

Article

Operational Large-Scale Segmentation of Imagery Based on Iterative Elimination

James D. Shepherd ^{1,*}, Pete Bunting ² and John R. Dymond ¹

¹ Landcare Research, Private Bag 11052, Palmerston North 4442, New Zealand; dymondj@landcareresearch.co.nz

² Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth SY23 3DB, UK; pete.bunting@aber.ac.uk

* Correspondence: shepherdj@landcareresearch.co.nz

Received: 28 February 2019; Accepted: 15 March 2019; Published: 18 March 2019



Abstract: Image classification and interpretation are greatly aided through the use of image segmentation. Within the field of environmental remote sensing, image segmentation aims to identify regions of unique or dominant ground cover from their attributes such as spectral signature, texture and context. However, many approaches are not scalable for national mapping programmes due to limits in the size of images that can be processed. Therefore, we present a scalable segmentation algorithm, which is seeded using k-means and provides support for a minimum mapping unit through an innovative iterative elimination process. The algorithm has also been demonstrated for the segmentation of time series datasets capturing both the intra-image variation and change regions. The quality of the segmentation results was assessed by comparison with reference segments along with statistics on the inter- and intra-segment spectral variation. The technique is computationally scalable and is being actively used within the national land cover mapping programme for New Zealand. Additionally, 30-m continental mosaics of Landsat and ALOS-PALSAR have been segmented for Australia in support of national forest height and cover mapping. The algorithm has also been made freely available within the open source Remote Sensing and GIS software Library (RSGISLib).

Keywords: image segmentation; remote sensing; land cover; iterative elimination; RSGISLib

1. Introduction

The classification and analysis of remotely-sensed optical data has become a key technology for the mapping and monitoring of land cover [1,2]. Traditionally, such analysis has been performed on a per pixel basis, but over the last 20 years, there has been a significant movement to embrace context and segment-based classifications [3] due to observed improvements in classification accuracy [3–7]. A significant driver for this adoption has been the availability of the eCognition software [8], which has provided user-friendly tools for these operations and made them widely available to the community. For national mapping programs, high-resolution remotely-sensed data such as RapidEye, Sentinel 1 and 2, SPOT (4–5) and Landsat (TM, ETM+, OLI), provide the majority of data. These data typically have a resolution of 5–30 m, thereby providing sufficient detail for monitoring land cover, such as forestry, agriculture and grasslands. The use of segmentation for land cover mapping at this resolution aims to identify spatially-homogeneous units such as fields and forestry blocks as single objects, or a small number of large objects. However, any under-segmentation will result in a poor classification result, as features are merged and cannot be identified. Therefore, where errors occur within the segmentation result, a small amount of over-segmentation is preferable to retain classification accuracy [9].

Building on the extensive literature on image segmentation (e.g., [10–14]) within the image processing and medical imaging communities, there have been a significant number of publications devoted to segmentation of remotely-sensed imagery. These can be categorised as either top-down or bottom-up approaches. Top-down approaches, such as multi-scale wavelet decomposition [15], start with a single region and progressively break the image into smaller regions until a termination condition is met. Bottom-up methods, such as region growing and spectral clustering [16], start with individual pixels and group them until a termination condition is met. Termination criteria vary between approaches, but they are generally a combination of the colour similarity, shape, and size of the regions generated. To group pixels within a bottom-up approach, a number of methods have been proposed, but include region growing [17], thresholding [18], statistical clustering [19], Markov random field models [20], fuzzy clustering [21], neural networks [22] and image morphology watersheds [23]. Thresholding techniques are generally applied to extract features from a distinctive background, with histograms commonly being used to identify the threshold values automatically [24]. Region growing techniques require seeds, which are difficult to generate reliably automatically and are time consuming to provide manually. Soille [17], Brunner and Soille [25] used an automated seeding method based on quasi-flat zones, as did Weber et al. [26], who applied it to multi-temporal imagery. Watershed techniques are one of the most commonly applied, but these are often over-segmented and, due to the filtering used to generate the gradient image(s) boundaries between features, are often blurred and poorly defined at the pixel level.

Segmentation techniques are often designed for specific applications, such as tree crown delineation [27], requiring considerable parameterisation that can be difficult to define and optimise [28]. The termination criteria used are also a common differentiating factor in the results, which are generated by the segmentation algorithm. For example, Baatz and Schäpe [28] aimed to minimise the spectral homogeneity of the image regions while creating regions of similar size and compact shape [8] to allow accurate statistical comparison between features for later classification. The key parameter within the eCognition multi-scale segmentation algorithm is the scale factor f [28]. The user-defined threshold of f controls the termination for the segmentation process where f is defined as:

$$\begin{aligned}
 f &= w \cdot h_{colour} + (1 - w) \cdot h_{shape} \\
 h_{colour} &= \sum_{b=1}^{n_{bands}} w_b \cdot \sigma_b \\
 h_{shape} &= w_s \cdot h_{cmpct} + (1 - w_s) \cdot h_{smooth} \\
 h_{cmpct} &= \frac{p_l}{\sqrt{n_{pxl}}} \\
 h_{smooth} &= \frac{p_l}{p_{bbox}}
 \end{aligned} \tag{1}$$

where w is a user-defined weight (range: 0–1) defining the importance of object colour versus shape within the termination criteria, n_{bands} is the number of input image bands, w_b is the weight for the image band b , σ_b is the standard deviation of the object for band b , w_s is the user-defined weight (range: 0–1) defining the importance of compactness versus smoothness, p_l is the perimeter length of the object, n_{pxl} is the number of pixels within the object and p_{bbox} is the perimeter of the bounding box of the object.

Others have used k-means clustering as an alternative approach. Lucchese and Mitra [29] used k-means clustering and median filtering for post-processing. While the method is simple and effective, some of the objects with sharper edges were over-smoothed from the median filtering. Wang et al. [19] first initialised the segmentation using a k-means clustering to remove noise and reduce the feature space from single pixels to clumps. The clumps were then merged using a combination of spectral colour (weighted by object size) and object shape. The resulting segmentation is similar to that derived from the eCognition software. Likewise, Brunner and Soille [25] executed a single merge of clumps based on spectral similarity.

Computational constraints are another issue commonly encountered with many segmentation algorithms. The size of the scene that can be processed is in some cases significantly limited. Often,

one whole satellite image cannot be processed in a single step, due to memory constraints. Additionally, methods commonly deployed within the field of computer vision expect either one (greyscale) or three (red, green and blue) image bands, or spectral measurements, and it is assumed that these data channels are scaled to the same range (i.e., 0–255). In remote sensing, this is commonly not the case, with many satellites having multi-spectral capabilities with at least four spectral wavelengths, including the near-infrared (NIR), which over vegetation, results in considerably larger range of values than other channels due to leaf cell structure [30].

Assessing the quality and effectiveness of a segmentation result has proven difficult [31]. In this context, the aim of segmentation is to subdivide the landscape into units of the same land cover or spectral colour, where ideally, the entire feature, such as a single species forest block, will be captured as a single segment so that these units can be classified within an appropriate context. However, if the scene is under-segmented, that is, multiple features of interest are contained within a single segment, then classification of these features is not possible. Therefore, under-segmentation is to be avoided. Therefore, a degree of over-segmentation is generally accepted, as the following classification of neighbouring objects of the same class can be merged. Additionally, the scale of the imagery and result being sought will significantly impact segmentation requirements. Segmentation quality can be assessed for specific tasks, such as delineation of tree crowns or buildings, with respect to a set of reference data. The generation of a reference set is a clear task with a relatively clear answer for specific features, but for more general segmentation tasks, such as splitting a whole SPOT 5 scene into units for land cover classification, where a large number of land covers, uses and vegetation conditions are present, there is no one “best” solution. In an attempt to quantify the differences between different segmentation results, various approaches exist [32] where parameters such as shape (e.g., shape index [33]), comparison to a set of reference features [34–36], synthetic images [37] and colour metrics [31,38] have been used.

Our aim was to provide a segmentation technique that is scalable to support the national land cover mapping program of New Zealand, while only using a few understandable parameters. Segments needed to be of uniform spectral response (i.e., colour) with a minimum size and suitable for land cover and land cover change classification, while boundaries between land covers should be sharp and accurate. The algorithm needs to be efficient over large datasets (i.e., multiple mosaicked scenes) and applicable to a wide range of sensors and image pixel resolution. We have therefore proposed a new iterative elimination algorithm to meet these aims. We describe and apply the method on Landsat ETM+ and SPOT5 imagery of typical landscapes in New Zealand.

2. Methods

The iterative elimination method we present here has four steps (Figure 1). First, is a seeding step, which identifies the unique spectral signatures within the image, second a clumping process to create unique regions, third an elimination process to remove regions below the minimum mapping unit, and, finally, the regions are relabelled so that they are sequentially numbered, providing the final segmentation. These steps require two parameters, the initial number of cluster centres to be produced and the minimum mapping unit defined for the elimination process.

When processing imagery with a large number of highly-correlated image bands (i.e., time series and hyper-spectral imagery), it is recommended that a subset or derived compression of the image data (e.g., principle component analysis) of the available input image bands be used to both minimise auto-correlation between the bands and reduce processing time. In this study, the segmentation of SPOT-5 data from two dates was undertaken using the near-infrared (NIR), shortwave infrared (SWIR) and red wavelengths (i.e., the green band was not included due to the significant correlation with the red band). Images were processed from DN to standardised reflectance before segmentation to reduce scene complexity [39].

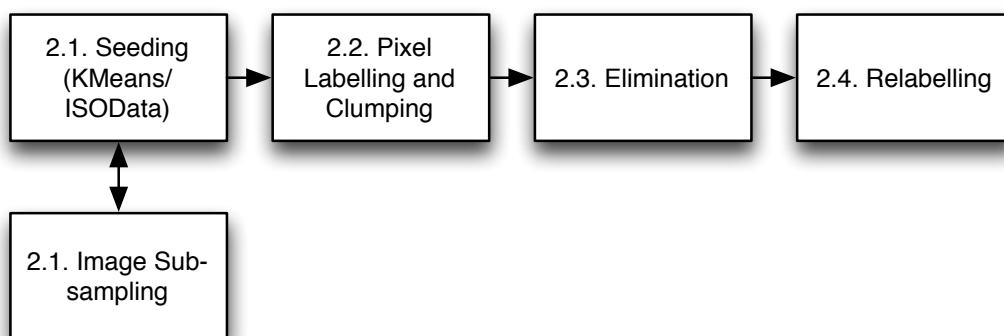


Figure 1. Flowchart for the segmentation algorithm.

2.1. Seeding

The first step in the algorithm is to seed an initial set of unique spectral signatures to represent the image. To seed, in an unsupervised fashion, the k-means [40] clustering algorithm was used. While there are many clustering algorithms within the literature such as Iterative Self-Organising Data (ISOData [41]), mean-shift [42], hierarchical clustering [43] and fuzzy k-means [44], it was found that k-means produced the best results while being computationally efficient. The k-means clustering algorithm is an iterative procedure requiring calculations in the order of nc ($\mathcal{O}(nc)$) where n is the number of pixels and c is the number of cluster centres. To minimise the number of calculations, the image was subsampled before clustering. It was found that subsampling to 10% or 1% of the image pixels was sufficient to provide similar cluster centres to those derived from the whole input dataset (Figure 2), while providing a significant performance improvement in computation time (Table 1). Another advantage of k-means is that it allows development of the model on one dataset (i.e., a subset) and application to another. As with any statistical sampling, a sufficiently large sample is required to ensure representativeness and that no features are missed. Experiments were carried out using a number of scenes (5 Landsat-7, 5 SPOT-5 and 3 WorldView-2) from each of the sensors to cover a range of environments.

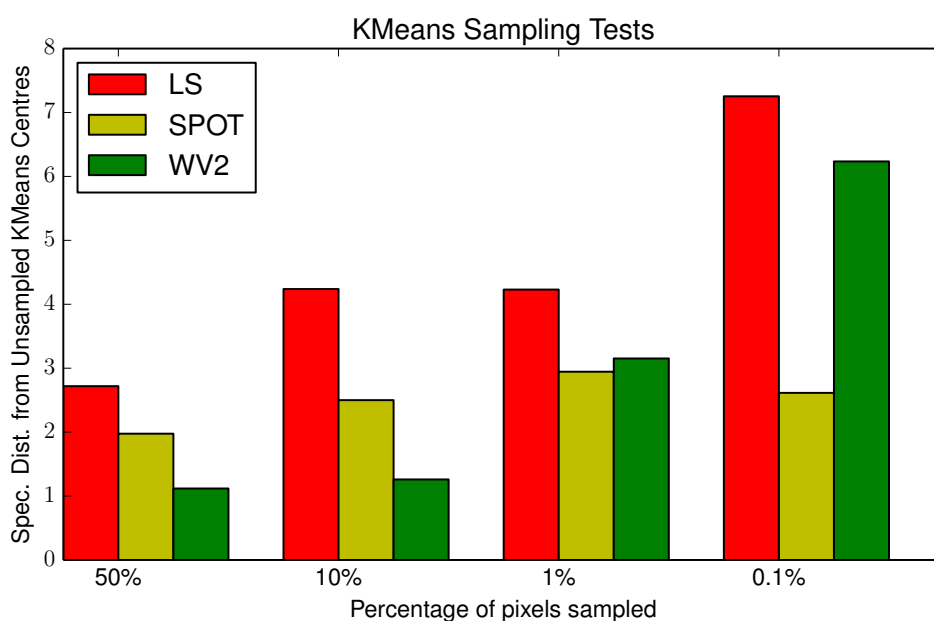


Figure 2. Demonstration that k-means cluster centres generated using 50%, 10%, 1% and 0.1% subsampled image data result in cluster centres close to the original data with considerably less computing time. The imagery used is Landsat ETM+, SPOT-5 and WorldView-2.

Table 1. Timing experiment generating 60 clusters using k-means employing resampling. Experiments were carried out on a 2.93-GHz Intel Xeon processor with 8 GB of RAM running Mac OSX 10.7 using a SPOT 5 scene with 60 million pixels and 4 images bands.

Sampling (Pixels)	Time (Minutes:Seconds)
100%	17:14
50%	11:10
20%	04:03
1%	02:39
0.1%	00:08

Within the k-means clustering process, the Euclidean distance metric is used to calculate the distance between features and cluster centres within feature space. Therefore, due to the differences in dynamic range between the input wavelengths (i.e., near-infrared has a larger range than red), some channels could be weighted higher than others and need to be rescaled. A number of options exist for rescaling the data, but it was found that normalising each image band (wavelength) independently within a range of two standard deviations from the mean, thresholded by the image minimum and maximum, produced the best results, increasing the contrast between the land cover classes.

2.2. Clumping

The k-means-classified pixels are then clumped to define a set of geographically uniquely-labelled regions. This result is over-segmented in many regions of the scene due to the per-pixel nature of the processing and spectral confusion between units. It is common that over half of the resulting clumps are only a single pixel in size (Figure 3).

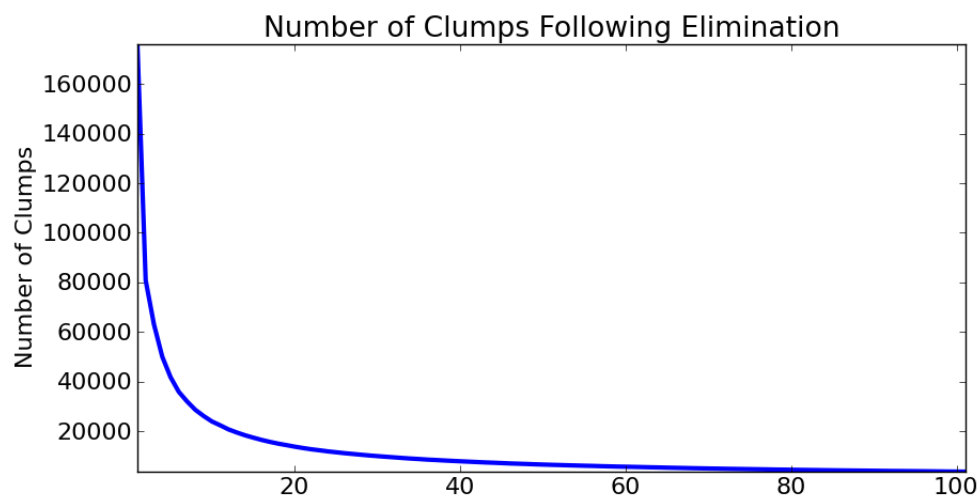


Figure 3. The number of clumps of each size at each elimination step from 1–100 pixels (1 ha) on a 10×10 km SPOT5 image (see Figure 4).

2.3. Local Elimination

The next stage eliminates these small clumps up to a minimum size, which will correspond with the minimum mapping unit for subsequently derived mapping products. The elimination is an iterative process (Algorithm 1) where regions are eliminated based on size, starting with the smallest (i.e., 1 pixel). To eliminate, a region is merged with its spectrally-closest (colour) neighbour, which must be larger than itself, but does not need to be larger than the minimum mapping unit. If there are no neighbouring regions larger than the region being eliminated, then it is left for the next iteration, thereby ensuring that regions are eliminated naturally into the most appropriate neighbour. A critical consideration of the elimination is that the elimination itself (i.e., merging of clumps) is not performed until the end of each iteration. Otherwise, the elimination will result in region growing from the first

clump reviewed for elimination as the clump spectral mean will change. Upon merging, the clump statistics are updated to provide the new mean spectral values used within the next iteration.

Algorithm 1 Pseudocode for the elimination process.

```

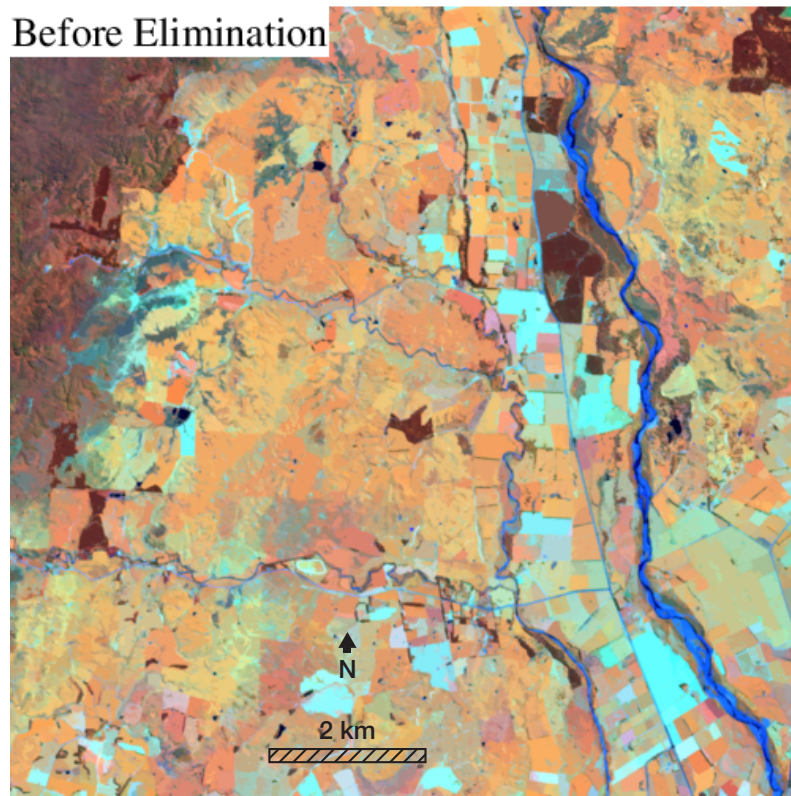
List elimPairs
for  $size = 1 \rightarrow minunitsize$  do
   $clumps \leftarrow GetUnitsOfSizeLessThan(size)$ 
  for all  $clump \leftarrow clumps$  do
     $neighbors \leftarrow GetNeighborsAreaGtThan(clump, size)$ 
    if  $neighbors.size() > 0$  then
       $neighbor \leftarrow FindSpectrallyClosestNeighbor(neighbors, clump)$ 
       $elimPairs \leftarrow pair(neighbors, clump)$ 
    end if
  end for
  for all  $pair \leftarrow elimPairs$  do
     $pair.neighbor \leftarrow MergeRegions(pair.neighbor, pair.clump)$ 
  end for
end for

```

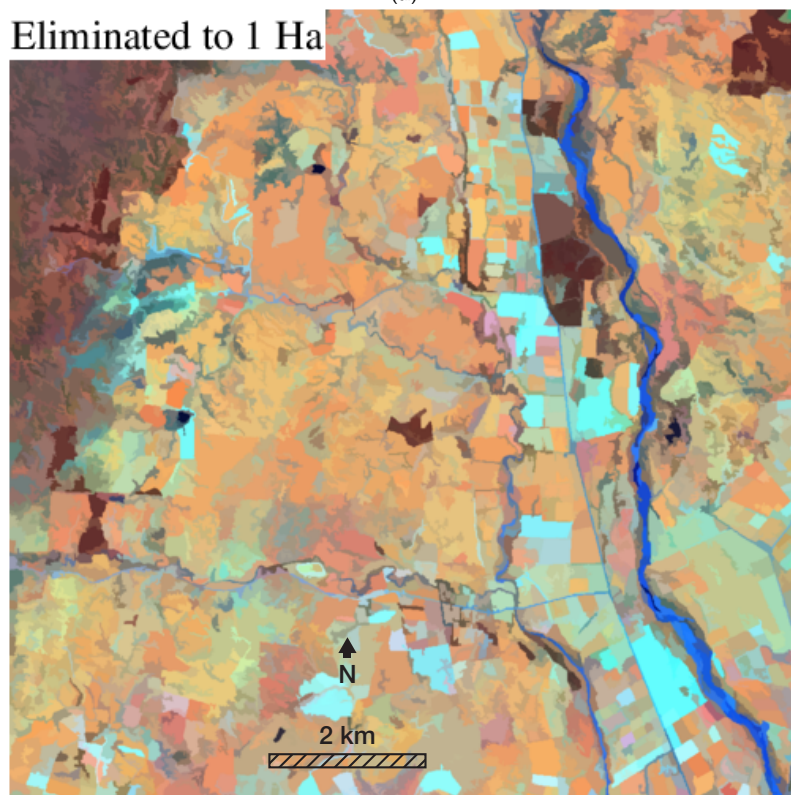
One problem with the elimination process is that features, which are highly spectrally distinct from their neighbours, such as water bodies, with an area smaller than the minimum clump size will be merged. This can result in undesirable spectral mixing and potentially reduces the likelihood of correct classification of land cover. To resolve this issue, an optional threshold was introduced that prevents clumps with a spectral distance (d) above the defined threshold from being merged with their neighbour, hence allowing regions below the minimum size to be retained. The spectral distance threshold was set at 0.1 reflectance. Whenever clumps differ by more than that, they do not merge. We found the 0.1 threshold effective, but this might need to be reviewed depending on the types of features found in some scenes.

As shown in Figure 3, the elimination process reduces the number of clumps representing a 10×10 km SPOT5 scene from over 170,000 to less than 4000. Before elimination (Figure 4a), there were 1137 segments over the 100-pixel (1 ha) threshold of the object size. Therefore, 31% of segments in the final result (Figure 4b) were based on a clump directly originating from the k-means clustering with the rest amalgamated from the smaller clumps, which will not correspond directly with the spectral regions defined by the k-means clustering.

The first step, which removes single pixels, results in a significant removal of clumps (170,000–80,000). For single pixels, a memory-efficient approximation of the elimination process can be adopted. Using a 3×3 window, all single pixel clumps can be identified and stored as a binary image mask. The single pixels can be iteratively eliminated by merging each pixel with the clump of the spectrally-closest pixel that is part of a larger clump (i.e., at least two pixels). If there is no neighbouring clump of at least two pixels, then the single pixel is considered in the next iteration. The spectrally-closest pixel may occasionally differ from the spectrally-closest clump, but this has not been found to be significant and greatly reduces the memory requirements of the elimination process, thus allowing larger datasets to be rapidly processed.



(a)



(b)

Figure 4. (a,b) Elimination from 1–100 pixels on a 10×10 km SPOT5 image (10-m resolution). Segments are coloured with their mean spectra.

2.4. Relabelling

During elimination, the features below the minimum mapping unit are removed and merged with their spectrally-closest larger neighbour. This results in a gap within the attribute table for each eliminated unit. Therefore, a relabelling process to collapse the attribute table is undertaken to ensure the feature identification numbers are sequential. Sequential numbering minimises the number of required rows within the attribute table and makes later classification more efficient.

2.5. Parameterisation

Using k-means seeding, the algorithm has two main parameters: (1) the number of seeds (k) (i.e., initial clusters in feature space) and (2) the minimum segment size for the elimination. There are no hard and fast rules for identifying these parameters, as with all segmentation algorithms, it is difficult to define a metric that completely quantifies the quality of the segmentation [31]. Nevertheless, the selection process may be guided by observing the resulting number of segments, their size, and their internal and inter-neighbour spectral variation [38]. It is worth noting that smaller segments will provide lower spectral variation both within the segments and to their neighbours.

The number of clusters selected for the k-means to be produced as seeds is the key parameter governing the spectral separation of the features. Too few clusters and features, which are close to one another within the feature space, will be under-segmented, while too many clusters will result in an over-segmented scene. Through experimentation, it was found that between 30 and 90 seeds, depending on the level of detail required by the user, were generally sufficient for multi-spectral imagery, such as RapidEye, Sentinel-2, SPOT5, WorldView-2 and Landsat (TM, ETM+, OLI), regardless of the number of image bands or the use of multi-date image stacks. Although, in some use cases, such as segmenting bright objects on a dark background (e.g., ships on the ocean), a significantly smaller number of seeds (e.g., 2 or 3) may be more suitable. For vegetation studies, a value of 60 seeds has been consistently used across a range of studies (e.g., [45,46]) using Sentinel-2, SPOT5 and Landsat data, either at a single or with multiple dates and found to produce good results.

The minimum segment size for the elimination can be related to the minimum mapping unit of the output product being derived using the resulting segments. Where a minimum mapping unit is not defined by the project, the user needs to assess the size of the smallest features they are interested in within the image. For example, if isolated trees within paddocks are of interest within high resolution imagery (i.e., WorldView2), then the minimum segment size needs to be set such that the isolated trees are above this threshold, otherwise they will be eliminated in the larger paddock segment.

3. Results

Assessment of segmentations was aided through the establishment of a set of reference regions, which were manually drawn with reference to the original satellite imagery. We also observed key features and the overall appearance of the segmentation result compared with input image. In this context, segmentation can be thought of as a means of image compression. Therefore, when segments are coloured with their mean spectral values, the resulting image should retain the same structures as the original image (i.e., as shown in Figure 4). We segmented a range of optical imagery, and the resulting number of segments, their size and their spectral variation were assessed to ascertain whether the segmentation results were fit for the purpose of land cover classification, including change.

3.1. Visual Assessment

Segmentations were produced using SPOT-5 scenes for the whole of New Zealand, and Figure 5 shows an example. The NIR, SWIR and red image bands were used and segmented with $k = 60$ and a minimum segment size of 100 pixels.

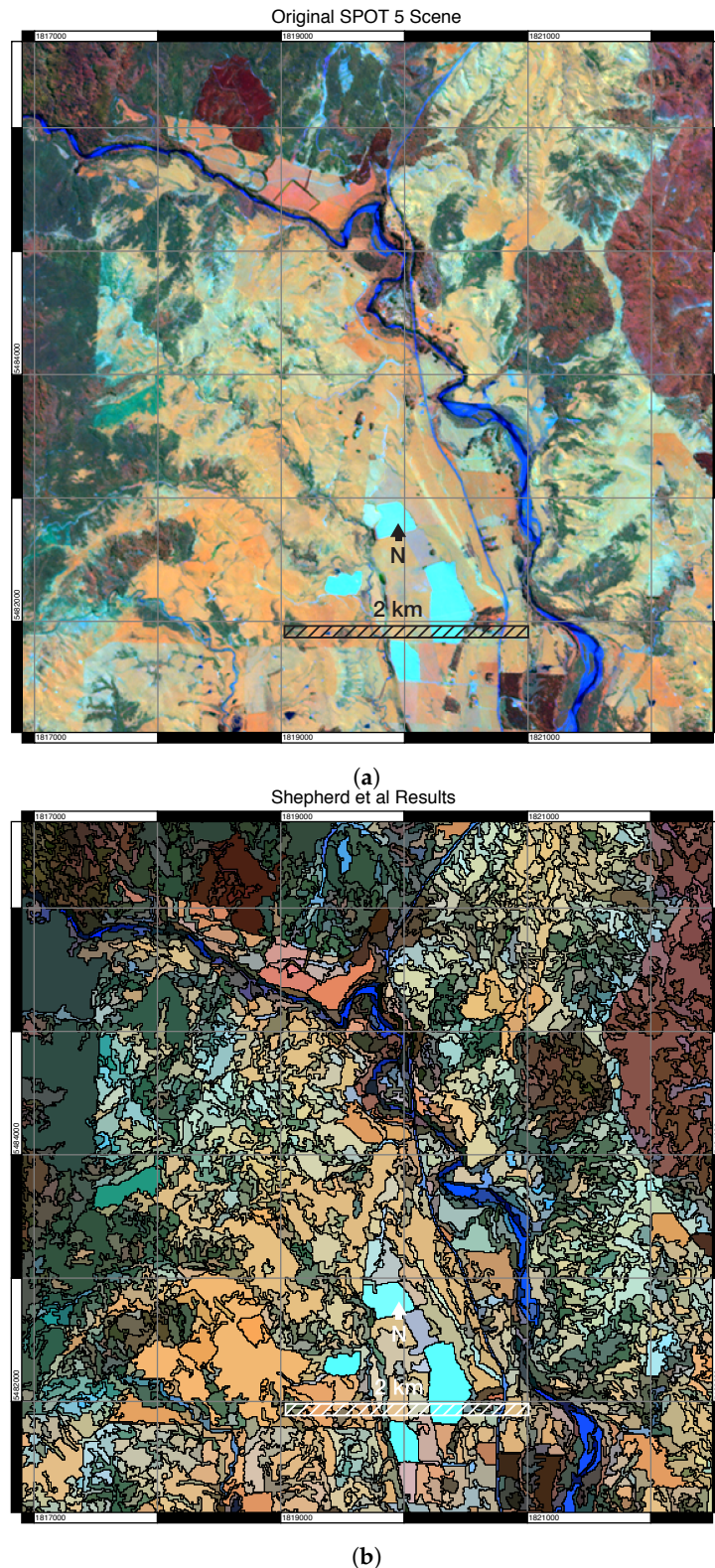


Figure 5. Example segmentation on a SPOT 5 scene (a), using a colour composite of red = NIR, green = SWIR and blue = Red. Segmented with 60 seeds and eliminated to a minimum mapping unit of 1 ha (b). The spatial pattern of land cover is well represented with the number of clumps being 0.6% of the number of pixels.

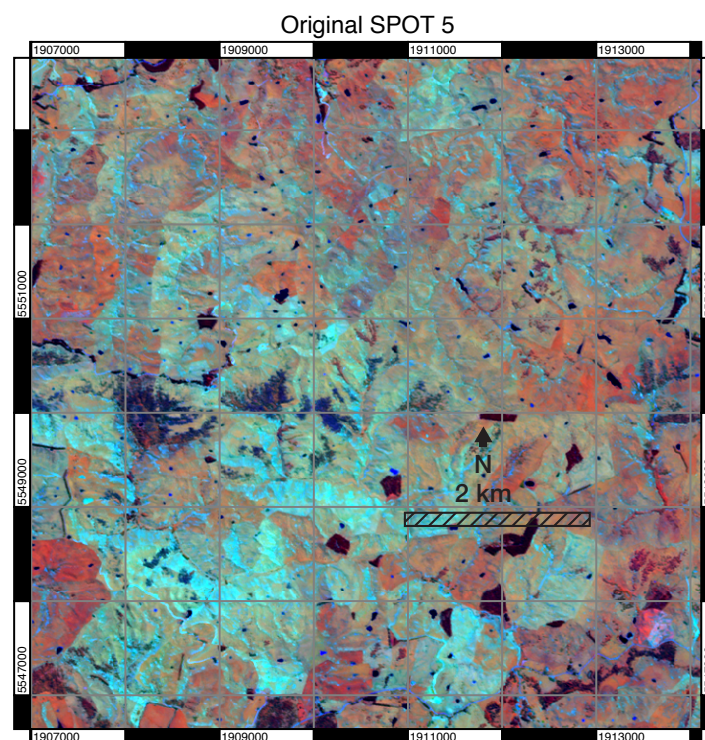
Figure 5 demonstrates the application of the segmentation process on a region of New Zealand typical to the North Island with forest, scrub, agricultural fields and grasslands at various conditions.

As demonstrated, the algorithm has maintained the spectral homogeneous forest and scrub blocks as larger continuous segments, while regions of higher spectral heterogeneity, such as the grasslands, have been segmented into smaller units. Overall, it is considered that this segmentation provides an ideal basis for a subsequent land cover classification. There is some moderate over-segmentation, particularly within the grasslands and some spectrally-heterogeneous forest, but this is often desirable to ensure under-segmentation is not present, which prevents the correct classification of those segments.

Where small spectrally-distinct features need to be retained, the spectral distance threshold on the elimination, which prevents features too far away from one another in the feature space from being merged, can be used. As demonstrated in Figure 6c, where a threshold of 0.1 reflectance has been applied, the small ponds and some additional shelter belts have been retained that are below the minimum mapping unit of 2.5 ha. This produces a segmentation result that more closely corresponds to the original image (Figure 6a) compared to the segmentation without the application of the threshold (Figure 6b).

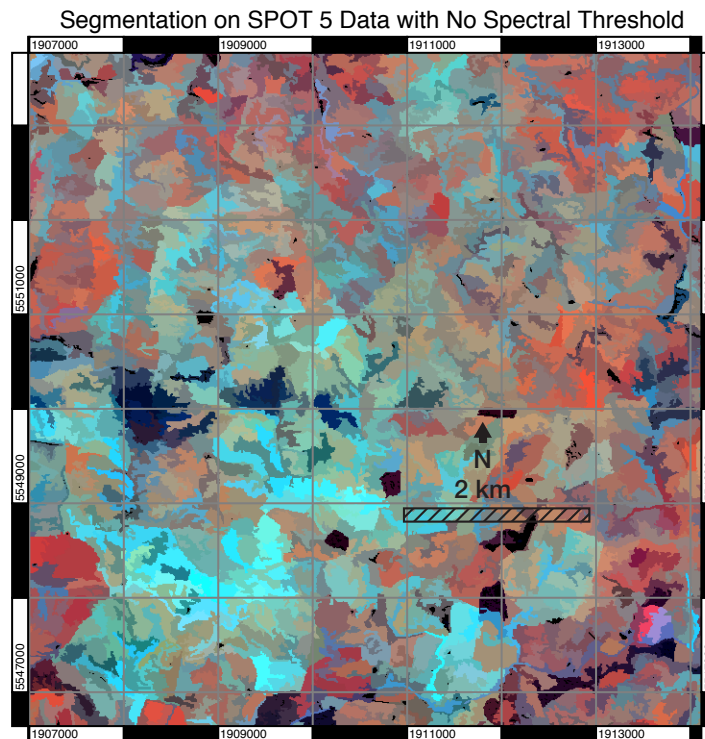
3.2. Parameterisation

To understand the effect of the number of seeds, a number of tests were undertaken for a range of sensors. The datasets used for these experiments (Table 2) were a pair of multi-spectral WorldView2 scenes for Wales, U.K., spectrally transformed to top-of-atmosphere reflectance. A pair of SPOT5 scenes and Landsat (L4/L7), covering regions of North Island, New Zealand, scenes were all corrected to surface reflectance and standardised for Sun angle and topography [39]. For each of these sensors, experiments were undertaken for an image at a single date and the combined multi-date pair. For the single date tests, all the image bands were used, while for the multi-date tests, a subset of the image bands deemed optimal for visualisation was used. For the SPOT5 and Landsat data, these were the red, NIR and SWIR (SWIR1 for Landsat) image bands; while for WorldView2 images, bands NIR (Band 8), red edge (Band 6) and blue (Band 2) were used. The minimum segment size was defined as 100 pixels for the WorldView2 and SPOT5 data and 30 pixels for Landsat 4/7 data.

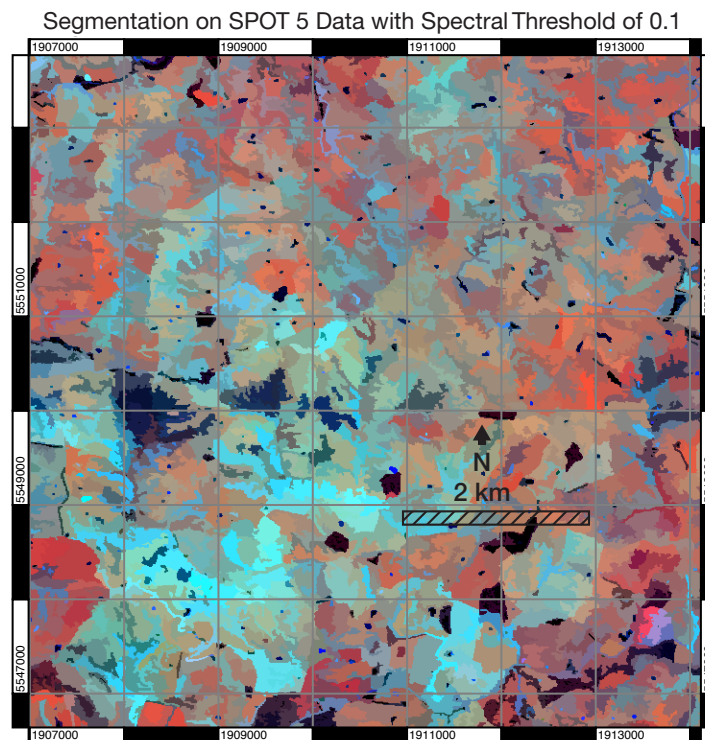


(a) The original SPOT-5 image before segmentation.

Figure 6. Cont.



(b) The segmented SPOT-5 image, shown as the mean-reflectance for a segment, without using a spectral distance threshold to allow smaller features to be retained.



(c) The segmented SPOT-5 image, shown as the mean-reflectance for a segment, where a threshold to allow smaller features to be retained has been used.

Figure 6. (a) The original SPOT 5 image, (b) the segmentation applied without a spectral distance threshold (60 seeds and eliminated to the 2.5-ha minimum mapping unit) and (c) the segmentation applied with a spectral distance threshold of 0.1 reflectance. Images are displayed with a band combination of NIR, SWIR and red.

Table 2. Datasets used for parameter testing. * indicates the image used for the single date tests.

Sensor	Year (s)	Image Size	Pixel Resolution
Worldview2	July 2011 *, November 2011	3539 × 3660	2 m
SPOT5	2008, 2012 *	7753 × 7703	10 m
Landsat 4/7	1990, 2002 *	12,000 × 12,000	15 m

Figure 7 shows the number of segments required to characterise 50% of the total area of the scenes. If a large number of segments are required to make up 50% of the area, it is likely the scene is over-segmented, while if small, it is likely to be under-segmented. These plots therefore provide an understanding of the relationship between the number and size of the segments with respect to the number of k clusters, from which the recommendation that the suitable range of values for k is 30–90 can be made. A concurrent visual examination of segments is important.

For the SPOT scenes (Figure 7A,B), the number of segments increased rapidly before plateauing. This shape suggests there is limited benefit in increasing the size of k once the plateau has been reached. In the test datasets, it was observed that a typical range of k was from 30–90 clusters (shown by the red lines). The 50% cumulative area profile (Figure 7C) for the WorldView2 scene (Figure 8) is a straight line. This is due to the number of large features within the scene, specifically the raised bog (Borth Bog, Wales, U.K.) in the centre of the scene, the estuary and sea, which even at high values of k , generated large segments due to their spectral homogeneity. When higher values of the cumulative area were considered (Figure 7D), the normal profile was seen, and the range of k from 30–90 clusters when visually assessed was still found to be appropriate.

3.3. Comparison to Other Algorithms

To compare the performance of the algorithm detailed in this paper to those of others, four alternatives were identified. The algorithms used for the comparison were the mean-shift algorithm [42] implemented within the Orfeo toolbox [47], as it is widely cited (e.g., [48–50]) as an approach that produced good results on a wide variety of Earth Observation (EO) data, the eCognition multi-resolution segmentation algorithm [28], as the algorithm most commonly used within the literature (e.g., [1,2,51]), and the Quickshift algorithm of Vedaldi and Soatto [52] and the algorithm of Felzenszwalb and Huttenlocher [53], implemented within the scikit-image library and interfaced within the RSGISLib library [54], as examples of more recent approaches from the computer vision community applicable to EO data. A SPOT-5 scene (a subset of which is shown in Figure 10A), which represents a range of land covers and uses, was used for the experiment. To identify the parameters for each of the segmentation algorithms, a grid search was performed for each algorithm across a range of parameters (Table 3).

Table 3. Parameters used for each of the segmentation algorithms.

Algorithm	Parameters	Number of Segmentations
eCognition	scale: [10–100], shape: [0–1], compact.: [0–1]	1210
Mean-Shift	range radius; [5–25], convergence thres.: [0.01–0.5], max. iter.: [10–500], min. size: [10–500]	625
Felzenszwalb	scale: [0.25–10], sigma: [0.2–1.4], min. size: [5–500]	343
Quickshift	ratio: [0.1–1], kernel size: [1–20], max. dist.: [1–30], sigma: [0–5]	1500
Shepherd et al.	k: [5–120], d: [10–10,000], min. size: [5–200]	540

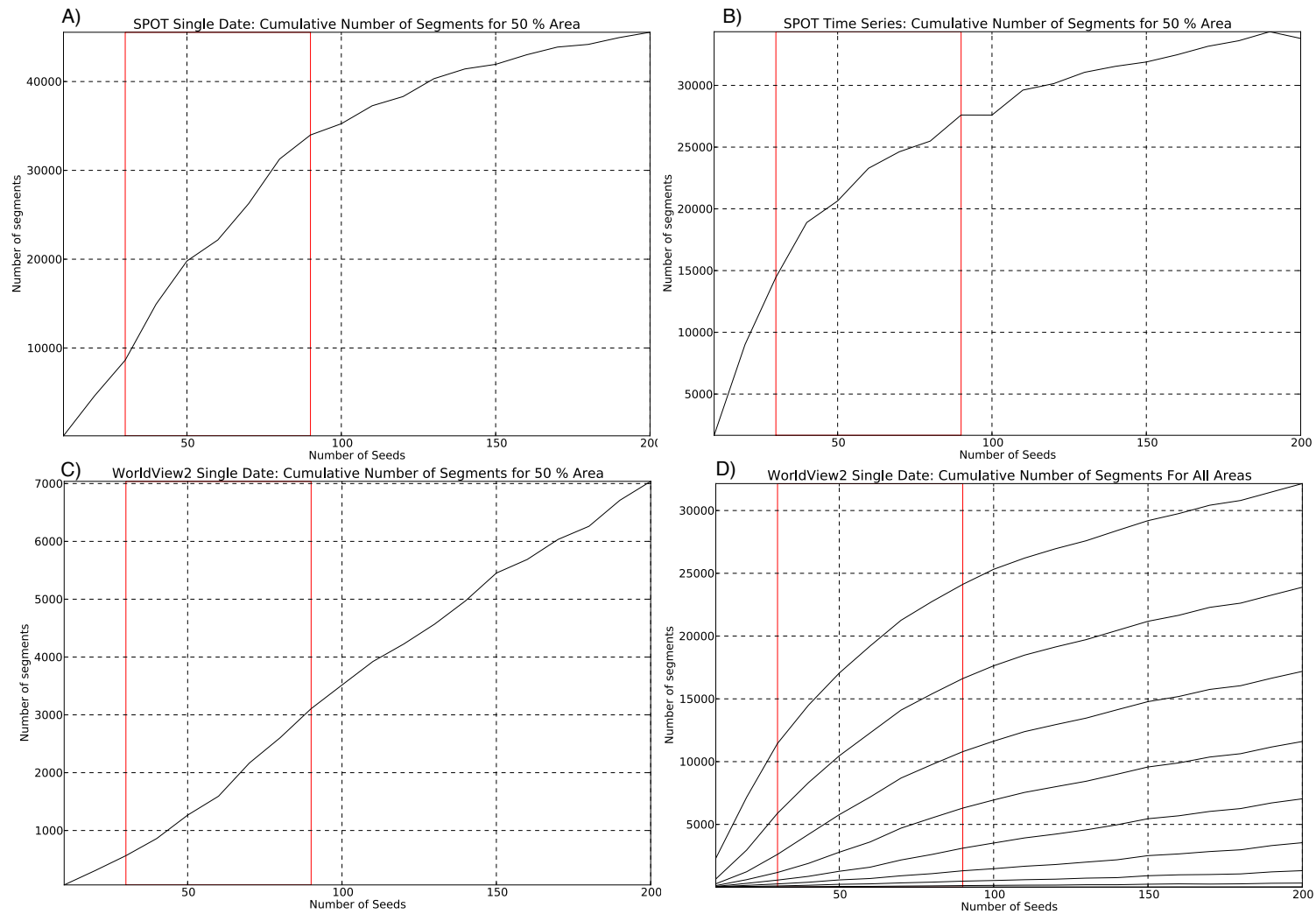


Figure 7. Cumulative number of segments required to characterise 50% of the area of (A) the single date SPOT5 scene, (B) multi-date SPOT5 time series, (C) the single date WorldView2 image and (D) the number of segments required to characterise 10–90% of the area of a single-date WorldView2 image (at 10% increments; the top line is 90%).

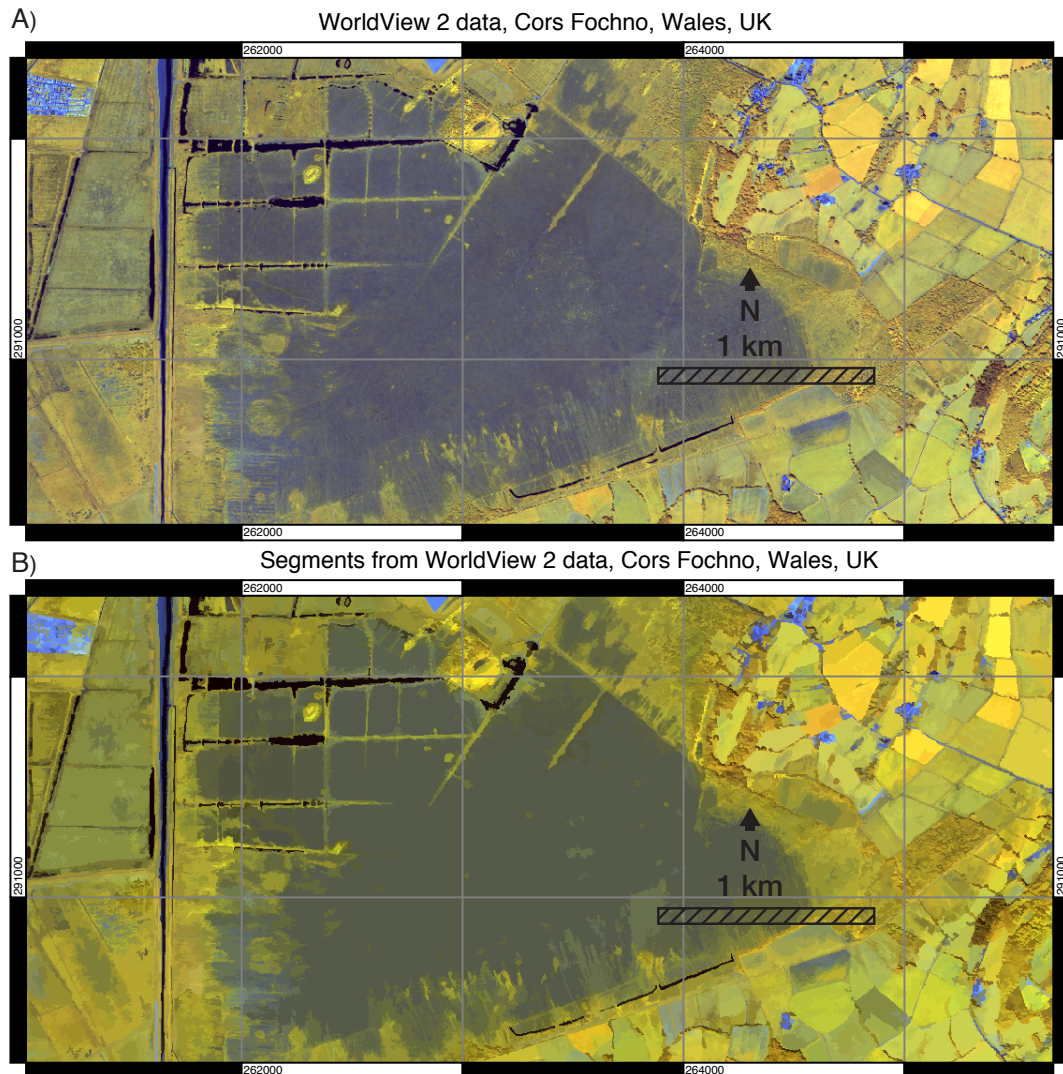


Figure 8. (A,B) Example segmentation on a WorldView2 scene, with a value of $k = 60$ and a minimum mapping unit of 0.4 ha, using a colour composite of red = NIR, green = red edge and blue = blue. The spatial pattern of land cover is well represented with the number of clumps being 0.2% of the number of pixels.

To compare the parameterisations and segmentation quality, two quantitative approaches were adopted. The first used a set of 200 reference segments (Figure 9), manually drawn on the SPOT-5 image, to calculate the precision and recall metrics of Zhang et al. [36]. The precision and recall metrics were combined to produce the f metric [36] as follows using an α of 0.5:

$$f = \frac{1}{\alpha \frac{1}{\text{precision}} + (1 - \alpha) \frac{1}{\text{recall}}} \quad (2)$$

The second approach was to use the metrics of Johnson and Xie [38] for estimating the over- and under-segmentation within the result. For this study, the metrics of Johnson and Xie [38] were calculated with all four spectral bands of SPOT-5 to create the overall global score (gs). To combine f and gs to generate a single ranked list, gs was normalised to a range of 0–1 (gs_norm) and added to f , using a weight ω :

$$gs_f = \omega gs_norm + (1 - \omega) f \quad (3)$$

For this analysis, ω was set at 0.25, giving more weight to reference segments metric f . The top parameter sets and metrics for each segmentation algorithm are given in Table 4.

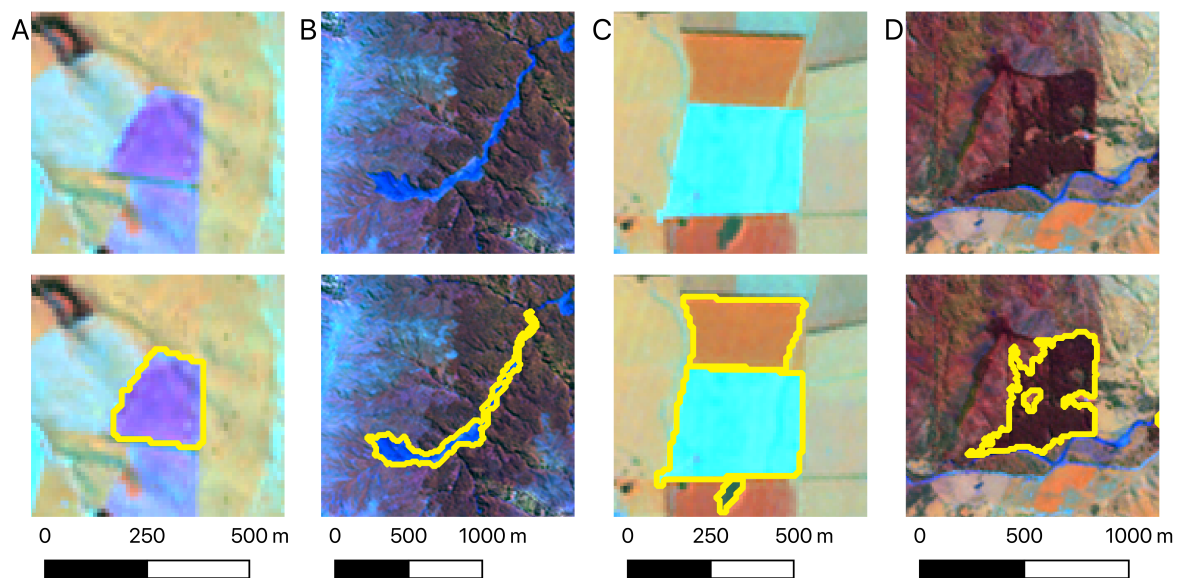


Figure 9. (A–D) Examples of the reference segments that were manually drawn.

Table 4. Comparison of the segmentation results.

Algorithm	Parameters	Rank	g_{s_f}	f	Precision	Recall	g_s	$g_{s_{green}}$	$g_{s_{red}}$	$g_{s_{nir}}$	$g_{s_{swir}}$
Shepherd et al.	k: 60, d: 10,000, min. size: 10	1	0.74	0.85	0.84	0.86	0.97	1.01	0.90	0.99	0.98
Quickshift	ratio: 0.75, kernel size: 10, max. dist.: 5, sigma: 0, lab colour space.	86	0.64	0.73	0.80	0.68	0.94	0.92	0.86	0.99	0.98
Mean-Shift	range radius; 15, convergence thres.: 0.2, max. iter.: 100, min. size: 10	253	0.56	0.61	0.55	0.70	0.93	0.95	0.87	0.96	0.94
eCognition	scale: 10, shape: 0.7, compact: 0.2	411	0.49	0.52	0.57	0.48	0.95	0.99	0.92	0.95	0.93
Felzenszwalb	scale: 10, sigma: 12, min. size: 20	539	0.47	0.46	0.49	0.43	1.10	1.14	1.04	1.17	1.04

The rank refers to the position in the order list of all performed segmentations. Therefore, using the g_{s_f} metric, there were 85 segmentations using different parameterisations of the Shepherd et al. algorithm (from this article) that were ranked higher than the highest Quickshift segmentation. For the mean-shift, there were 252 Quickshift and Shepherd et al. segmentations ranked higher than the highest mean-shift. Of those Shepherd et al. segmentations, the parameters, particularly k , were clustered with a k value of 60 being identified for the top 38 segmentations, corresponding with the analysis shown in Figure 7.

Figure 10 illustrates the segmentation results from the comparison using the parameters outlined in Table 4. The results were similar; however, two were visually more appropriate, these being the Shepherd et al. (Figure 10B) and mean-shift (Figure 10D), as they both captured the majority of the homogeneous regions as continuous segments while sub-dividing the areas of spectral variation. Both the Shepherd et al. (Figure 10B) and mean-shift (Figure 10D) algorithms demonstrated some artefacts associated with the smooth gradient transitions, as do the other methods tested. In regions of transition, the Quickshift algorithm (Figure 10C) produced a large number of very small segments, and these are not desirable within a segmentation approach; however, the homogeneous regions were well delineated with good correspondence with the reference regions. The eCognition multi-resolution segmentation (Figure 10E) aims to produce segments of similar size, and therefore, had some over-segmentation of the homogeneous features present in the Shepherd et al. (Figure 10B), Quickshift (Figure 10C) and mean-shift (Figure 10D) segmentations.

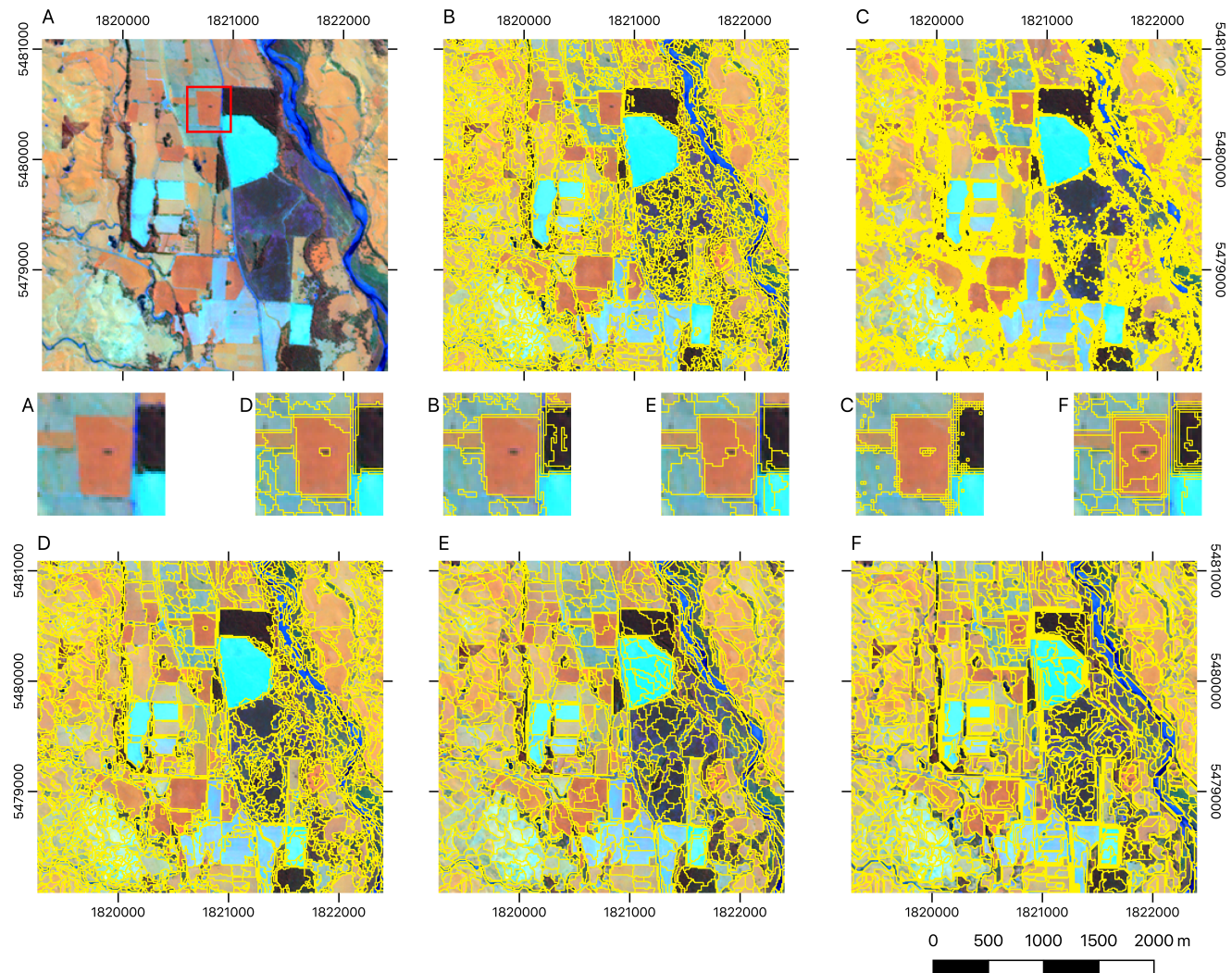


Figure 10. Comparison of segmentation algorithms with the parameters found in Table 4; the red box in (A) highlights the region of the zoom-in area. (A) SPOT-5 image with NIR, SWIR, red band combination, (B) Shepherd et al. (this article), (C) Quickshift [52], (D) mean-shift [42], (E) eCognition [28] and (F) Felzenszwalb [53].

4. Discussion

It is difficult to compare segmentation algorithms because comparison involves multiple criteria and depends on the specific application and scale [51]. However, we have demonstrated that the algorithm proposed in this paper compares favourably in most assessment metrics with the commonly-used mean-shift [42] and multi-resolution [28] segmentation approaches when applied to a typical rural and forest landscape in New Zealand. In addition, our algorithm has attributes that make it more useful in certain applications. It is readily scalable to large areas, such as nations or regions (e.g., [55]), which is desirable for preventing hard boundaries on tiles. It uses a small number of parameters, which may be consistently used across a large range of geographic areas and data types. It permits the direct setting of a minimum mapping unit, and it is freely available in open-source software.

The technique presented in this paper is being used operationally for national land-cover and land-cover change mapping of New Zealand using a single parametrisation (60 seeds and 100-pixel minimum mapping unit) on SPOT-5 and Sentinel-2 data. This is because the iterative elimination algorithm is highly scalable, being operationally applied to each of the regions of New Zealand, the largest of which is Canterbury. The Canterbury regional mosaic comprised 36 SPOT5 scenes and produced a single image with $36,533 \times 35,648$ pixels. Computationally, the segmentation of the Canterbury region required 12 GB of RAM and ran in approximately three hours on a 3-GHz x86 processor running Linux, producing 1,222,885 segments. This level of scalability is advantageous for operational use, as image tiling is not required, thus avoiding complex and time-consuming operations to merge tile boundaries, and the results are consistent across the scene. However, for larger areas, we have also implemented a tiled version of the algorithm [56], which has been used to produce a single segmentation for the Australian continent [55] using a mosaic ($141,482 \times 130,103$ pixels) of Landsat and ALOS PALSAR. This resulted in 33.7 million segments.

In hilly and mountainous terrain topographic shadowing can significantly affect the result of the image segmentation, as the shadowing alters the spectral response of those pixels. We therefore recommend that imagery be standardised to a consistent Sun and view angle before segmentation [39]. For example, a SPOT5 scene of 7753×7703 pixels, segmented with 60 clusters and eliminated to 100 pixels, generated 7715 fewer segments when standardised imagery was used, but more significantly, the distribution of those features only corresponded with patterns in land cover as opposed to land cover and topographic shadowing.

We have also successfully applied our segmentation algorithm to multi-date imagery. In this case, the segmentation is required to capture the information present in both scenes and the change between the scenes for subsequent classification. The change segments can be easily differentiated from the rest by the change in spectral reflectance between early and later dates. Automatic identification of these change segments has proven useful in developing a semi-automated method for updating land cover.

Segments are now being successfully provided across large regions using this algorithm. Therefore, new methods of assessing the quality of segmentation results are required, so that optimal parameters can be estimated. This will become increasingly important as datasets for segmentation become larger. Additionally, when segmenting large regions, the number of segments being generated is large, and their classification remains challenging, primarily due to the significant memory requirements for attribution. Existing software and methods are commonly unable to cope with this number of segments. Therefore, a new and advanced attribute table file and image format (using HDF5, called KEA [57]) and software API have been implemented within the RSGISLib [54] (<http://www.rsgislib.org>) to support rule-based object-oriented classification of the segments. For full details, see Clewley et al. [56].

5. Conclusions

This paper has outlined a new technique for the segmentation of remotely-sensed imagery, which is highly scalable through an innovative iterative elimination process. It is being used

operationally at national scales across New Zealand. The technique has clear and simple parameters and can be consistently applied across a large range of geographic areas and data types. Segmentation quality is similar to that achieved by other commonly-used algorithms such as mean-shift implemented within the Orfeo toolbox and the multi-resolution segmentation algorithm widely used within the eCognition software. The algorithm provides good performance in both memory usage and computation time. The algorithm is recommended for segmenting all modalities of remotely-sensed imagery, although pre-processing of SAR data for speckle reduction is required. Additionally, a free and open implementation of the algorithm has also been provided as part of the remote sensing and GIS software library (RSGISLib; [54]) and can be downloaded from <http://www.rsgislib.org>.

Author Contributions: J.D.S. was responsible for algorithm development. P.B. was responsible for algorithm application. J.R.D. contributed to algorithm application. All authors were involved in reviewing and approving the final version of the manuscript.

Funding: The Ministry of Business Innovation & Employment (New Zealand) funded this research under Contract C09X1709.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lucas, R.; Rowlands, A.; Brown, A.; Keyworth, S.; Bunting, P. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 165–185. [[CrossRef](#)]
- Lucas, R.; Medcalf, K.; Brown, A.; Bunting, P.; Breyer, J.; Clewley, D.; Keyworth, S.; Blackmore, P. Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 81–102. [[CrossRef](#)]
- Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [[CrossRef](#)]
- Lobo, A. Image segmentation and discriminant analysis for the identification of land cover units in ecology. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 1136–1145. [[CrossRef](#)]
- Stuckens, J.; Coppin, P.; Bauer, M. Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sens. Environ.* **2000**, *71*, 282–296. [[CrossRef](#)]
- Fuller, R.; Smith, G.; Sanderson, J.; Hill, R.; Thomson, A. The UK Land Cover Map 2000: Construction of a parcel-based vector map from satellite images. *Cartogr. J.* **2002**, *39*, 15–25. [[CrossRef](#)]
- Tarabalka, Y.; Benediktsson, J.; Chanussot, J. Spectral-Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2973–2987. [[CrossRef](#)]
- Definiens. *eCognition Version 5 Object Oriented Image Analysis User Guide*; Technical Report; Definiens AG: Munich, Germany, 2005.
- Carleer, A.P.; Debeir, O.; Wolff, E. Assessment of Very High Spatial Resolution Satellite Image Segmentations. *Photogramm. Eng. Remote Sens.* **2005**, *71*. [[CrossRef](#)]
- Fu, K.S.; Mui, J.K. A survey on image segmentation. *Pattern Recognit.* **1981**, *13*, 3–16. [[CrossRef](#)]
- Haralick, R.M.; Shapiro, L.G. Image segmentation techniques. *Comput. Vis. Graph. Image Process.* **1985**, *29*, 100–132. [[CrossRef](#)]
- Pal, N.; Pal, S. A review on image segmentation techniques. *Pattern Recognit.* **1993**, *26*, 1277–1294. [[CrossRef](#)]
- Cheng, H.; Jiang, X.; Sun, Y.; Wang, J. Color image segmentation: Advances and prospects. *Pattern Recognit.* **2001**, *34*, 2259–2281. [[CrossRef](#)]
- Dey, V.; Zhang, Y.; Zhong, M. A review on image segmentation techniques with remote sensing perspective. In Proceedings of the ISPRS TC VII Symposium—100 Years ISPRS, Vienna, Austria, 5–7 July 2010; Volume 38, pp. 5–7.
- Chen, Q.; Luo, J.; Zhou, C.; Pei, T. A hybrid multi-scale segmentation approach for remotely sensed imagery. In Proceedings of the 2003 IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003; Volume 6, pp. 3416–3419.

16. Tilton, J.C. Image segmentation by region growing and spectral clustering with a natural convergence criterion. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium Proceedings (IGARSS), Seattle, WA, USA, 6–10 July 1998; Volume 4, pp. 1766–1768.
17. Soille, P. Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1132–1145. [[CrossRef](#)] [[PubMed](#)]
18. Ghamisi, P. A Novel Method for Segmentation of Remote Sensing Images based on Hybrid GAPSO. *Int. J. Comput. Appl.* **2011**, *29*, 7–14. [[CrossRef](#)]
19. Wang, Z.; Jensen, J.R.; Im, J. An automatic region-based image segmentation algorithm for remote sensing applications. *Environ. Model. Softw.* **2010**, *25*, 1149–1165. [[CrossRef](#)]
20. Yang, Y.; Han, C.; Han, D. A Markov Random Field Model-based Fusion Approach to Segmentation of SAR and Optical Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Boston, MA, USA, 7–11 July 2008; Volume 4.
21. Zhao, F.; Jiao, L.; Liu, H. Spectral clustering with fuzzy similarity measure. *Digit. Signal Process.* **2011**, *21*, 701–709. [[CrossRef](#)]
22. Tasdemir, K. Neural network based approximate spectral clustering for remote sensing images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, BC, Canada, 24–29 July 2011; pp. 2884–2887.
23. Tarabalka, Y.; Chanussot, J.; Benediktsson, J. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognit.* **2010**, *43*, 2367–2379. [[CrossRef](#)]
24. El Zaart, A.; Ziou, D.; Wang, S.; Jiang, Q. Segmentation of SAR images. *Pattern Recognit.* **2002**, *35*, 713–724. [[CrossRef](#)]
25. Brunner, D.; Soille, P. Iterative area filtering of multichannel images. *Image Vis. Comput.* **2007**, *25*, 1352–1364. [[CrossRef](#)]
26. Weber, J.; Petitjean, F.; Gançarski, P. Towards Efficient Satellite Image Time Series Analysis: Combination of Dynamic Time Warping and Quasi-Flat Zones. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012.
27. Culvenor, D. TIDA: An algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Comput. Geosci.* **2002**, *28*, 33–44. [[CrossRef](#)]
28. Baatz, M.; Schäpe, A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *J. Photogramm. Remote Sens.* **2000**, *58*, 12–23.
29. Lucchese, L.; Mitra, S.K. Unsupervised segmentation of color images based on k-means clustering in the chromaticity plane. In Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'99), Fort Collins, CO, USA, 22 June 1999.
30. Rouse, J.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the NASA Goddard Space Flight Center, 3d ERTS-1 Symposium (NASA SP-351 I), Greenbelt, MD, USA, 1 January 1974; pp. 309–317.
31. Zhang, H.; Fritts, J.E.; Goldman, S.A. Image segmentation evaluation: A survey of unsupervised methods. *Comput. Vis. Image Underst.* **2008**, *110*, 260–280. [[CrossRef](#)]
32. Costa, H.; Foody, G.M.; Boyd, D. Supervised methods of image segmentation accuracy assessment in land cover mapping. *Remote Sens. Environ.* **2018**, *205*, 338–351. [[CrossRef](#)]
33. Novack, T.; Fonseca, L.; Kux, H. Quantitative comparison of segmentation results from ikonos images sharpened by different fusion and interpolation techniques. In Proceedings of the GEOgraphic Object Based Image Analysis for the 21st Century (GEOBIA), Calgary, AB, Canada, 5–8 August 2008.
34. Neubert, M.; Herold, H. Evaluation of remote sensing image segmentation quality—Further results and concepts. In Proceedings of the Bridging Remote Sensing and GIS 1st International Conference on Object-Based Image Analysis (OBIA), Salzburg, Austria, 4–5 July 2006.
35. Neubert, M.; Herold, H. Assessment of remote sensing image segmentation quality. In Proceedings of the GEOgraphic Object Based Image Analysis for the 21st Century (GEOBIA), Calgary, AB, Canada, 5–8 August 2008.
36. Zhang, X.; Feng, X.; Xiao, P.; He, G.; Zhu, L. Segmentation quality evaluation using region-based precision and recall measures for remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 73–84. [[CrossRef](#)]

37. Marcal, A.R.S.; Rodrigues, S.A. A framework for the evaluation of multi-spectral image segmentation. In Proceedings of the GEOgraphic Object Based Image Analysis for the 21st Century (GEOBIA), Calgary, AB, Canada, 5–8 August 2008.
38. Johnson, B.; Xie, Z. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 473–483. [[CrossRef](#)]
39. Shepherd, J.D.; Dymond, J.R. Correcting satellite imagery for the variance of reflectance and illumination with topography. *Int. J. Remote Sens.* **2003**, *24*, 3503–3514. [[CrossRef](#)]
40. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
41. Ball, G.; Hall, D. A clustering technique for summarizing multivariate data. *Behav. Sci.* **1967**, *12*, 153–155. [[CrossRef](#)]
42. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [[CrossRef](#)]
43. Ward, J.H., Jr. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [[CrossRef](#)]
44. Bezdek, J.C. Fuzzy Mathematics in Pattern Classification. Ph.D. Thesis, Cornell University, Ithaca, NY, USA, 1963.
45. Dymond, J.R.; Shepherd, J.D.; Newsome, P.F.; Gapare, N.; Burgess, D.W.; Watt, P. Remote sensing of land-use change for Kyoto Protocol reporting: The New Zealand case. *Environ. Sci. Policy* **2012**, *16*, 1–8. [[CrossRef](#)]
46. Lucas, R.; Clewley, D.; Accad, A.; Butler, D.; Armston, J.; Bowen, M.; Bunting, P.; Carreiras, J.; Dwyer, J.; Eyre, T.; et al. Mapping forest growth stage in the Brigalow Belt Bioregion of Australia through integration of ALOS PALSAR and Landsat-derived Foliage Projected Cover (FPC) data. *Remote Sens. Environ.* **2014**, *155*, 42–57. [[CrossRef](#)]
47. Michel, J.; Feuvrier, T.; Inglada, J. Reference algorithm implementations in OTB: Textbook cases. *IEEE Int. Geosci. Remote Sens. Symp.* **2009**. [[CrossRef](#)]
48. Chaabouni-Chouayakh, H.; Datcu, M. Coarse-to-fine approach for urban area interpretation using TerraSAR-X data. *IEEE Geosci. Remote Sens. Soc. Newsl.* **2010**, *7*, 78–82. [[CrossRef](#)]
49. Chehata, N.; Orny, C.; Boukir, S.; Guyon, D.; Wigneron, J.P. Object-based change detection in wind storm-damaged forest using high-resolution multispectral images. *Int. J. Remote Sens.* **2014**, *35*, 4758–4777. [[CrossRef](#)]
50. Bellakanji, A.C.; Zribi, M.; Lili-Chabaane, Z.; Mougenot, B. Forecasting of Cereal Yields in a Semi-arid Area Using the Simple Algorithm for Yield Estimation (SAFY) Agro-Meteorological Model Combined with Optical SPOT/HRV Images. *Sensors* **2018**, *18*, 2138. [[CrossRef](#)]
51. Mikes, S.; Haindl, M. Benchmarking of remote sensing segmentation methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2240–2248. [[CrossRef](#)]
52. Vedaldi, A.; Soatto, S. Quick shift and kernel methods for mode seeking. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2008.
53. Felzenszwalb, P.; Huttenlocher, D. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [[CrossRef](#)]
54. Bunting, P.; Clewley, D.; Lucas, R.M.; Gillingham, S. The Remote Sensing and GIS Software Library (RSGISLib). *Comput. Geosci.* **2014**, *62*, 216–226. [[CrossRef](#)]
55. Scarth, P.; Armston, J.; Lucas, R.; Bunting, P. A Structural Classification of Australian Vegetation Using ICESat/GLAS, ALOS PALSAR, and Landsat Sensor Data. *Remote Sens.* **2019**, *11*, 147. [[CrossRef](#)]
56. Clewley, D.; Bunting, P.; Shepherd, J.; Gillingham, S.; Flood, N.; Dymond, J.; Lucas, R.; Armston, J.; Moghaddam, M. A Python-Based Open Source System for Geographic Object-Based Image Analysis (GEOBIA) Utilizing Raster Attribute Tables. *Remote Sens.* **2014**, *6*, 6111–6135. [[CrossRef](#)]
57. Bunting, P.; Gillingham, S. The KEA image file format. *Comput. Geosci.* **2013**, *57*, 54–58. [[CrossRef](#)]

