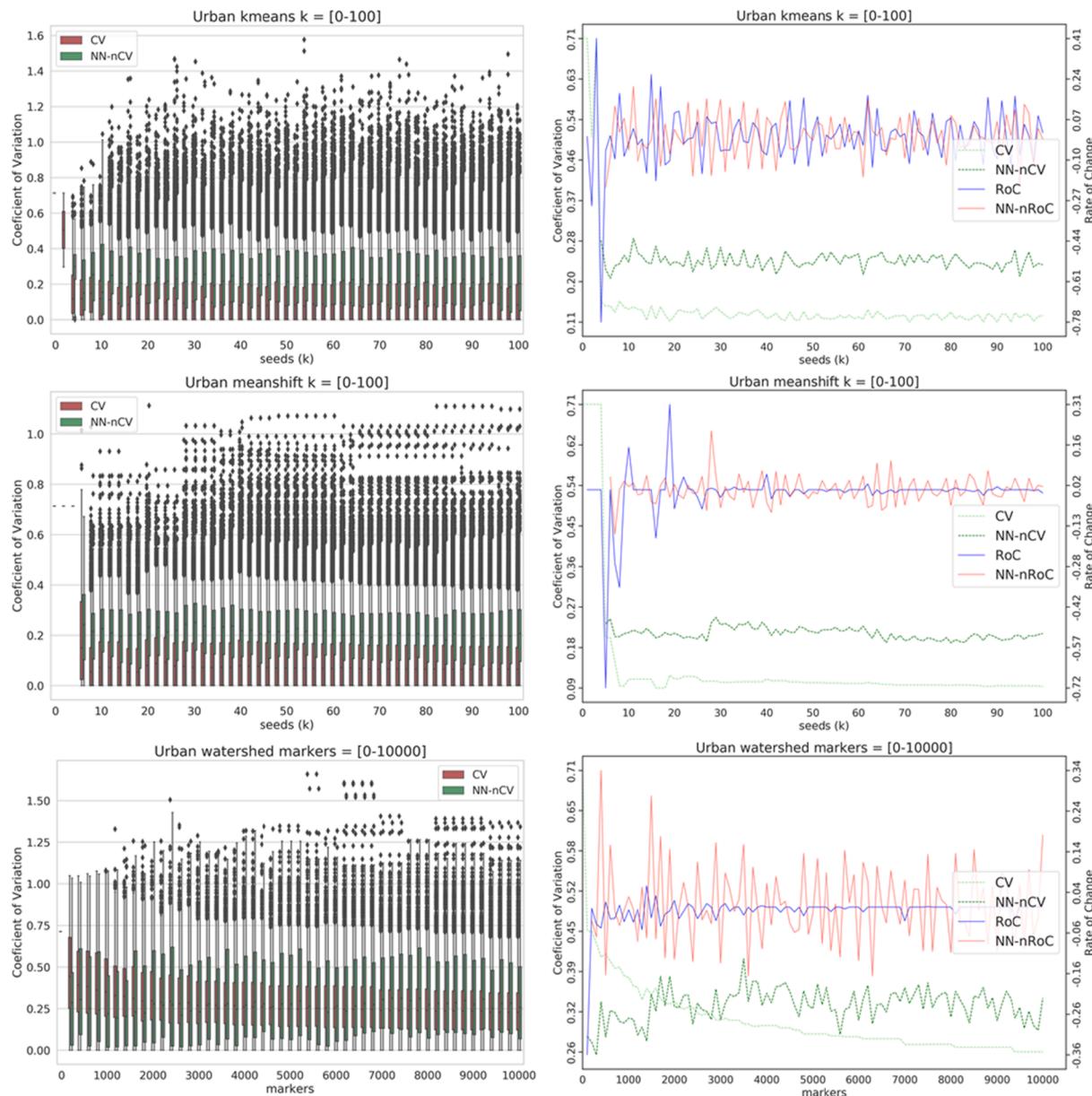


Fig. 7. The statistics of CV, NN-nCV, RoC, and NN-nRoC across large segmentation scales in all three segmentation algorithms for the urban image. The left figures are the combined boxplots of CV and NN-nCV and the right figures are the combined CV, NN-nCV, RoC, and NN-nRoC of k-means ($k = 0-100$), mean-shift (seeds = 0–100), and watershed (markers = 0–10,000), respectively.

OSs, and EDs outperformed the RoC approach in all algorithms and assessment measures. The over-segmentation was higher in k-means and mean-shift over rooftops. This would be because a roof with sun-facing and shaded sides with different spectral properties could be segmented into a different object, or the shaded sides could be grouped with shadows. However, as mentioned earlier, over-segmentation does not affect the subsequent image classification results. Over different land cover types, the results of rooftops were the most accurate with low US values, followed by that in lawns and shadows. The under-segmentation in shadows can be explained by the improper segmentation of a portion of shadow objects to the shaded sides of buildings with similar spectral properties. Under-segmentation of shadows was particularly high (USs greater than 0.9) in k-means and mean-shift of the RoC method. It should be noted that many small objects were created in shadows in k-means, making them much smaller than the reference polygons. Hence, shadow segments of k-means were not evaluated for this image.

The validation and comparison of the forest image are shown in

Fig. 14. As seen, these algorithms performed well in separating tree canopies from shadows. The segmentation results of the NN-nRoC method were exceptional in k-means and mean-shift with low USs (most USs were smaller than 0.2) and EDs (smaller than 0.4), especially in segmenting shadows. The segmentation was less efficient in the watershed with high USs (0.61 for trees and 0.39 for shadows) and EDs (0.35 for trees and 0.38 for shadows). Surprisingly, the forest image was better segmented compared to the suburban and urban images, demonstrating the effectiveness of our approach in segmenting a complex natural forest scene. It was also difficult for human eyes to accurately delineate individual tree canopy in this heterogeneous natural forest with sunlit and shaded portions within canopies. This finding highlights the capability of our approach and the hyperspectral data in delineating small but spectrally distinct objects in a complicated landscape. It is helpful when used in species and individual tree classification in a heterogeneous vegetated region. Again, the segmentation results of the NN-nRoC method were better than those of the RoC method in most

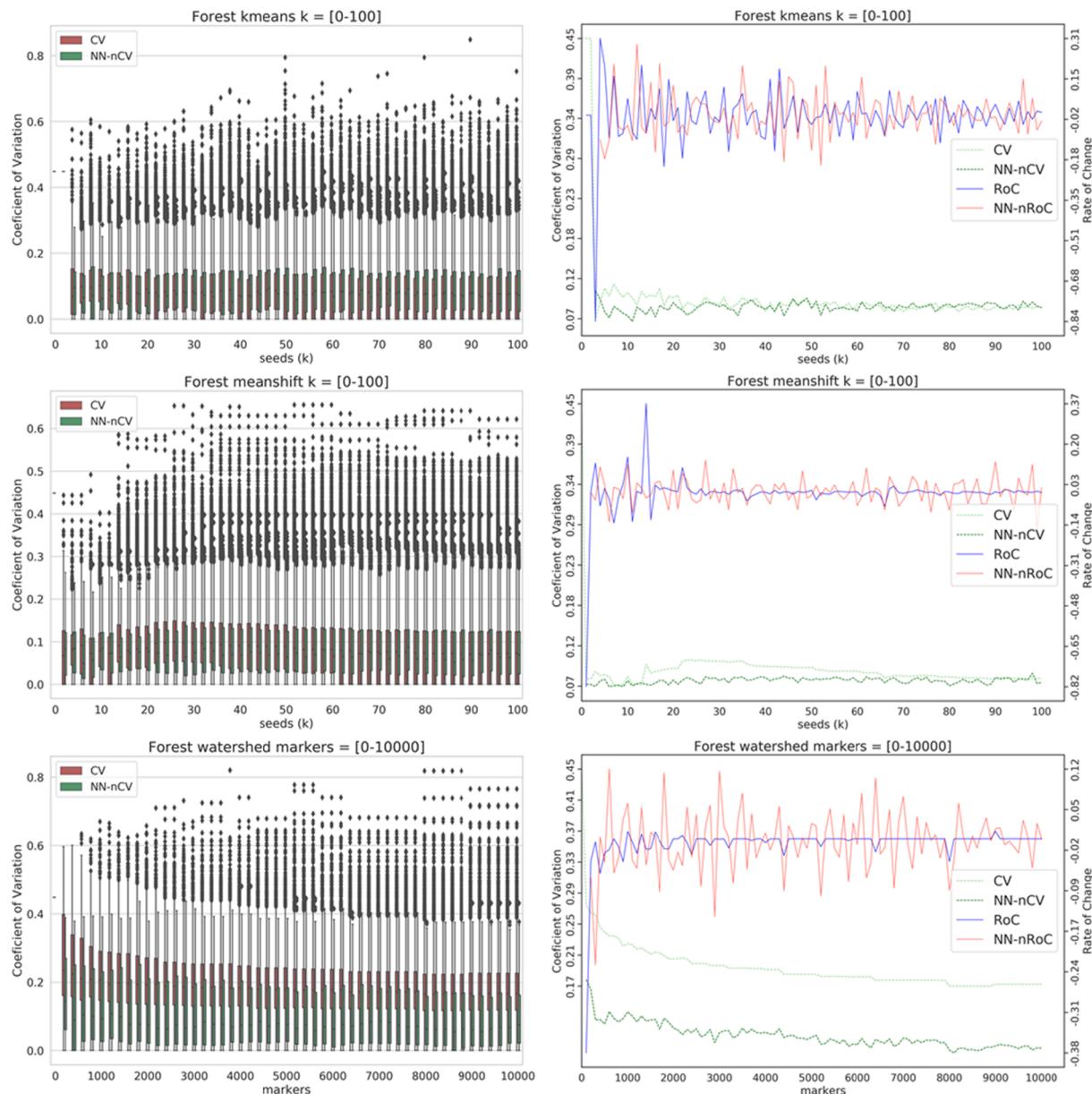


Fig. 8. The statistics of CV, NN-nCV, RoC, and NN-nRoC across large segmentation scales in all three segmentation algorithms for the forest image. The left figures are the combined boxplots of CV and NN-nCV and the right figures are the combined CV, NN-nCV, RoC, and NN-nRoC of k-means ($k = 0\text{--}100$), mean-shift (seeds = 0–100), and watershed (markers = 0–10,000), respectively.

Table 1

The optimal scales of k-means, mean-shift, and watershed of the three images using RoC and NN-nRoC graphs.

	Using RoC Graphs			Using NN-nRoC Graphs		
	KM	MS	WS	KM	MS	WS
Suburban Image	11	15	500	11	7	1,100
Urban Image	4	8	1,500	6	26	1,300
Forest Image	19	15	500	11	10	1,500

segmentation algorithms and ground features.

The overall evaluation of the entire image, including the US, OS, and ED over all segmentation algorithms, is illustrated in Fig. 15. The top figures are the results of RoC, while the bottom figures are the results of the NN-nRoC method. For the NN-nRoC method, the suburban image was best segmented with the lowest USs (0.12–0.21) and EDs (0.32–0.37), despite its high OS (0.52–0.65) values across three methods. The

segmentation performance in the forest image was intermediate in k-means and mean-shift with medium USs (0.16–0.19) and EDs (0.30–0.31), except in the watershed with US = 0.49 and ED = 0.36. The segmentation accuracy of the urban image was slightly lower; however, the over- and under-segmentation were more balanced – USs (0.22–0.36), OSs (0.26–0.49), EDs (0.27–0.36). For comparison, the segmentation results of the NN-nRoC method were improved compared to those from the RoC method in all images and algorithms, especially in mean-shift and watershed and in the urban and forest segmentation results. Overall, despite mild variation in these indices over different methods and images, the segmentation performed well with low US, OS, and ED indices in most cases. This quantitative evaluation suggests that by applying inverse noise weighting and outlier removal, our newly proposed method produced greatly improved segmentation results, with high quality and reliability, compared to previous studies.

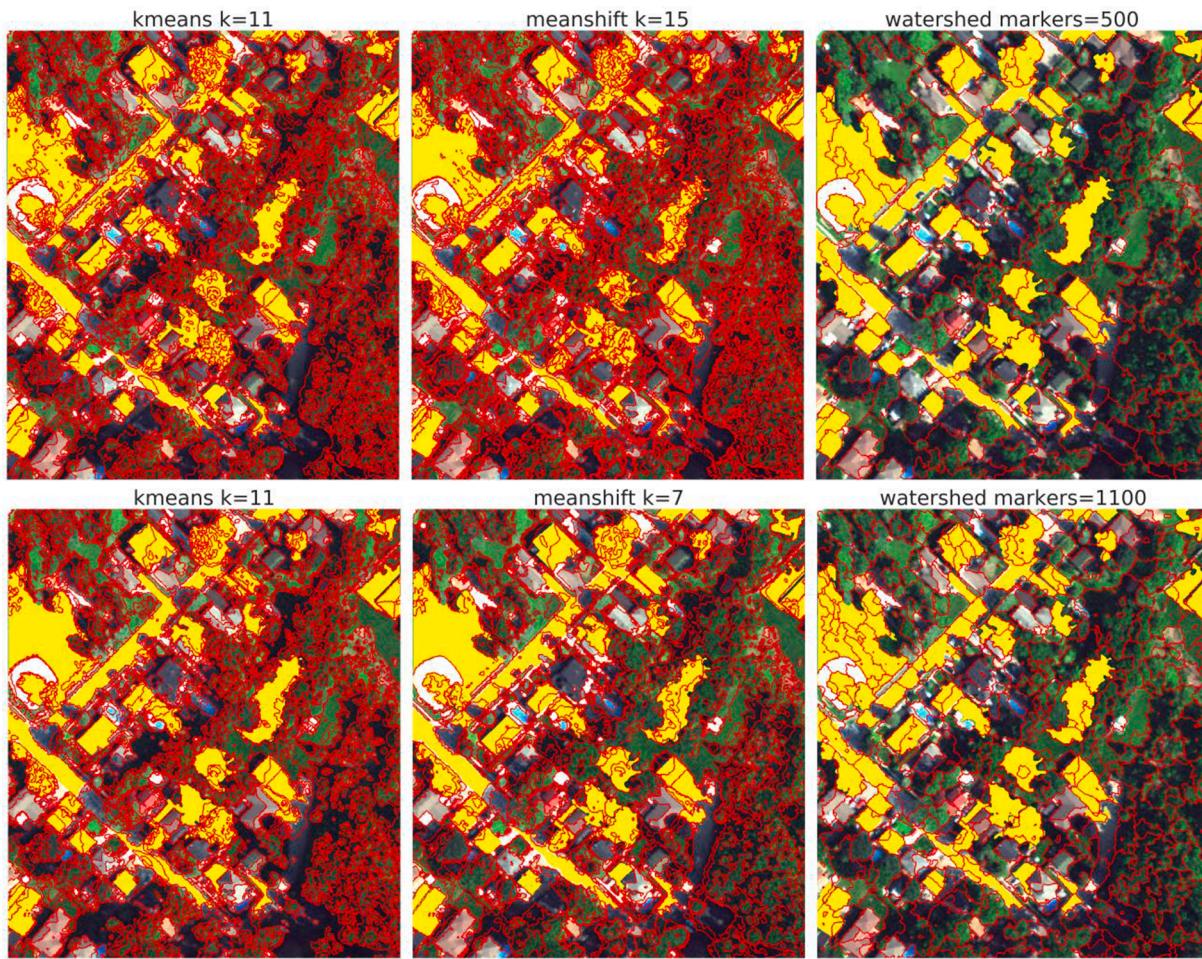


Fig. 9. Segmentation results of the suburban image using optimal scales selected in Table 1. From top to bottom are the results of RoC and NN-nRoC methods, respectively. From left to right, the original image with reference polygon, k-means, mean-shift, and watershed segmentation results. Yellow polygons are manually digitized reference polygons for validation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Scale selection performance

The computing performances of the segmentation algorithms on the three datasets are shown in Fig. 16. As expected, k-means run faster by a great margin (269, 394, and 301 s for the suburban, urban, and forest images, respectively), corroborating the assumption that this method is of one the most suitable for hyperspectral, high resolution and large scales image processing. Mean-shift performances were intermediate while watershed is the most computationally expensive algorithm with the slowest performance observed in the urban image (2,861 s). The larger runtime of the watershed algorithm was due to the processing the gradient images for each band before the segmentation process, which made the complexity of this algorithm in the order of $m^*O(n^2)$, where m is the number of spectral bands and n the size of the image, therefore the performances in different images were controlled by not only the size but also the number of bands and segmentation method. For example, mean-shift ran faster in the urban image than in the forest image while watershed experienced the opposite trend.

To confirm the performance of our method, we also compared the results with the widely used estimation of scale parameter (ESP) method for the multiresolution segmentation (MRS) method (ESP-MRS), developed by Drăguț et al. (2014), in eCognition software – an industry standard OBIA software in remote sensing. The major parameters of this method are scale parameter, shape, and compactness. In this experiment, shape = 0.1 and compactness = 0.5 were used as the default, and scale parameter, the deciding factor of MRS, were automatically tuned

for the best value using the ESP tool – based on the procedure (optimal scale is at the highest peak on the graph of RoC of LV) proposed Drăguț et al. (2010). The comparison is shown in Table 2. As seen, US values of ESP-MRS were much lower than those of our NN-nRoC method, indicating the under-segmentation in ESP-MRS was marginal. However, over-segmentation in the ESP-MRS method was much more severe (OSs were much higher) than in our NN-nRoC method, leading to higher ED values. This indicate that ESP-MRS was not good at balancing the over- and under-segmentation.

In terms of computation efficiency, our methods also ran much faster and were more applicable compared to the MRS in the eCognition software – runtimes are 18,900, 43,8001, and 4,201 s for the suburban, urban, and forest images, respectively. This comparison indicates that despite being a popular method and offering good segmentation results, the scale selection for MRS was computationally expensive (even for these three small images) and it may not be good to use it for large hyperspectral images in practice.

4. Discussion

4.1. General discussion

In this study, we did not separately evaluate the impacts of noise and outliers on the scale selection because the scale selection was influenced by the combined effects of these two factors. Specifically, on the one hand, noise affected the image segmentation (through the input of high-

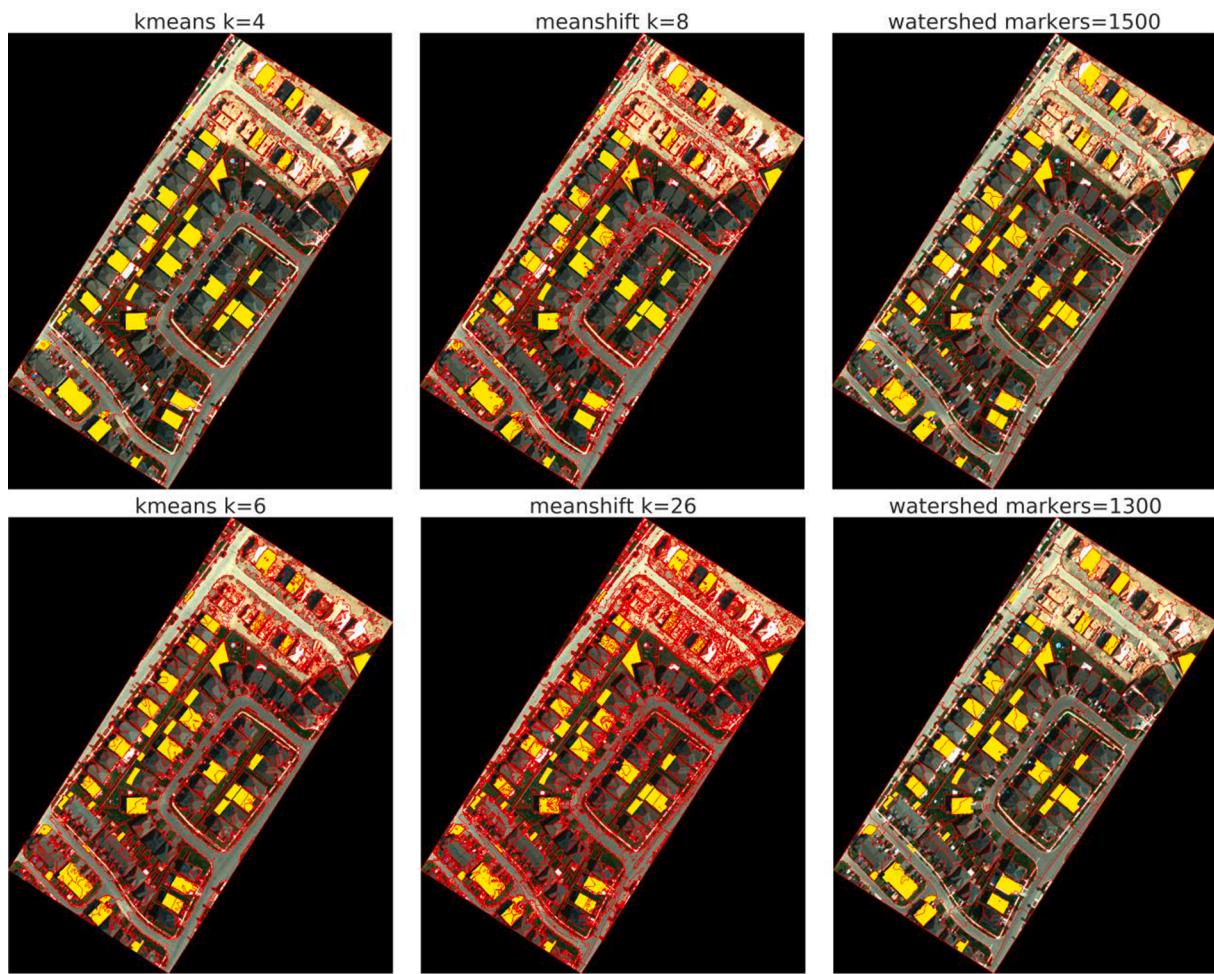


Fig. 10. Segmentation results of the urban image using optimal scales selected in Table 1. From top to bottom are the results of RoC and NN-nRoC methods, respectively. From left to right, the original image with reference polygon, k-means, mean-shift, and watershed segmentation results. Yellow polygons are manually digitized reference polygons for validation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

noise bands). On the other hand, noise also caused an increase in the number of outlier segments across scales, leading to the bias in the calculation of averaged NN-nCV and NN-nRoC. Meanwhile, outliers (segments with extreme NN-nCV) solely affected the scale selection through the bias on the calculation of the averaged NN-nCV and NN-nRoC. However, even the impacts of these two factors were not separately evaluated, the individual impact of noise can be seen in k-means segmentation of the suburban image, in which the NN-nRoC approach performed better than the RoC method despite the segmentation scales were the same ($k = 11$). The assessment of the combined effect of both factors was to make the method more straightforward and easier for practical use.

Our NN-nRoC method is useful for data with high noise, especially for hyperspectral images (He et al., 2015), Lidar (Fang and Huang, 2004), or Radar data (Dekker, 1998). The normalization of noise and the removal of outliers before optimal scale selection and segmentation would improve the segmentation and subsequent classification results. In addition to these data, the approach can also be applied for digital elevation model (DEM) data or multispectral optical imagery with lower noise levels. However, extensive tests on large data may be required to evaluate how noise and outlier affect the scale selection and segmentation and how the results are improved by solving these problems using this NN-nRoC method. This scale selection method could be easily reproducible because it is an unsupervised method that is solely based on the NN-nRoC of NN-nCV of the segmented image with the quantitative determination criterion – the peak with the greatest change in the

RoC graph. With this criterion, the method can be conducted without the involvement of expert knowledge. This optimal scale selection procedure can also be easily automated, similar to the above-mentioned method that was developed by Drăguț et al. (2014).

In comparison with the ESP-MRS method proposed by (Drăguț et al., 2014), our NN-nRoC method exhibited slightly higher under-segmentation; however, it achieved a better balance between over- and under-segmentation. The primary purpose of this study is to propose a new scale selection method that incorporates the noise normalization and outlier removal to produce more reliable, optimal scales for segmenting hyperspectral imagery. We used widely-used segmentation techniques to study the proposed segmentation scale selection method that applied both noise normalization and outlier removal. We used the RoC and NN-nRoC metrics to demonstrate the benefits of using our method for segmentation scale selection. The ESP scale selection for the MRS method was infeasible to use on hyperspectral imagery due to computational considerations. Runtimes presented in Table 2 attest to this observation. In hyperspectral image analysis, the trade-off between the scale selection accuracy and the computation efficiency must be considered to choose the most suitable approach that meets the technical requirements while it can be applied to the segmentation and classification of large images in large-scale studies. The consideration for these two factors depends on the applications and the computational resources available for the project. This study suggests that the NN-nRoC method for k-means, mean-shift, and watershed balances between accuracy and computation efficiency and is more applicable for

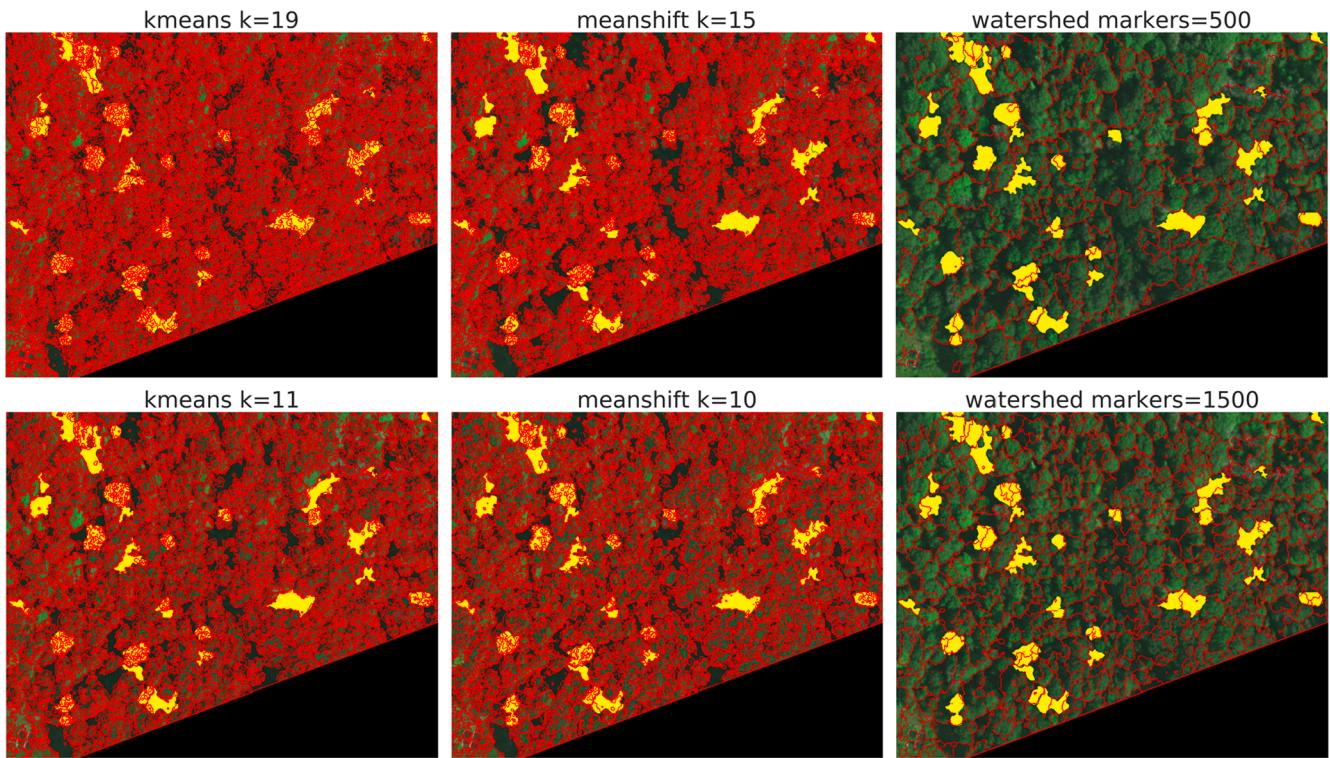


Fig. 11. Segmentation results of the forest image using optimal scales selected in Table 1. From top to bottom are the results of RoC and NN-nRoC methods, respectively. From left to right, the original image with reference polygon, k-means, mean-shift, and watershed segmentation results. Yellow polygons are manually digitized reference polygons for validation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

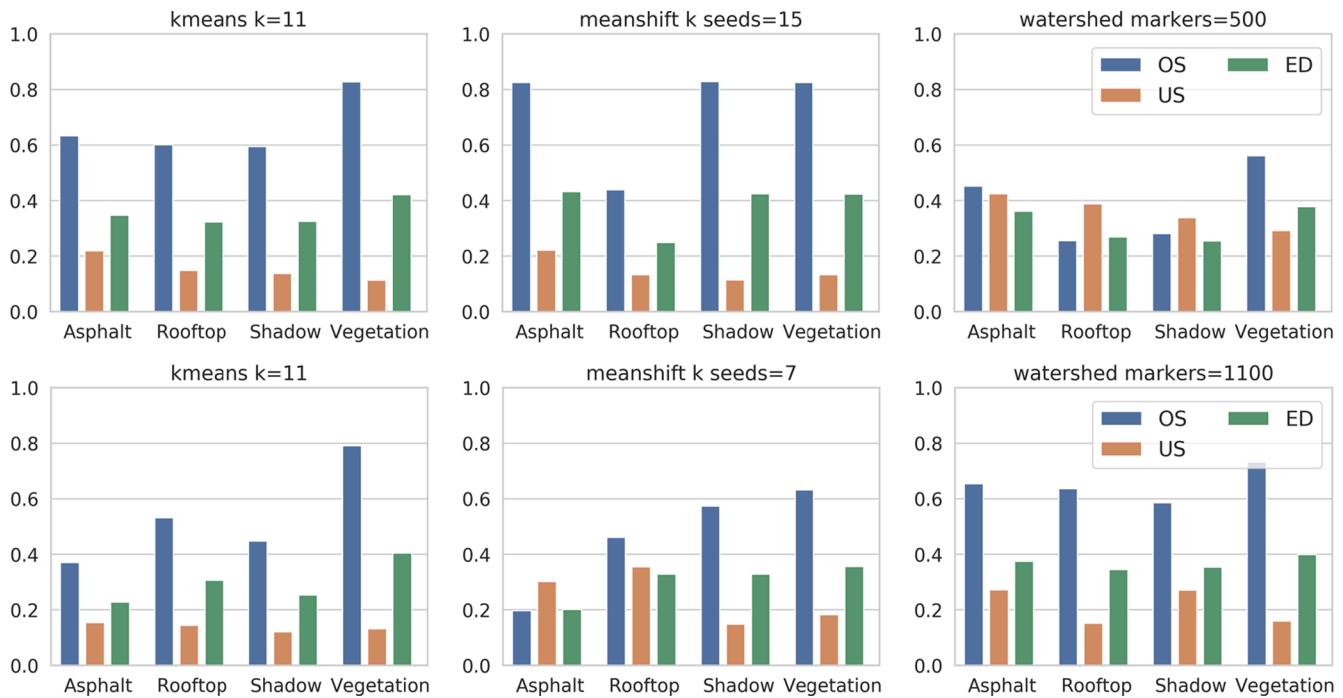


Fig. 12. Segmentation evaluation statistics of US, OS, and ED over different land cover types by comparing to reference polygons for the suburban image. From top to bottom are the results of the RoC and NN-nRoC methods, respectively.

hyperspectral images than the ESP tool for the MRS algorithm.

It should be noted that the NN-nRoC method uses the noise-normalized and outlier-removed averaged values of NN-nCV and its NN-nRoC of all bands and segments in the entire image for the scale selection. This global measure may not be able to represent local

variation and distribution of the NN-nCV in each band and across all bands. Advanced machine and deep learning algorithms with a more robust local measure may be used to understand the local NN-nCV pattern that may lead to a more robust scale selection method for hyperspectral images in future studies.

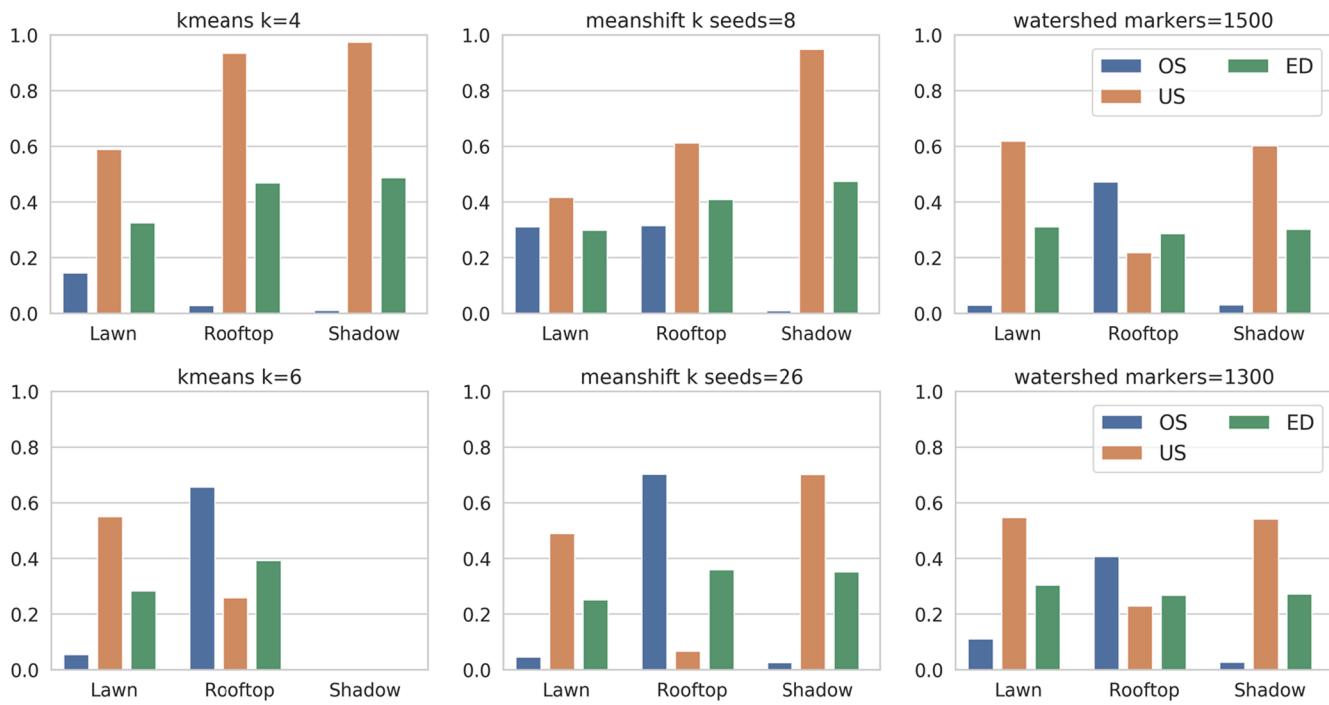


Fig. 13. Segmentation evaluation statistics of US, OS, and ED over different land cover types by comparing to reference polygons for the urban image. From top to bottom are the results of the RoC and NN-nRoC methods, respectively.

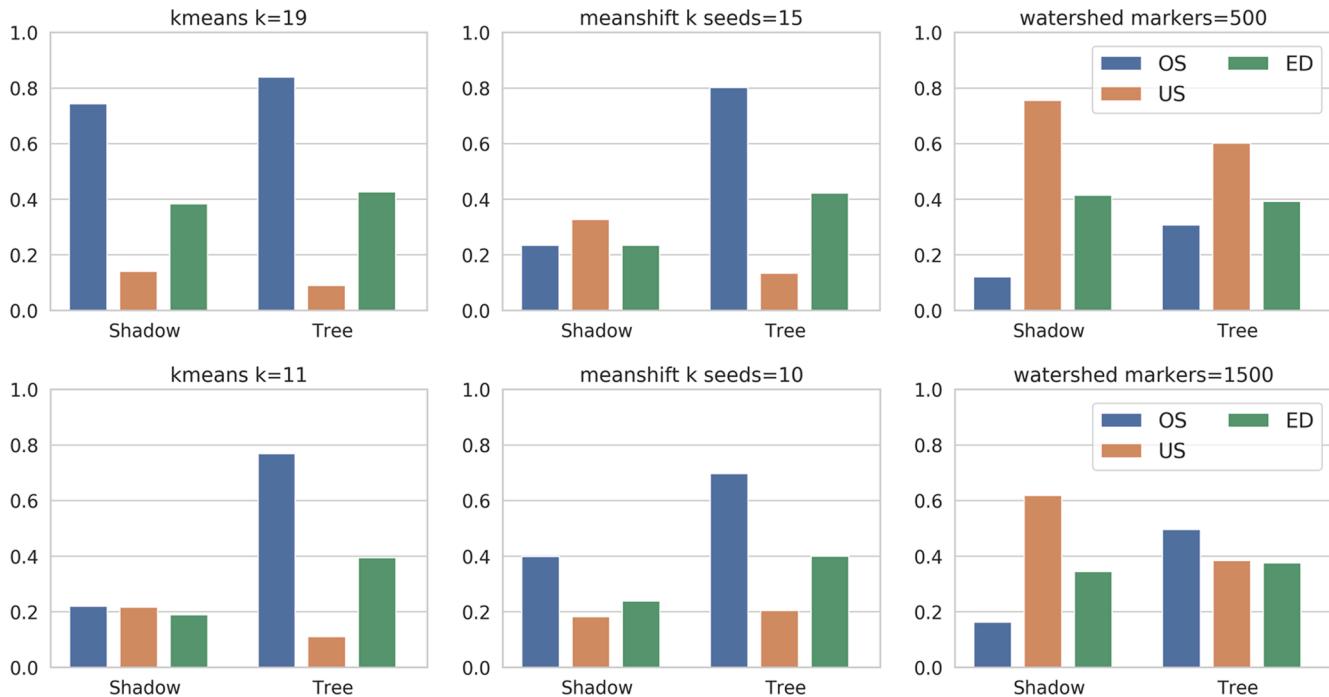


Fig. 14. Segmentation evaluation statistics of US, OS, and ED over different land cover types by comparing to reference polygons for the forest image. From top to bottom are the results of the RoC and NN-nRoC methods, respectively.

4.2. Computational complexity

The algorithm pipeline is composed of several stages: 1) segmentation, 2) aggregation of segments (statistics), 3) outlier detector, and 4) scale selection. Each stage has its own computational complexity study.

The k-means and mean-shift algorithms are known to have a time complexity of $O(n^2)$, while watershed is known to have $O(n)$, where n is the size of the image. Despite the linear computational complexity of the

watershed algorithm, its implementation depends on several internal functions to delineate the watershed boundaries based (derivative filters), followed by the dilation function in the search for homogeneous pixels. Implementation of the watershed for hyperspectral datasets has a complexity of $b^*O(n)$, where b is the number of bands (Kornilov and Safonov, 2018). The aggregation algorithm is also linear complexity and depends only on the number of segments from the previous step. The computational complexity can be resumed to $2^*O(s)$, where s is the

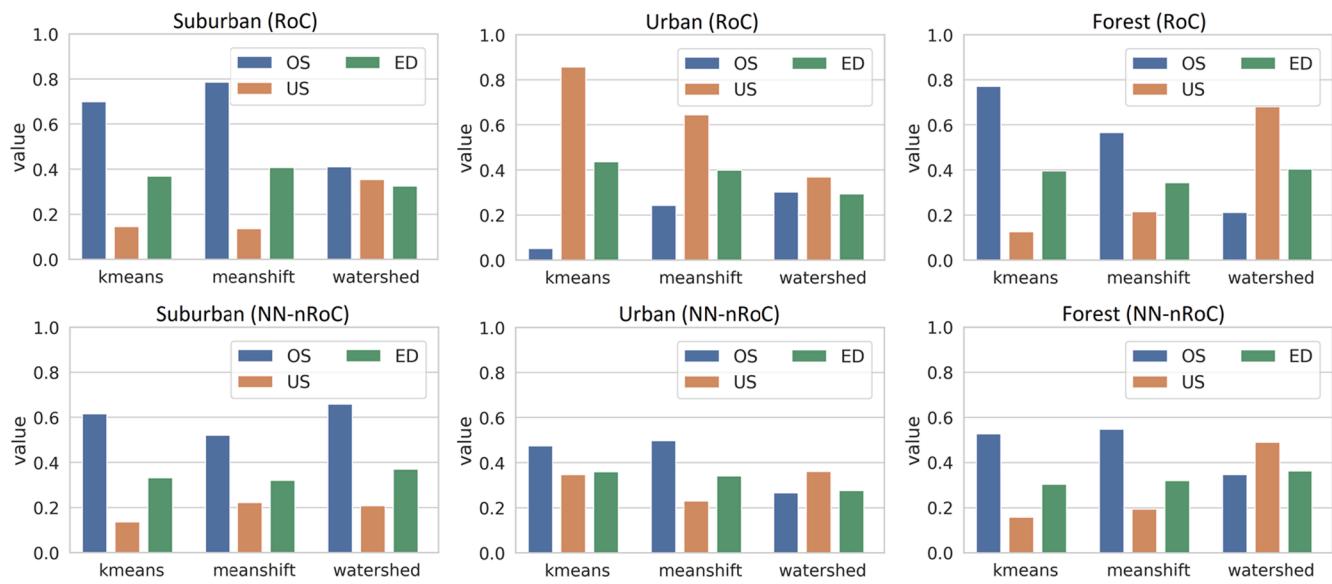


Fig. 15. The overall evaluation of the performance of three algorithms over three images with the averaged US, OS, and ED measures. From left to right are the evaluation results of the suburban, urban, and forest images.

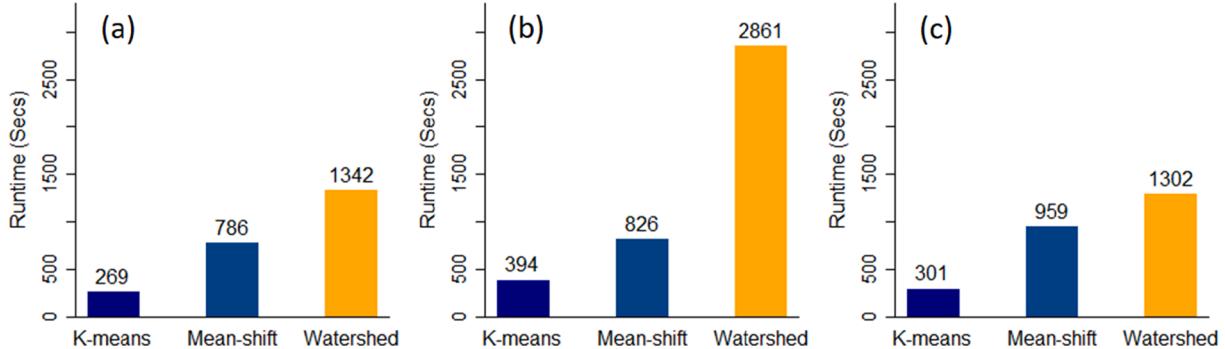


Fig. 16. The segmentation and scale selection runtime for different algorithms across three different images. (a) the suburban image, (b) the urban image, and (c) the forest image.

Table 2

Comparing the performance of the NN-nRoC scale selection method for k-means, mean-shift, and watershed with the ESP-MRS approach.

Image	Index	NN-nRoC Method			ESP
		KM	MS	WS	
Suburban Image	OS	0.61	0.52	0.65	0.94
	US	0.12	0.22	0.21	0.17
	ED	0.34	0.32	0.37	0.48
	Time (sec)	269	786	1,342	18,900
Urban Image	OS	0.47	0.49	0.26	0.97
	US	0.35	0.22	0.36	0.08
	ED	0.36	0.33	0.27	0.49
	Time (sec)	394	826	2,861	43,8001
Forest Image	OS	0.52	0.55	0.34	0.99
	US	0.16	0.19	0.49	0.10
	ED	0.30	0.31	0.36	0.50
	Time (sec)	301	959	1,302	4,201

number of segments found and the constant occurs because we calculate only to descriptions (standard deviation and mean).

The isolation forest outlier detection algorithm is known to have a time complexity of $O(e * d * \log(s))$, where e is the number of estimators (number of trees in the forest), d is the number of dimensions in the feature vector, in our case fixed to only 2 (standard deviation and mean) and n is the number of segments found in the first stage (segmentation

step). In our implementation, we fixed the number of estimators to 100.

The scale selection parameter depends only on the profile curves of NN-nCV and NN-nRoC and its derivative, which makes it linear complexity $2^*O(x)$ where x is the number of scales used in the analysis. In our case, we fixed this number to 100 scales, which makes constant time. In summary the total complexity $T(n, \theta)$ can be written as:

$$T(n, \theta) = 2^*O[f(n, \theta)] + O[2^*100*\log(s)] + 2^*O(100) \quad (17)$$

where $f(n, \theta)$ is the complexity of the segmentation algorithm, n is the image size in pixels, θ is the parameter of the algorithm (initial points for k-means and mean-shift and number of markers for watershed), s is the number of segments found by algorithm f . The worst-case scenario occurs when $f(n, \theta)$ is quadratic type; however, our experiments show less execution time of k-means due to optimizations implemented in the FAISS Python library in such as for the initialization of centroids, also known as k-means++ (Arthur and Vassilvitskii, 2006).

The memory footprint accumulates at every stage. The segmentation stage is the most offender as it requires the storage of the segment id for each pixel, i.e., $M(n) = \Theta(n^2)$. At the final stages, we only have segment descriptors which are aggregated data and depend only on the number of segments $M(s) = \Theta(n)$.

5. Conclusions

This study proposed a robust and repeatable approach for the selection of optimal segmentation scale parameters in high-resolution hyperspectral images. We applied a noise normalization procedure for spectral bands before segmentation and used the curves generated from the CV and its RoC after outlier removal. From the Isolation Forest outlier detection results of CV, it was clear that outlier segments existed at all scales. The more fine-grained the segmentation, the more outliers it produced. The impact of noise in image spectra and outliers on the calculation of averaged CV, and the CV and RoC graphs was substantial, since its CV values were remarkably higher than those of normal segments, even when they were in small amounts. These issues are even more severe for hyperspectral image segmentation, as the data experience higher noise in many spectral bands and the number of small and speckle-noise segments are more likely to occur. It affects the selection negatively, when based only on RoC graphs, making experts select unreliable and inaccurate scales.

By applying noise normalization and removing the outlier segments in the calculation, the NN-nRoC curves were more robust, resulting in a more reliable selection of the optimal segmentation scales. Since the peaks in the NN-nRoC graphs were not always the same, the optimal scales determined from these graphs were different from those determined from the RoC graphs. Also, this underlying distinction is true in different image segmentation algorithms and in images of various ground features and scene complexity.

The visual inspection and quantitative evaluation using OS, US, and ED metrics indicate that the new segmentation procedure worked well for k-means, mean-shift, and watershed algorithms. The results from our new method were greatly improved in comparison to traditional approaches without inverse noise weighting and outlier removal. The segmentation well-delineated various land cover types, especially in segmenting features like rooftops, shadows, roads, and vegetation in the suburban, urban, and forest images. The only mild over- and under-segmentation were observed in all images owing to the complexity of the evaluated scenes and the high variability in the spectral properties of different land cover types in hyperspectral data. However, the evaluation of different segmentation algorithms and images of various land cover types confirmed the excellent performance of this approach for different applications.

The computational complexity of the presented pipeline occurs mostly due to the inherent steps required for segmentation. The overall complexity exposes the simplicity of the scale selection step, which in turn can also be combined with different algorithms, including Deep Learning-Based segmentation and anomaly detector.

The proposed segmentation scale selection approach and the procedure can be applied to other types of data such as digital photos, multispectral and radar, and DEM data for various applications in computer science and remote sensing fields. Further tests with larger images and different types of data with high variability can be conducted for further evaluation and confirmation of the performance of the approach. Besides, visual inspection and manual selection of optimal scales based on the NN-nRoC graphs are time-consuming and can be subject to human errors. As the method is easily reproducible, an automated line graph interpretation can be explored for a faster and more objective segmentation scale selection that can be repeated for any type of data. Furthermore, a random search and multi-parameter tuning should be conducted for an exhaustive search of the best parameter combination.

Declaration of Competing Interest

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

Acknowledgments

This project was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the NSERC Discovery Program (RGPIN-2016-05614 awarded to Dr. Yuhong He) and the Graduate Expansion Fund from the Department of Geography, Geomatics and Environment at the University of Toronto Mississauga. The airborne hyperspectral imaging was supported by the Centre for Global Change Science, the University of Toronto, through the Graduate Student Research Awards (awarded to Phuong D. Dao).

References

- Abdi, H., 2010. Coefficient of variation. Encyclopedia of research design 1, 169–171.
- Arthur, D., Vassilvitskii, S., 2006. k-means++: The advantages of careful seeding. Stanford InfoLab, pp. 1027–1035.
- Baatz, M., Schäpe, A. (Eds.). 2010. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. T. B. J. Strobl, and G. Grieshaber, Angewandte Geographische Informations-Verarbeitung XII. Wichmann, Karlsruhe, Germany, 12-23.
- 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.
- Bellens, R., Gautam, S., Martinez-Fonte, L., Philips, W., Chan, J.C.-W., Canters, F., 2008. Improved classification of VHR images of urban areas using directional morphological profiles. IEEE Trans. Geosci. Remote Sens. 46 (10), 2803–2813.
- Beucher, S., Lantuejoul, C., 1979. Use of watersheds in contour detection. Paper presented at the International Workshop on Image Processing, Rennes, France.
- Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65 (1), 2–16.
- Böck, S., Immitzer, M., Atzberger, C., 2017. On the objectivity of the objective function—Problems with unsupervised segmentation evaluation based on global score and a possible remedy. Remote Sensing 9 (8), 769–769.
- Bondur, V.G., 2014. Modern approaches to processing large hyperspectral and multispectral aerospace data flows. Izv. Atmos. Oceanic Phys. 50 (9), 840–852.
- Camps-Valls, G., Tuia, D., Bruzzone, L., Benediktsson, J.A., 2013. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. IEEE Signal Process Mag. 31 (1), 45–54.
- Chen, S., Yang, X., You, Z., Wang, M., 2016. Innovation of aggregate angularity characterization using gradient approach based upon the traditional and modified Sobel operation. Constr. Build. Mater. 120, 442–449.
- Clinton, N., Holt, A., Scarborough, J., Yan, L.I., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. Photogramm. Eng. Remote Sens. 76 (3), 289–299.
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 5, 603–619.
- Dao, P.D., He, Y., Lu, B., 2019a. Maximizing the quantitative utility of airborne hyperspectral imagery for studying plant physiology: An optimal sensor exposure setting procedure and empirical line method for atmospheric correction. Int. J. Appl. Earth Obs. Geoinf. 77, 140–150.
- Dao, P.D., Liou, Y.-A., 2015. Object-based flood mapping and affected rice field estimation with Landsat 8 OLI and MODIS data. Remote Sensing 7 (5), 5077–5097.
- Dao, P.D., Liou, Y.-A., Chou, C.-W., 2015. Detection of flood inundation regions with Landsat/MODIS synthetic data. Paper presented at the International Symposium on Remote Sensing.
- Dao, P.D., Mong, N.T., Chan, H.-P., 2019b. Landsat-MODIS Image Fusion and Object-based Image Analysis for Observing Flood Inundation in a Heterogeneous Vegetated Scene. GIScience & Remote Sensing 56 (8), 1148–1169.
- Dekker, R., 1998. Speckle filtering in satellite SAR change detection imagery. Int. J. Remote Sens. 19 (6), 1133–1146.
- Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet shrinkage. biometrika, 81, 3, 425–455.
- 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.
- 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. Geograph. Inform. Sci. 24 (6), 859–871.
- Dronova, I., Gong, P., Clinton, N.E., Wang, L., Fu, W., Qi, S., Liu, Y., 2012. Landscape analysis of wetland plant functional types: The effects of image segmentation scale, vegetation classes and classification methods. Remote Sens. Environ. 127, 357–369.
- Epshtain, B., Ofek, E., Wexler, Y., 2010. Detecting text in natural scenes with stroke width transform. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 13–18 June, 2963–2970.
- Espindola, G.M., Câmara, G., Reis, I.A., Bins, L.S., Monteiro, A.M., 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. Int. J. Remote Sens. 27 (14), 3035–3040.
- Fang, H.-T., Huang, D.-S., 2004. Noise reduction in lidar signal based on discrete wavelet transform. Opt. Commun. 233 (1–3), 67–76.
- Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory 21 (1), 32–40.

- Gao, Q., Lim, S., Jia, X., 2018. Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sensing* 10 (2), 299.
- Ghosh, A., Joshi, P.K., 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *Int. J. Appl. Earth Obs. Geoinf.* 26, 298–311.
- Guo, B., Gunn, S.R., Damper, R.I., Nelson, J.D.B., 2006. Band Selection for Hyperspectral Image Classification Using Mutual Information. *IEEE Geosci. Remote Sens. Lett.* 3 (4), 522–526.
- Guo, X., Liu, X., Zhu, E., Yin, J., 2017. Deep Clustering with Convolutional Autoencoders. Paper presented at the International Conference on Neural Information Processing, Guangzhou, China, 26 October, 373–382.
- Gupta, S., Mazumdar, S.G., 2013. Sobel edge detection algorithm. *Int. J. Comp. Sci. Manage. Res.* 2 (2), 1578–1583.
- Hay, G.J., Castilla, G., Wulder, M.A., Ruiz, J.R., 2005. An automated object-based approach for the multiscale image segmentation of forest scenes. *Int. J. Appl. Earth Obs. Geoinf.* 7 (4), 339–359.
- He, W., Zhang, H., Zhang, L., Shen, H., 2015. Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8 (6), 3050–3061.
- Hossain, M.D., Chen, D., 2019. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS J. Photogramm. Remote Sens.* 150, 115–134.
- Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS J. Photogramm. Remote Sens.* 66 (4), 473–483.
- Johnson, J., Douze, M., Jégou, H., 2019. Billion-scale similarity search with GPUs. *IEEE Trans. Big Data*, 1–1.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* 7, 881–892.
- Karl, J.W., Maurer, B.A., 2010. Spatial dependence of predictions from image segmentation: A variogram-based method to determine appropriate scales for producing land-management information. *Ecol. Inf.* 5 (3), 194–202.
- Kim, M., Madden, M., Warner, T. (Eds.). 2008. Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. T. Blaschke, S. Lang G. J. Hay, Object-Based Image Analysis. Lecture Notes in Geoinformation and Cartography. Berlin, Heidelberg: Springer Berlin Heidelberg, 291–307.
- Kornilov, A.S., Safonov, I.V., 2018. An overview of watershed algorithm implementations in open source libraries. *J. Imaging* 4 (10), 123.
- Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* 44 (2–3), 145–163.
- Kurnaz, M.N., Dokur, Z., Ölmez, T., 2005. Segmentation of remote-sensing images by incremental neural network. *Pattern Recognit. Lett.* 26 (8), 1096–1104.
- Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49 (4), 764–766.
- Liu, F.T., Ting, K.M., Zhou, Z.-H., 2012. Isolation-based anomaly detection. *ACM Trans. Knowl. Discovery Data* 6 (1), 1–39.
- Liu, Y., Gao, Q., Miao, S., Gao, X., Nie, F., Li, Y., 2017. A Non-Greedy Algorithm for L1-Norm LDA. *IEEE Trans. Image Process.* 26 (2), 684–695.
- Liu, Y., Nie, F., Gao, Q., Gao, X., Han, J., Shao, L., 2019a. Flexible unsupervised feature extraction for image classification. *Neural Netw.* 115, 65–71.
- Liu, Y., Shan, C., Gao, Q., Gao, X., Han, J., Cui, R., 2019b. Hyperspectral image denoising via minimizing the partial sum of singular values and superpixel segmentation. *Neurocomputing* 330, 465–482.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition.
- Lu, B., Dao, P.D., Liu, J., He, Y., Shang, J., 2020. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sensing* 12 (16), 2659.
- Lu, B., He, Y., Dao, P.D., 2019. Comparing the Performance of Multispectral and Hyperspectral Images for Estimating Vegetation Properties. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (6), 1784–1797.
- Martin, D.R., Fowlkes, C.C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* 5, 530–549.
- Ming, D., Li, J., Wang, J., Zhang, M., 2015. Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. *ISPRS J. Photogramm. Remote Sens.* 106, 28–41.
- Mitra, P., Shankar, B.U., Pal, S.K., 2004. Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recogn. Lett.* 25 (9), 1067–1074.
- Mou, L., Ghamisi, P., Zhu, X.X., 2017. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 55 (7), 3639–3655.
- Mou, L., Ghamisi, P., Zhu, X.X., 2018. Unsupervised Spectral-Spatial Feature Learning via Deep Residual Conv–Deconv Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 56 (1), 391–406.
- Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sens. Environ.* 115 (5), 1145–1161.
- Nalepa, J., Myller, M., Imai, Y., Honda, K., Takeda, T., Antoniak, M., 2019. Unsupervised Segmentation of Hyperspectral Images Using 3-D Convolutional Autoencoders. *IEEE Geosci. Remote Sens. Lett.* 1–5.
- Neubert, P., Protzel, P., 2014. Compact watershed and preemptive slice: On improving trade-offs of superpixel segmentation algorithms. Paper presented at the 22nd International Conference on Pattern Recognition.
- Rodarmel, C., Shan, J., 2002. Principal component analysis for hyperspectral image classification. *Survey. Land Inform. Sci.* 62 (2), 115–122.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36 (8), 1627–1639.
- Smith, A., 2010. Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. *J. Spatial Sci.* 55 (1), 69–79.
- Sorensen, A.T., 2000. Equilibrium price dispersion in retail markets for prescription drugs. *J. Polit. Econ.* 108 (4), 833–850.
- Su, T., Zhang, S., 2017. Local and global evaluation for remote sensing image segmentation. *ISPRS J. Photogramm. Remote Sens.* 130, 256–276.
- Tarabalka, Y., Chanussot, J., Benediktsson, J.A., 2010. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recogn.* 43 (7), 2367–2379.
- Theis, L., Shi, W., Cunningham, A., Huszár, F., 2017. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*.
- Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Yu, T., 2014. scikit-image: image processing in Python. *PeerJ* 2 e453–e453.
- Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intelligence* 6, 583–598.
- Wang, Y., Qi, Q., Liu, Y., Jiang, L., Wang, J., 2019. Unsupervised segmentation parameter selection using the local spatial statistics for remote sensing image segmentation. *Int. J. Appl. Earth Obs. Geoinf.* 81, 98–109.
- Wang, Z., Song, C., Wu, Z., Chen, X., 2005. Improved watershed segmentation algorithm for high resolution remote sensing images using texture. Paper presented at the IEEE International Geoscience and Remote Sensing Symposium.
- Weber, E.U., Shafir, S., Blais, A.-R., 2004. Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychol. Rev.* 111 (2), 430–430.
- Winter, S., 2000. Location similarity of regions. *ISPRS J. Photogramm. Remote Sens.* 55 (3), 189–200.
- Witharana, C., Civco, D.L., 2014. Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2). *ISPRS J. Photogramm. Remote Sens.* 87, 108–121.
- Woodcock, C.E., Strahler, A.H., 1987. The factor of scale in remote sensing. *Remote Sens. Environ.* 21 (3), 311–332.
- Yang, J., He, Y., Weng, Q., 2015. An automated method to parameterize segmentation scale by enhancing intrasegment homogeneity and intersegment heterogeneity. *IEEE Geosci. Remote Sens. Lett.* 12 (6), 1282–1286.
- Yang, J., Li, P., He, Y., 2014. A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation. *ISPRS J. Photogramm. Remote Sens.* 94, 13–24.
- Yin, D., Du, S., Wang, S., Guo, Z., 2015. A direction-guided ant colony optimization method for extraction of urban road information from very-high-resolution images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8 (10), 4785–4794.
- Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., Schirokauer, D., 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm. Eng. Remote Sens.* 72 (7), 799–811.
- Zhang, X., Xiao, P., Song, X., She, J., 2013. Boundary-constrained multi-scale segmentation method for remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 78, 15–25.
- Zhong, P., Gong, Z., Li, S., Schönlieb, C., 2017. Learning to Diversify Deep Belief Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 55 (6), 3516–3530.