Accurate Area Estimation by Data-Driven Decomposition of Mixed Pixels

Maurice S. klein Gebbinck, Theo E. Schouten

University of Nijmegen, Computing Science Institute
P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
E-mail: mauricek@cs.kun.nl

December 2, 1996

Abstract

There are many image processing applications where the area of an object has to be estimated as accurately as possible. A well-known example is the area estimation of agricultural fields, which is of great importance for the management of the agricultural subsidy system of the European Union. The area of an object can be estimated using both classification, which allocates a pixel to a single class, and decomposition, which divides a pixel between several classes. Since decomposition is better at handling mixed pixels—pixels comprising multiple classes—which are often found at object boundaries, area estimation by decomposition is expected to be more accurate. To test this hypothesis, a data-driven decomposition method was developed and applied to a series of artificial satellite images of increasing complexity. Data-driven decomposition was able to estimate the percentage of each component of a pixel with an average error of 5%. Narrow structures were processed correctly, and isolated pixels were detected using a simple threshold. A quantitative comparison with the results of three other methods found in the literature showed that the area estimates of data-driven decomposition were significantly more accurate. This study suggests that data-driven decomposition is an accurate area estimation method which is worth further research using real satellite images.

Keywords: (crop) area estimation, subpixel resolution, mixed pixels, linear mixture model, classification, decomposition, (artificial) satellite imagery.

Classification: AMS-1991: 68U10, 68U05; CR-1991: I.4.7, I.4.6, I.3.3.

1 Introduction

Every year the European Union spends tens of milliards of dollars on subsidies for the agricultural sector. The amount of money that is granted to each farmer depends on the type and acreage of the crops that are grown. Since it is impossible to check the claims of Europe's six million farmers using human controllers only, the initial crop classification and area estimation are done by computer using satellite remotely sensed images. One of the tasks that have to be automated is the generation of a ground cover map of the area under observation. Two approaches that have been suggested in the literature are pixel-based (model-driven) and region-based (data-driven) classification. The first method classifies each pixel individually, whereas the second method first partitions the image into spectrally similar regions and then classifies each region [3, 8]. Although the segmentation requires a considerable computational effort, the region-based approach is usually preferred because misclassifications due to the within-crop variation are reduced.

Since the spatial resolution of space-borne scanners is relatively low compared to the size of the agricultural fields, many pixels at the field boundaries are made up of more than one ground cover type. In scenes where the average field size is small, like in some Mediterranean countries, the percentage of these so-called mixed pixels can be as high as 30% [8]. Mixed pixels are not only difficult to classify, they also pose a problem for the segmentation algorithm. Fortunately, with the use of a linear mixture model it is possible to estimate the fractions of the ground cover types a mixed pixel is composed of. While in recent years this technique was applied on a pixel basis in several studies [1, 7, 10], it has not been combined with the region-based approach yet. The purpose of this study was to develop a data-driven decomposition method and to investigate whether its area estimates are more accurate than the results of an area estimator based on classification.

Although the suggested method was designed with the area estimation of agricultural fields in mind, it can also be used for other applications. A few examples of applications where accurate area estimates are required are tumor size measurement from medical images, and non-fill detection of chip molds by an industrial vision system. In the present report the data-driven decomposition method is described in detail, therefore it should not be too difficult to tailor the system for these applications. To analyse the resulting system, supervised information at subpixel level about the image is needed. For the present study a series of satellite images was generated, which resembled real Landsat Thematic Mapper images. With these images the operation of the data-driven decomposition method was analysed. Furthermore, they were used for a quantitative comparison of the accuracies of the suggested method and three other area estimation techniques, i.e. model- and data-driven classification, and model-driven decomposition.

2 Data-driven decomposition

An object that is large compared to the resolution cell size consists of a cluster of pure pixels, representing the interior of the object, surrounded by layer of mixed pixels, which are positioned at the object boundaries. Data-driven decomposition attempts to divide each mixed pixel between the pure pixel clusters that are in its neighbourhood. The linear mixture model upon which the decomposition technique is based is briefly explained in Section 2.1. Section 2.2 describes which image processing techniques were used to identify the objects that were part of the mixed pixel.

2.1 Linear mixture modelling

For remotely-sensed images the linear mixture model has shown to be a good abstraction for mixed pixels. Although there are many variants on this model (Pech *et al.* [6], Settle and Drake [10], Klein Gebbinck and Schouten [4]), the classical linear mixture model by Horwitz *et al.* [2] is still much used. Following the nomenclature of Settle and Drake, this model is given by

$$\mathbf{x} = \mathbf{M}\mathbf{f} + \mathbf{e}, \tag{1}$$

where vector $\mathbf{x}$ denotes the n-dimensional pixel, and vector $\mathbf{f}$ represents the proportions of the c objects that take part in the mixed pixel. The columns of the $(n\text{x}c)$ matrix $\mathbf{M}$ list the mean reflectance of the contributing objects. The variation in the reflectance of the objects, which are assumed to have multivariate normal distributions, is covered by the error vector $\mathbf{e}$. In theory $\mathbf{e}$ is also distributed multivariate normally, having a zero mean and a variance-covariance matrix $\mathbf{N}$ depending on the distributions as well as the fractions of the objects. However, it has been argued (Horwitz *et al.* [2]) that $\mathbf{N}$ can be approximated by the averaged variance-covariance matrices of the objects without much loss of accuracy, but resulting in a set of equations that is much more tractable.

If the number of components is at most equal to the dimensionality of the data ($c \leq n$), Equation (1) can be solved in infinitely many ways. From these solutions the one with the highest probability is chosen, which according to the maximum likelihood approach minimizes

$$\mathbf{e}^T\mathbf{N}^{-1}\mathbf{e} + \ln|\mathbf{N}| = (\mathbf{x} - \mathbf{M}\mathbf{f})^T\mathbf{N}^{-1}(\mathbf{x} - \mathbf{M}\mathbf{f}) + \ln|\mathbf{N}| \tag{2}$$

with $\ln|\mathbf{N}|$ representing the natural logarithm of the determinant of $\mathbf{N}$. For the case that the variance-covariance matrix $\mathbf{N}$ is independent of the fractions vector $\mathbf{f}$, it can be derived analytically that the optimal solution is given by

$$\hat{\mathbf{f}} = \mathbf{U}\mathbf{M}^T\mathbf{N}^{-1}\mathbf{x} \quad \text{with} \quad \mathbf{U}^{-1} = \mathbf{M}^T\mathbf{N}^{-1}\mathbf{M}. \tag{3}$$

Imposing the condition that the proportions should sum to one gives a slightly different solution which can be found through standard Lagrangian analysis. The other constraint that should be satisfied is that all proportions should be positive (or zero); although there are many ways to accomplish this, the fastest and simplest method, which is used most often, is to set negative proportions to zero and normalise the proportions vector. For further reading about both constraints is referred to Settle and Drake [10].

2.2 Data-driven aspects

Before mixed pixels can be divided between the objects they are composed of, the pure pixel regions have to be determined. If no ancillary information (e.g. cadastral) is available, the regions have to be found through segmentation of the image. Two techniques that are often used for segmentation are edge detection, which looks for local discontinuities in the image, and region growing, which groups neighbouring pixels that are similar. These techniques do not only find the pure pixel regions, but they also suggest which pixels are mixed and need to be decomposed. For further information on the segmentation of optical satellite images is referred to [8] since an in-depth discussion of this subject is considered to be beyond the scope of the present report.

Every ground cover class that is to be distinguished must be characterised by a mean vector and a variance-covariance matrix. If the pixel-based approach is followed, supervised data is needed to derive the parameters of these multivariate normal distributions. In this case decomposition is said to be model-driven. The region-based approach, however, calculates the means and variance-covariance matrices from the regions themselves, which are assumed to contain pure pixels of a single class only. Since all parameters are derived from the satellite image and no supervised data is needed, decomposition in this case is called data-driven. An important advantage of data-driven decomposition is that local variations in the statistics of the ground cover classes (e.g. due to variations in ground slope, elevation, orientation, etc.) can be reckoned with. However, for small regions the estimation of the parameters will tend to be unreliable, especially the variance-covariance matrix. For regions containing no more pixels than the number of spectral bands, the variance-covariance matrix will even be singular, which makes decomposition impossible. A solution for this problem is to classify these small regions first and subsequently use supervised data to determine the parameters, similar to the model-driven method. Other solutions are to use the parameters of a spectrally similar but larger region or to calculate some sort of class-independent variance-covariance matrix from all regions.

When working on a pixel basis, the surroundings of a pixel usually are not considered. Therefore, decomposition of a pixel can only be done regarding all ground cover classes at the same time. Since the number of classes the decomposition algorithm can handle simultaneously is at most one more than the number of spectral bands, the resulting ground cover map—in this case called mixture map—can display no more ground cover types either. With the region-based approach on the other hand, spatial information present in the image can be used to determine which ground cover classes a mixed pixel probably is composed of. This way, for each mixed pixel different classes can be considered, so the number of ground cover types shown by the ground cover map is no longer restricted. More important though is that the number of classes that have to be handled simultaneously is reduced significantly, which results in a much more accurate and reliable decomposition. In fact, most mixed pixels in agricultural areas are surrounded by only two regions, hence often just two classes have to be considered. However, agricultural fields are usually separated by a very narrow strip of grass, a ditch, etc; these edge classes have to be considered as a component of the mixed pixel as well. Since such small structures do not form a neighbouring region, their statistics cannot be extracted from the image, so supervised data is needed.

```
FOR each suspected mixed pixel DO
  construct a set containing all pairs of two 8-connected regions;
  add all pairs of one 8-connected region and one edge class to the set;
  add all triplets of two 8-connected regions and one edge class to the set;

  FOR each element of the set DO
    calculate the unreliability of decomposing the mixed pixel by this combination;
  OD

  IF the lowest unreliability is below the threshold THEN
    divide the pixel according to the corresponding combination;
  ELSE
    mark the pixel isolated;
  FI
OD
```

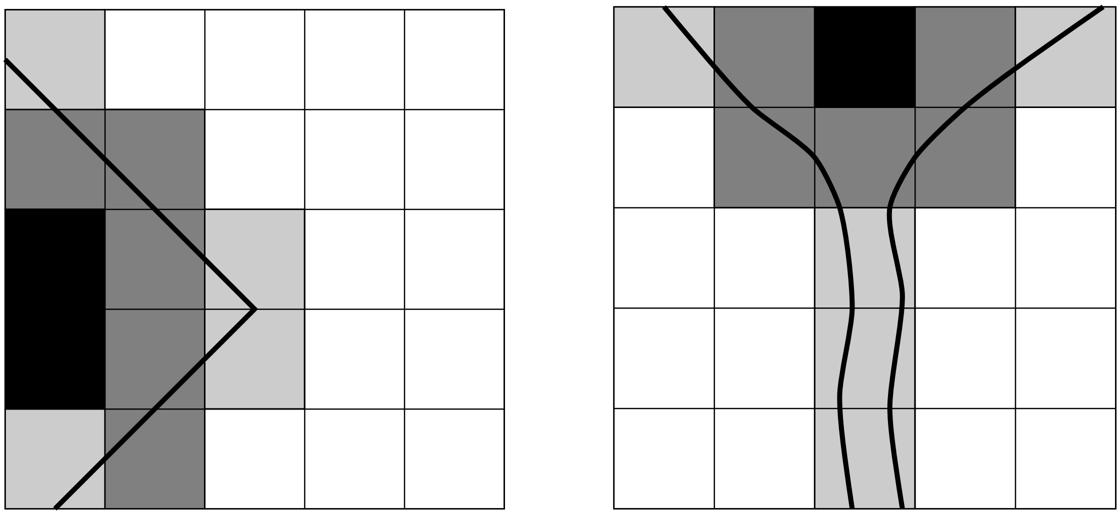
Algorithm 2.1: First stage of the data-driven decomposition method.

As mentioned at the beginning of this section, sometimes the regions in an image have to be found through segmentation. In this case all pixels that are not assigned to a region are possibly mixed pixels. However, some of these pixels may represent a small object like a house, a high tension mast, etc and should not be decomposed as usual. Fortunately, the linear mixture model provides a measure for the unreliability of each decomposition which can be used to decide whether a pixel is probably mixed or not. If the unreliability is below a certain threshold, the pixel should be divided between its neighbouring regions; if the unreliability is too high, the pixel should remain isolated. Since the unreliability measure is based on the Mahalanobis distance between the pixel and its approximation by the decomposition algorithm, a rough indication of the threshold setting can be given. Suppose the classes the mixed pixel is to be decomposed into show no correlation between the different spectral bands. If for each band a deviation of say 2σ is allowed, the Mahalanobis distance is equal to

$$\begin{bmatrix} 2\sigma_1 & 2\sigma_2 & \ldots & 2\sigma_n \end{bmatrix} \begin{bmatrix} \sigma_1^{-2} & & & 0 \\ & \sigma_2^{-2} & & \\ & & \ddots & \\ 0 & & & \sigma_n^{-2} \end{bmatrix} \begin{bmatrix} 2\sigma_1 \\ 2\sigma_2 \\ \vdots \\ 2\sigma_n \end{bmatrix} = 4n. \tag{4}$$

Of course the distance is different in case the spectral bands are correlated, but a threshold setting equal to four times the number of bands seems to be a well-educated guess.

The unreliability measure can also be used to further reduce the number of classes that have to be handled simultaneously, increasing both accuracy and reliability. Since a pixel usually is composed of only a few regions with possibly an edge class, not all neighbouring regions have to be considered at the same time. By successively decomposing the pixel using all combinations of two or three neighbouring regions and edge classes, the most reliable mix can be selected. An implementation of this strategy is given in Algorithm 2.1. In this algorithm, only a small fraction of all possible combinations are tried in order to save time as well as to prevent undesirable mixes of classes from being chosen. For instance, combinations of three or more surrounding regions could be investigated, but they would give only marginal improvement because few such pixels exist, and probably only the classes that contribute little to the pixel are wrongly estimated. Furthermore, no more than one edge class at the same time is used in order to stick as closely as possible to the data-driven approach. So far the pixel's neighbourhood that is scanned to determine the bordering regions has not clearly been defined. In Algorithm 2.1 only regions which are directly connected to the pixel are considered. This strategy guarantees that a region remains contiguous even after assignation of mixed pixels.



(a)

(b)

Figure 1: The white and black squares are pure pixels of different regions, the grey squares are mixed pixels. For the light grey pixels the correct regions cannot be found by looking at the pure pixels in their direct neighbourhood alone.

Preliminary experiments showed that with Algorithm 2.1 the majority of the mixed pixels could successfully be decomposed. Unfortunately, the correct regions were not always found, which led to erroneous results. If the unreliability of the decomposition was coincidentally below the threshold, the mixed pixel was divided between incorrect regions, but if no regions were found at all or unreliability was too high, the pixel was marked isolated. In Figure 1 two situations showing the flaws of Algorithm 2.1 are given. In both cases the direct neighbourhood of each light grey pixel consists of some other mixed pixels and some pure pixels, but it does not contain all contributing regions at the same time. Figure 1(a) shows the common problem that mixed pixels hardly ever occur isolated but nearly always appear in clusters. Expansion of the neighbourhood with another or even two layers of pixels could solve the problem, but it could also result in regions that are no longer contiguous. For narrow structures like the one in Figure 1(b), which depicts the flow of a river into a lake, this strategy is not applicable since the necessary expansion of the neighbourhood could be arbitrarily large. A much better solution to these problems is given by Algorithm 2.2, which can be considered as the second stage of an overall algorithm started by Algorithm 2.1. In the second stage, the spatial information of the mixed pixels which were processed in the first stage is used to find regions that were not considered before because the neighbourhood did not contain pure pixels of those regions. The second stage is executed iteratively, using new spatial information obtained in the previous iteration. An advantage of this strategy is that the contingency of each region is guaranteed since there always exists a path from the mixed pixel to a pure pixel of each contributing region. Another advantage is that narrow structures can now also be handled without the need to expand the neighbourhood enormously. Although the iterative nature of the second part may lead to a significant increase in computation time in theory, this proved to be not the case in practice. Not only was the vast majority of the mixed pixels handled by the first stage, most remaining mixed pixels were of the type shown in Figure 1(a) and could nearly all be dealt with in the first iteration. Pixels belonging to narrow structures may need more iterations—Figure 1(b) needs three—depending on the position of the closest pure pixel of the main region, but they are small in number.

```
WHILE previous iteration decomposed at least one pixel DO
  FOR each pixel marked isolated DO
    determine new regions from 8-connected pixels decomposed in the previous iteration;
    construct a set containing all pairs of two new regions;
    add all pairs of one new region and one edge class to the set;
    add all triplets of two new regions and one edge class to the set;
    add all pairs of one old and one new region to the set;
    add all triplets of one old region, one new region and one edge class to the set;

    FOR each element of the set DO
      calculate the unreliability of decomposing the mixed pixel by this combination;
    OD

    IF the lowest unreliability is below the threshold THEN
      divide the pixel according to the corresponding combination;
    ELSE
      mark the pixel isolated for the next iteration;
    FI
  OD
OD
```

Algorithm 2.2: Second stage of the data-driven decomposition method.

```
FOR each pixel marked isolated DO
  determine all regions from the 8-connected pixels;
  construct a set containing all pairs of one region and one arbitrary class;

  FOR each element of the set DO
    calculate the unreliability of decomposing the mixed pixel by this combination;
  OD

  divide the pixel according to the combination with the lowest unreliability;
OD
```

Algorithm 2.3: Third stage of the data-driven decomposition method.

Pixels that are still marked isolated after the second stage do not resemble their surroundings enough and are considered to be truly isolated pixels. Each such pixel is assigned to a new region or joined with a neighbouring isolated pixel in case their characteristics are similar. However, these isolated pixels are probably still mixed pixels that contain a fraction of one of their neighbouring regions. Therefore, the data-driven decomposition algorithm is extended with a third stage given by Algorithm 2.3. In this final stage isolated mixed pixels are divided between an arbitrary class and a neighbouring region. This time no threshold is used, so the pixel is always decomposed according to the most likely combination.

3 Application to artificial satellite images

To evaluate the data-driven decomposition method, it was applied to artificial satellite images. The reason for using artificial instead of real satellite images was that for the real images no ground truth at subpixel level was available, which was needed for a quantitative analysis of the accuracy of the method. For the simulated images this information was available because of the way the images were constructed—see Section 3.1. Apart from being necessary for a quantitative analysis, the ground truth was also used to obtain a reference segmentation of the image. This way, no segmentation had to be made to identify the pure regions and the mixed pixels which would probably influence the results significantly. The experiments that were carried out can be divided into two categories. The first set of experiments—described in Section 3.2—was designed to determine strong and weak points of the data-driven decomposition algorithm itself. The goal of the remaining experiments—see Section 3.3—was to compare the results of data-driven decomposition with those of other area estimation algorithms.

3.1 Simulating satellite images

The generation of artificial satellite images can be done in many ways with varying degrees of realism. A method which preserves both first- and second-order statistics has been described by Schouten *et al.* [9] and will be briefly summarised in this section.

The basis of the artificial image was a binary edge image of high resolution, obtained by scanning and processing a hand-drawn sketch. After identification of the regions using a floodfill algorithm, a mapping was made between the region numbers and the available ground cover types; edges could be assigned a ground cover type as well. For each ground cover type a template was compiled using regions from real satellite images, resulting in the preservation of simple point statistics (e.g. mean, variance-covariance) as well as texture. These regions were selected using supervised data and were processed to ensure they could be seamlessly connected when composing the template. Next, the composition of each pixel of the low resolution satellite image was determined by looking at a square block of pixels from the edge image. Finally, the image itself was generated by filling each region using the template determined by the mapping. If a pixel was mixed, the contributions of several templates were added according to their proportion. In case no ground cover type was chosen for the edges, they were simply ignored.

The intermediate data structures created when simulating the satellite image were also used for the generation of a reference segmentation. The exact format in which the segmentation was saved can be found in Appendix A, but basically each region was described by run-length encoding pure pixels sequences, followed by listing the locations of the mixed pixels together with their fractions. If a ground cover type was assigned to the edges, all pixels containing edge subpixels were considered to belong to a so-called global region which was saved as well. The global regions were not needed by the data-driven decomposition method but were used solely for evaluation purposes. A problem appears when the pure pixels of a narrow structure like a river are not all connected but are separated by mixed pixels. Such a structure could be described using a single region, but a segmentation program could never achieve this. Whether these structures should therefore be split or not is not clear yet.

3.2 Evaluation of the data-driven decomposition method

In these experiments three images of increasing complexity were processed by the data-driven decomposition algorithm. In the first image the edges were ignored which led to simple transitions between the neighbouring regions. In the second image the edges were assigned the ground cover type bare soil, which gave mixed pixels that were more complex. In the third image a number of isolated mixed pixels were added in order to determine the ability to detect these pixels and to evaluate the threshold setting. During the decomposition process, the performance was monitored using several measures.

3.2.1 Measuring accuracy and unreliability

The measure used to evaluate the accuracy of the decomposition of each mixed pixel had to reflect the two types of errors that could be made. The first type was that the estimated fraction of a region deviated from the true fraction, and the second type was that the chosen region itself was incorrect. However, it was much more practical to calculate the accuracy based on ground cover types than on region numbers. Suppose for instance that two regions with the same ground cover type contributed to the same mixed pixel. While the fraction of each individual region could never be determined, calculation of the total fraction of that ground cover type was well possible.

This problem could be circumvented by assigning different ground cover types to neighbouring regions and using edge classes differing from all regions in the entire image, but a measure based on ground cover types was much more flexible. This way, the two errors mentioned above could be unified by considering a mixed pixel to be composed of all ground cover types present in the image, taking a fraction equal to zero for classes that do not belong to a pixel's composition. The accuracy measure for pixel p could now be given as

$$e_p = \frac{1}{2} \sum_i \left| \hat{\mathrm{f}}_{p,i} - \mathrm{f}_{p,i} \right| , \tag{5}$$

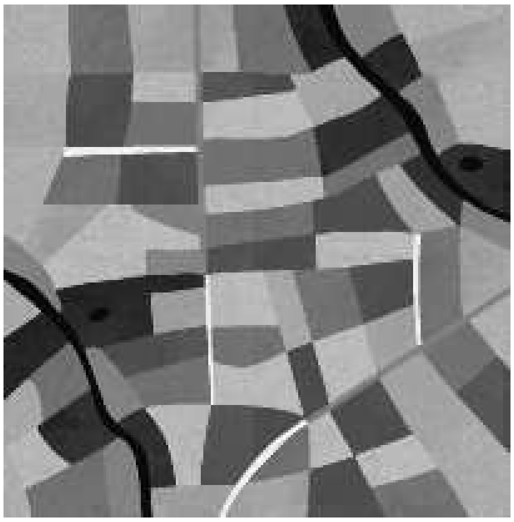
where $\hat{\mathbf{f}}_\mathbf{p}$ and $\mathbf{f}_\mathbf{p}$ were the estimated and real fractions vector respectively, and i ranged over all ground cover types. The summed absolute error was multiplied with $\frac{1}{2}$ to compensate for the fact that each deviation of the true fraction of one class was balanced by an opposite deviation of the true fraction of one or more of the other classes. These errors were presented in a histogram to show the global statistics and in an image to view the spatial properties.

To evaluate the unreliability of each decomposition, a measure was used which was based on the Mahalanobis distance. An exact description of the measure as calculated by the decomposition method can be found in [4]. Similar to the accuracy of the decomposition process, the unreliability of each decomposition was also presented using both a histogram and an image. With these means the threshold setting was assessed and certain problem areas were identified.

3.2.2 Results

Experiments were carried out on three simulated satellite images of 203x205 pixels with six spectral bands. All images contained pixels constructed out of 4x4 subpixels, using 10 different ground cover types. In image 1 no ground cover type was assigned to the edges, resulting in 2479 mixed pixels that were relatively easy to decompose. Figure 2(a) displays band 5 of the image, and the positions of the mixed pixels are shown in Figure 2(b). Data-driven decomposition of the mixed pixels in image 1 resulted in an average error in accuracy of 2.7%. The threshold was set at 4x the number of bands, i.e. 24.0. At stage 2, two iterations were needed during which subsequently 1 and 0 mixed pixels were processed; 12 pixels were left isolated. The majority of the pixels which could not be decomposed accurately consisted of three or more regions, something the algorithm could not handle. Another source of error was that some mixed pixels could be decomposed a little bit more reliably using a wrong combination of regions than using the correct combination. A last cause of error was that in a few cases the pixel could be decomposed correctly but with an unreliability that was just above the threshold. As a result, these pixels were subjected to the third stage, which was meant for truly isolated pixels only. Figure 3 shows how the errors in accuracy and the unreliability were distributed. Figure 3(b) confirms that a threshold setting of 24.0 was a good choice.

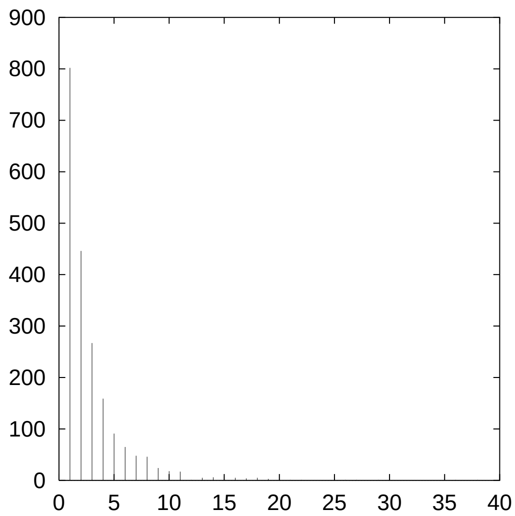
In image 2 the edges were assigned the ground cover type soil, resulting in an image very similar to image 1. The difference between these images can only be seen by zooming in on for instance the lower left corner, as is done in Figure 4. Not only did the composition of the mixed pixels become more complex, the number of mixed pixels also increased to 4928—compare Figures 2(b) and 5(a). As a result, image 2 contained several narrow structures consisting largely of mixed pixels—see Figure 5(b)—which were good test cases for the previously described algorithm. With a threshold setting of 24.0, the mixed pixels were decomposed with an average error in accuracy of 4.9%. For image 2 a total of 5 iterations were needed at stage 2 during which subsequently 29, 14, 4, 4 and 0 mixed pixels were decomposed. Iterations 3 and 4 were required to process the narrow structure depicted in Figure 5(b), which contained a massive block of mixed pixels of height 10. A total of 5 mixed pixels remained after the second stage and were marked isolated. In addition to the reasons already mentioned in the previous paragraph, spectral confusion was another source of error. Several pixels being composed of rice (±90%) and soil (±10%) were thought to have a much smaller fraction of rice (±60%), and more soil (±20%) and water (±20%). Much more frequent, however, was the confusion between soil, barley and grass. It appeared that the class



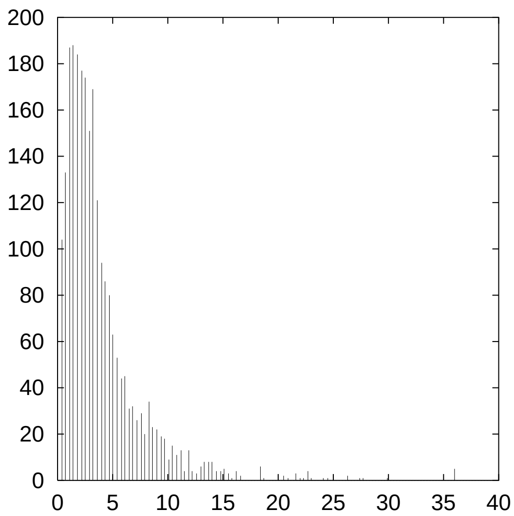
(a)

(b)

Figure 2: Image 1, the image in which edges are ignored. (a) The most contrast-rich band 5. (b) Positions of the mixed pixels shown in white.

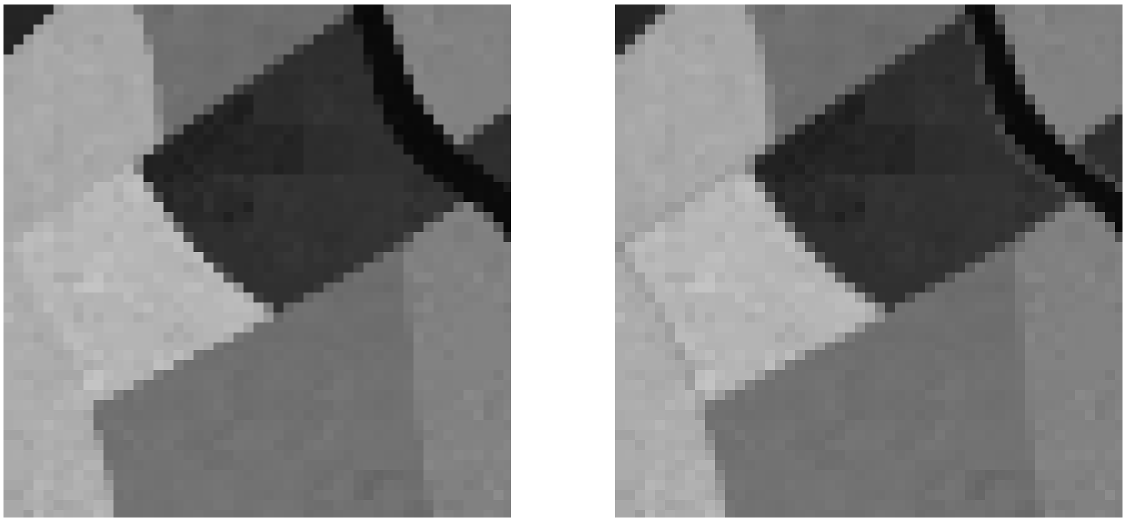


(a)



(b)

Figure 3: Statistics on decomposing image 1 using a threshold setting of 24.0. (a) Histogram of the errors in accuracy (40%–100% contained 8 pixels). (b) Histogram of the unreliability between stages 2 and 3; the last bar holds the number of pixels with an unreliability over 36.0.



(a)

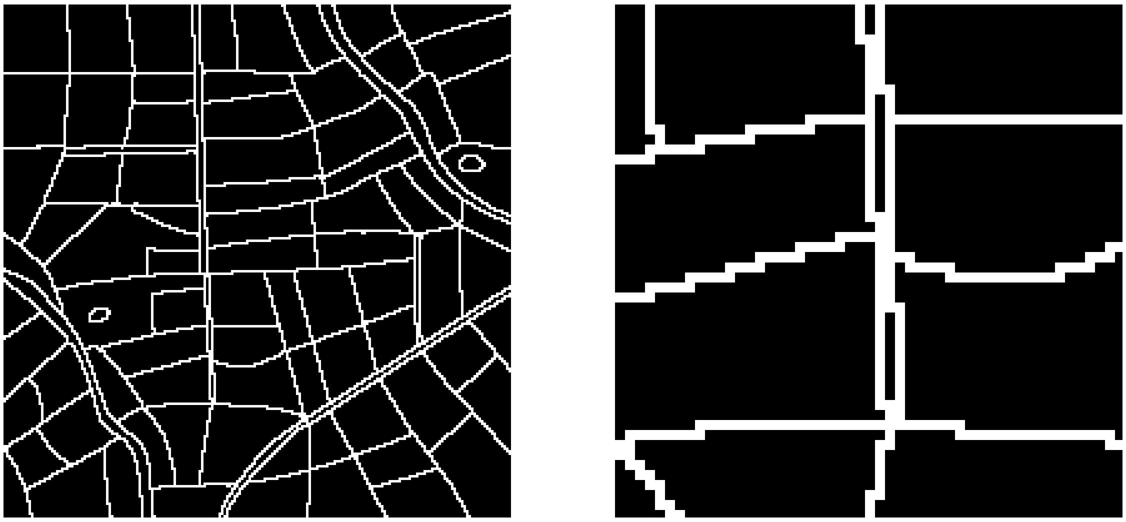
(b)

Figure 4: The difference between image 1 and 2 can be found at the field boundaries. (a) Lower left corner of image 1. (b) Lower left corner of image 2.

mean of grass was a linear combination of the class means of soil and barley, leading to spectral confusion for all pixels containing both soil and barley. Since the variance-covariance of grass was higher than the ones of soil and barley, the area covered by grass was overestimated while soil and barley were underestimated. In Figure 6 the histograms of the errors in accuracy and the unreliability are presented, while Figure 7 shows these measures for each pixel individually. The line fragments in Figure 7(a) (e.g. the diagonal line west of the center) were located at the borders of fields with barley and fields with grass. With edges of soil, spectral confusion occurred along the entire border. As can be seen in Figure 7(b), these pixels could not be identified based on the unreliability of their estimations.

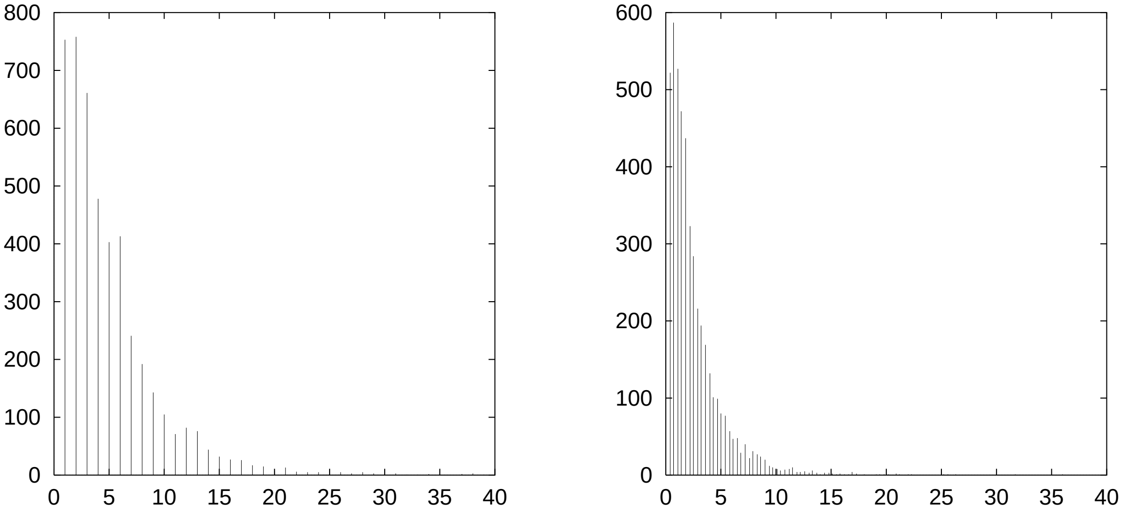
Image 3 was similar to image 2, differing only in the isolated pixels that were added—see Figure 8. An unrealistically high number of 1.0% of the image pixels was selected at random to be replaced by isolated mixed pixels. Each isolated pixel was constructed of a class not present in its neighbourhood with a fraction of at least 75.0%, completed with the ground cover type of a neighbouring region. Of these 416 pixels, 47 replaced pixels that were already mixed, resulting in 5297 mixed pixels to be processed. Data-driven decomposition of these pixels with a threshold setting of 24.0 gave an average error in accuracy of 5.0%. During the 5 iterations of the second stage, 36, 14, 4, 4, and 0 mixed pixels were decomposed, leaving 369 pixels for stage 3. Experiments with images where 0.1% of the pixels was isolated showed that most isolated mixed pixels were recognised as such. Those that were not identified were frequently placed at positions already occupied by mixed pixels. Since their neighbourhoods usually contained multiple regions, often an erroneous decomposition in stage 1 or 2 was accepted because of spectral confusion. Another reason was that some isolated pixels had soil as their main component. Because soil was also the class of the edges and thus was always considered as a possible component, a correct decomposition could already be achieved in stage 1. As can be seen in Figures 9(a) and 10(a), isolated mixed pixels could be decomposed as accurately as ordinary mixed pixels. Figures 9(b) and 10(b) show that with a threshold of 24.0 isolated mixed pixels could easily be identified.

The data-driven decomposition algorithm used a threshold to determine whether a mixed pixel could be decomposed reliably enough or not. If the threshold was set too low, pixels which could be decomposed correctly were marked isolated and erroneously processed in stage 3. If on the



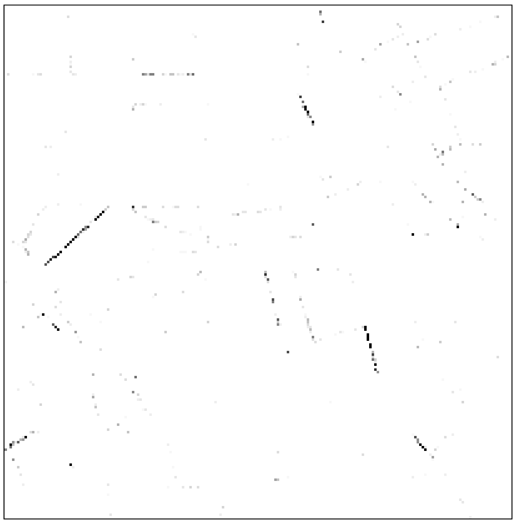
(a) (b)

Figure 5: Image 2 with edges consisting of soil. (a) Positions of the mixed pixels shown in white. (b) Blowup of the narrow structure in the south-west quadrant.

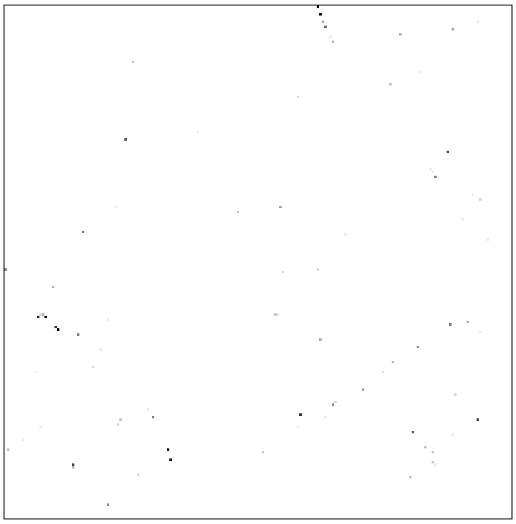


(a) (b)

Figure 6: Histograms regarding image 2 with a threshold setting of 24.0. (a) Errors in accuracy (40%–100% contained 28 pixels). (b) Unreliabilities after stage 2.

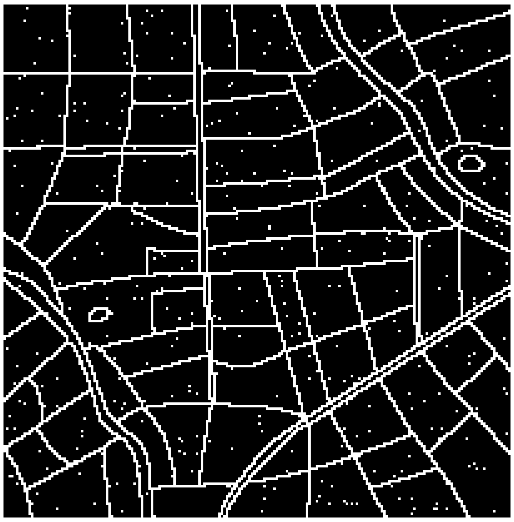


(a)

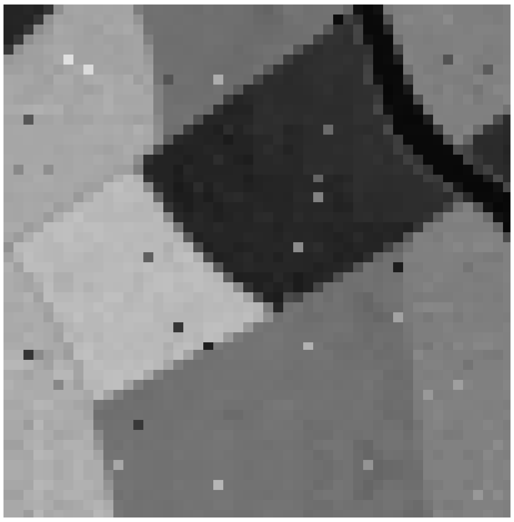


(b)

Figure 7: Spatial distributions of image 2 using a threshold of 24.0. (a) Errors in accuracy of more than 10%. The pixels get darker with decreasing accuracy. (b) Pixels decomposed in stages 1 and 2 with an unreliability of more than 10.0. The more unreliable the decomposition, the darker the pixel.

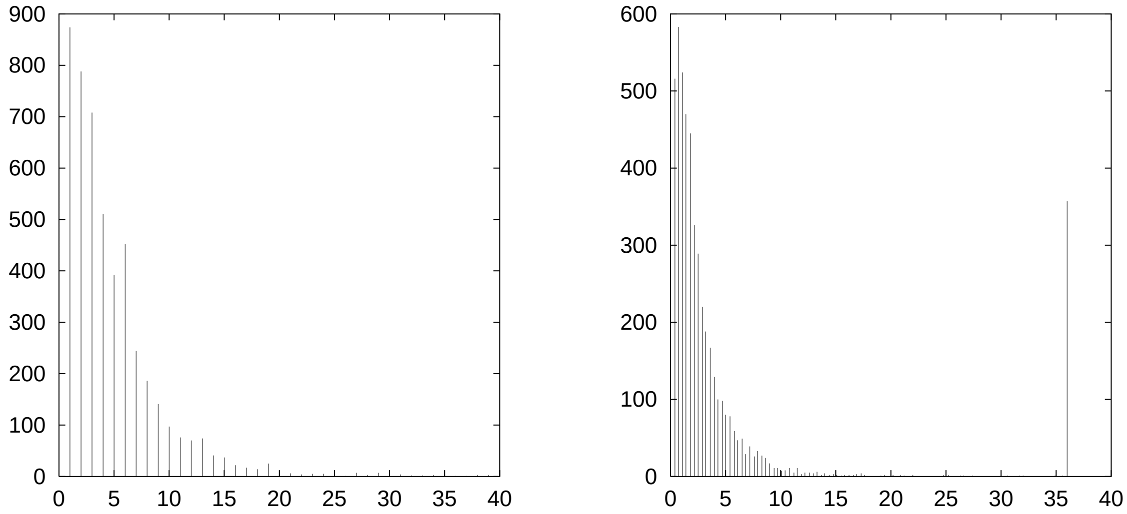


(a)



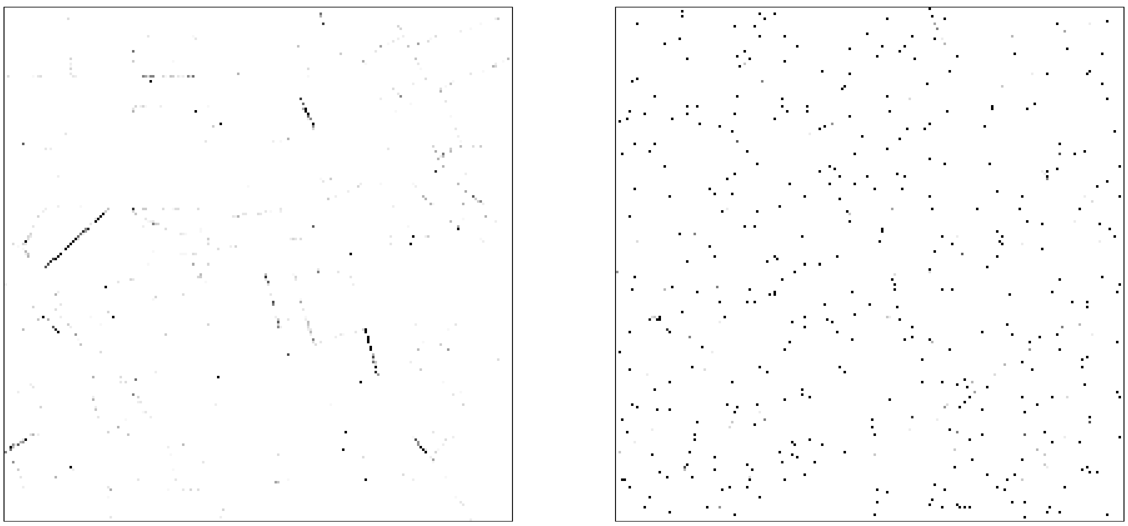
(b)

Figure 8: Image 3 with isolated mixed pixels. (a) Positions of the mixed pixels shown in white. (b) Lower left corner of image 3.



(a) (b)

Figure 9: Statistics on image 3 (threshold is 24.0). (a) Errors in accuracy (41 pixels in 40%–100%). (b) Unreliability before stage 3. Last bar holds pixels with an unreliability of 36.0 to several thousands.



(a) (b)

Figure 10: Spatial distributions of image 3 using a threshold of 24.0. (a) Errors in accuracy of more than 10%. The pixels get darker with decreasing accuracy. (b) Pixels decomposed in stages 1 and 2 with an unreliability of more than 10.0. Many of the truly isolated mixed pixels (black) can be identified.

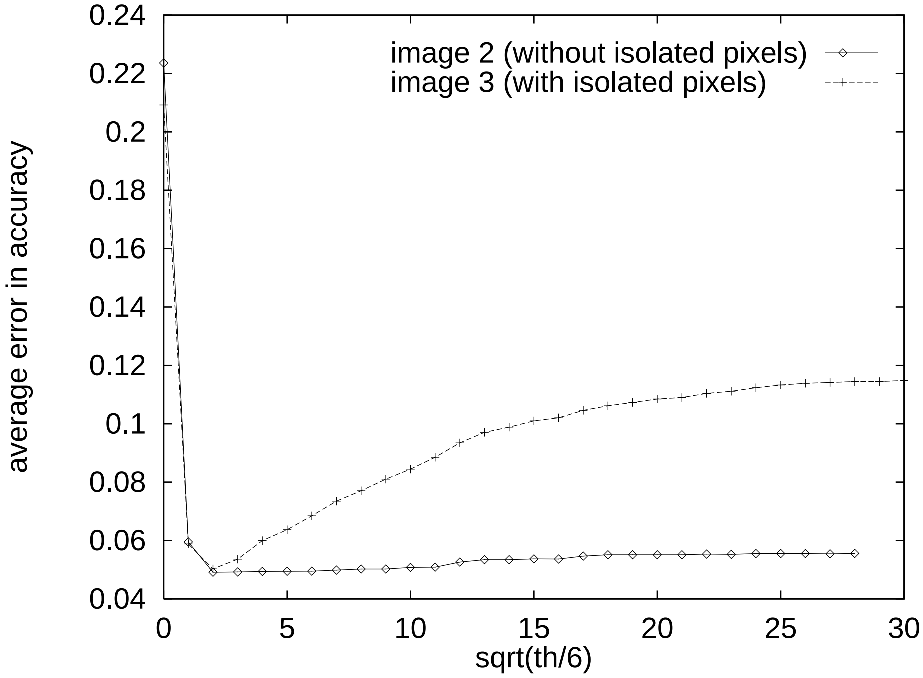


Figure 11: Decomposition of images with and without isolated pixels using different thresholds.

other hand the threshold was set too high, isolated pixels were not recognised but were incorrectly decomposed in stages 1 and 2. Furthermore, mixed pixels with contributing regions that could only be determined in the second stage were decomposed prematurely as well. In Section 2 is was derived that 2^2 times the number of spectral bands might be a good setting for the threshold, something which proved to be the case for the previous experiments. In the last experiment of this section, images 2 and 3 were repeatedly decomposed using different thresholds. The results shown in Figure 11 confirm that a threshold of 24.0 (Landsat-TM has 6 bands, $2^2 \cdot 6 = 24$) was an excellent setting. The effect of prematurely decomposing pixels whose regions could only be determined in stage 2 can be observed in the slightly rising curve of image 2. The much stronger effect of incorrectly decomposing isolated pixels is shown by the curve of image 3.

3.3 Comparison with other area estimation methods

Apart from data-driven decomposition, basically there exist three other methods for area estimation. The most simple approach is model-driven classification—sometimes called pixel-based classification—where each pixel is completely assigned to one class based on its spectral signature only. To reduce misclassifications due to significantly overlapping class distributions, data-driven (region-based) classification has been invented. After distinguishing the individual regions through segmentation, all pixels inside a region are assigned to the same class. The last approach is model-driven (pixel-based) decomposition, which divides each pixel between a pre-selected number of classes based solely on its spectral characteristics. To be able to compare the performance of the different methods, they were applied to images 2 and 3, which were described in Section 3.2.2, as well as two other images.

3.3.1 Measuring the performance

To measure how accurately a method could estimate the proportions of the components of a mixed pixel, the same error in accuracy e_p as given by Equation (5) was determined. The average error of all (mixed) pixels is denoted by $< e_p >$. But since the methods using classification could at best assign a pixel to the class having the largest proportion, the accuracy of the estimation per pixel was expected to be poor. However, the accuracy of the estimation per ground cover type could be better because a ground cover type of which the proportion was underestimated in one pixel could be compensated by an overestimation of its proportion in another pixel. Therefore, an error measure based on ground cover acreage was also determined. The true and estimated area covered by crop i are given by Equations (6) and (7), respectively:

$$A_i = \sum_p \mathrm{f}_{p,i}, \text{ and} \tag{6}$$

$$\hat{A}_i = \sum_p \hat{\mathrm{f}}_{p,i}. \tag{7}$$

Index p ranged over the mixed pixels only; pure pixels were ignored because their vast number would obscure the results of the mixed pixels. With these figures the following error measure was calculated:

$$e_A = \frac{1}{2} \sum_i \left| \hat{A}_i - A_i \right|. \tag{8}$$

In this equation, i ranged over all ground cover types to be distinguished in the image. Just like in Equation (5), the sum was divided by 2 because an underestimation of the acreage of one class was balanced by an overestimation of the acreage of (an)other class(es).

3.3.2 Results

The images that were used to carry out the experiments of this section were image 2 (edges of soil, no isolated pixels) and image 3 (edges of soil, 1.0% of the pixels isolated), see Section 3.2.2. In addition, images 2a (without isolated pixels) and 3a (with isolated pixels) were generated containing only 5 instead of 10 different ground cover types. These images were needed because model-driven decomposition could handle no more than five classes at the same time. This number was determined from own experiences but has also been confirmed by Mather [5] who stated that the number of Landsat-TM wavebands can be reduced to four without any significant reduction in separability. Since the number of classes model-driven decomposition could handle simultaneously was one more than the intrinsic dimensionality of the data, five classes was the maximum.

Both model-driven as well as data-driven classification used the maximum-likelihood approach to classify the mixed pixels. For data-driven classification the neighbouring regions were determined in the same way as for data-driven decomposition, see Section 2, using the same three stages. In the first stage the edge classes were also considered as a possible class of the mixed pixel, while in the last stage all classes were considered. To determine whether a pixel could be classified in stage 1 or 2, or only in stage 3, a reliability measure based on the Mahalanobis distance was used. In the following experiments the threshold was set at a value which resulted in the lowest average error in accuracy per pixel. For data-driven decomposition the threshold was always set at 24.0.

The results of the experiments on image 2a are presented in Table 1. Data-driven decomposition needed one iteration in which the remaining 16 pixels were processed. The average error in accuracy per pixel was 3.9%, and the summed error in acreage was 11.737. Model-driven decomposition achieved an average error in accuracy of 13.0% and a summed error in acreage of 203.462, which both were substantially higher. The classes that were wrongly considered during decomposition did not only increase the possibility of spectral confusion, usually they were assigned a small fraction as well. This combination of major and minor errors decreased the accuracy of the estimation. Data-driven classification gave an average error in accuracy of 32.7% and a summed error in acreage of 915.143. With the threshold set at 864.0, three iterations were needed in which subsequently 1, 1, and 0 pixels were classified, leaving 7 pixels isolated. Although it was already expected that the accuracy per pixel would be poor, the high error in acreage per crop was not anticipated.

	reference	data-driven decomposition	model-driven decomposition	data-driven classification	model-driven classification
$< e_p >$	–	0.039	0.130	0.327	0.406
e_A	–	11.737	203.462	915.143	1409.328
$\hat{A}_{soil}$	1853.786	1844.578	1699.259	1159	820
$\hat{A}_{barley}$	973.185	973.498	957.060	1293	826
$\hat{A}_{maize}$	947.660	943.910	913.562	725	718
$\hat{A}_{vineyards}$	772.629	782.739	775.979	1296	1467
$\hat{A}_{deciduous\ forest}$	383.346	383.253	582.170	455	1097

Table 1: Results of decomposing/classifying the mixed pixels of image 2a. Thresholds were set at 24.0 (data-driven decomposition) and 848.0 (data-driven classification).

The main reason for this surprising result was that certain combinations of ground cover types resembled a totally different class much more than the ground cover type of the pixel's main component. Especially the class vineyards seemed to be overestimated by this mechanism while the ground cover type soil—already underestimated because it usually was a small component—was underestimated. Model-driven classification achieved the worst results, giving an average error in accuracy of 40.6% and a summed error in acreage of 1409.328. Because no longer only classes of neighbouring regions were considered, spectral confusion was maximal. While the acreage of soil was further underestimated, the area covering deciduous forest was extremely overestimated.

Although model-driven decomposition could not be applied to image 2 because it contained too many classes, it still was interesting to compare the other three methods—see Table 2. The results of data-driven decomposition were already presented in Section 3.2.2 and included an average error in accuracy of 4.9%. The summed error in acreage was 26.558. Both error measures were a little bit higher than for image 2a, but they were still acceptable. The average error in accuracy for data-driven classification increased to 38.2%, which was to be expected since the doubled number of classes led to more possibilities for spectral confusion. The lower summed error in acreage of 734.554, however, was unexpected and could only be explained by estimations cancelling each other's error. With a threshold setting of 1734.0, only two iterations of stage 2 were needed in which respectively 12 and 5 pixels were processed without leaving any pixels isolated. Model-driven classification gave an average error in accuracy of 48.0% and a summed error in acreage of 1720.244. Both were substantially higher, as was to be expected.

Image 3a was similar to image 2a except for the 1.0% of its pixels that were converted into isolated mixed pixels. Another way of putting it is that image 3a was equal to image 3 with the exception that only 5 instead of 10 ground cover types were used. The results of the four methods are presented in Table 3. For the methods based on decomposition the introduction of isolated pixels was no problem judging by the minor increase of both errors. Data-driven decomposition used five iterations of stage 2 during which 36, 14, 4, 4, and 0 pixels were decomposed, marking 369 pixels as isolated. Data-driven classification had more difficulties, giving an average error in accuracy that was 2.5% higher than for image 2a. With the threshold set at 384.0, in 10 iterations 23, 8, 6, 5, 4, 2, 2, 2, 1, and 0 pixels were processed. A total of 492 pixels was marked isolated for stage 3. The problem was that data-driven classification had a hard time distinguishing ordinary mixed pixels from isolated mixed pixels. For ordinary mixed pixels, a higher threshold gave a more accurate estimation of the proportions per pixel, as was indicated by the results of image 2a. Accurate processing of isolated mixed pixels, on the other hand, demanded that the threshold was set at a lower value to prevent them from being processed before stage 3. The error in accuracy for model-driven classification improved from 40.6% for image 2a to 38.9%, but the error in acreage slightly increased to 1426.784. The improvement was caused by the isolated pixels, which always had a component with a proportion of at least 75.0%. Since isolated mixed pixels were much closer to the mean of a class than ordinary mixed pixels, classification was more accurate which led to a lower $< e_p >$.

	reference	data-driven decomposition	data-driven classification	model-driven classification
$< e_p >$	–	0.049	0.382	0.480
e_A	–	26.558	734.554	1720.244
$\hat{A}_{sand}$	173.666	175.072	148	112
$\hat{A}_{soil}$	1428.990	1420.572	937	350
$\hat{A}_{fresh\ water}$	251.130	252.523	115	76
$\hat{A}_{wheat}$	663.604	663.430	771	787
$\hat{A}_{barley}$	497.690	489.944	686	206
$\hat{A}_{maize}$	485.071	484.398	403	371
$\hat{A}_{rice}$	409.513	406.140	546	667
$\hat{A}_{vineyards}$	207.784	205.038	285	476
$\hat{A}_{grass}$	547.172	569.630	715	1142
$\hat{A}_{deciduous\ forest}$	265.986	261.258	322	741

Table 2: Decomposition and classification results of image 2 with 10 classes. Threshold settings were 24.0 for data-driven decomposition and 1734.0 for data-driven classification.

	reference	data-driven decomposition	model-driven decomposition	data-driven classification	model-driven classification
$< e_p >$	–	0.039	0.131	0.353	0.389
e_A	–	12.565	204.745	961.930	1426.784
$\hat{A}_{soil}$	1905.998	1898.427	1760.371	1179	872
$\hat{A}_{barley}$	1047.854	1046.726	1025.930	1279	889
$\hat{A}_{maize}$	1017.326	1013.830	978.753	781	782
$\hat{A}_{vineyards}$	847.853	859.004	847.915	1372	1540
$\hat{A}_{deciduous\ forest}$	480.757	478.974	684.060	686	1214

Table 3: Results of image 3a. Thresholds were 24.0 for decomposition and 384.0 for classification.

	reference	data-driven decomposition	data-driven classification	model-driven classification
$< e_p >$	–	0.050	0.411	0.464
e_A	–	22.916	738.703	1743.664
$\hat{A}_{sand}$	216.115	217.798	176	155
$\hat{A}_{soil}$	1446.312	1439.215	1027	378
$\hat{A}_{fresh\ water}$	290.206	291.353	101	87
$\hat{A}_{wheat}$	699.761	700.007	770	813
$\hat{A}_{barley}$	541.961	535.941	717	243
$\hat{A}_{maize}$	521.464	520.820	430	408
$\hat{A}_{rice}$	435.717	433.121	675	726
$\hat{A}_{vineyards}$	244.361	243.058	330	522
$\hat{A}_{grass}$	594.140	612.580	742	1181
$\hat{A}_{deciduous\ forest}$	309.752	303.097	329	784

Table 4: Results achieved for decomposition and classification of image 3. Threshold settings were 24.0 (data-driven decomposition) and 726.0 (data-driven classification).

The results achieved on image 3—10 classes and 416 isolated pixels—are presented in Table 4. For data-driven decomposition most results were already discussed in Section 3.2.2, except the summed error in acreage which equalled 22.916. Data-driven classification with a threshold setting of 726.0 resulted in five iterations of stage 2 as well, classifying 17, 12, 3, 2, and 0 pixels. The remaining 243 pixels were processed in stage 3. Compared to the results of image 3a, the same observation can be made as discussed when comparing the results of images 2 and 2a. With the number of ground cover types doubled, correct classification was more difficult due to the increased possibilities for spectral confusion. When comparing the performance with that of image 2, the same conclusions can be drawn as in the previous paragraph. Data-driven classification resulted in a higher error in accuracy because isolated and ordinary mixed pixels could not be properly distinguished, and model-driven classification showed a lower error in accuracy because isolated pixels could easily be classified, decreasing the average error.

4 Discussion

Data-driven decomposition can best be discussed in the light of the strong and weak points of other area estimation methods. The classification methods based on the maximum likelihood principle had great difficulties classifying mixed pixels. As expected, the estimation of the proportions was not very accurate since at best a pixel was assigned to the ground cover type having the largest proportion. As a result, all components with smaller proportions were underestimated which led to a high average error in accuracy $< e_p >$. If a ground cover type was overestimated for one pixel and underestimated for another, the area covered by that crop could still be accurately estimated. However, in many cases a mixed pixel did not resemble the ground cover type with the highest proportion, but a totally different class. Because of this spectral confusion, the acreage of some ground cover types was structurally overestimated, while the area of others was underestimated. As a consequence, the summed error in acreage e_A was high as well. Using spatial information, as data-driven classification did, lower errors were achieved. However, this approach was unable to clearly discriminate between ordinary and isolated mixed pixels based on the measure for the reliability of the estimation. Therefore, the optimal threshold setting was dependent on the image to be processed: the higher the percentage of isolated mixed pixels, the lower the threshold should be set to prevent isolated pixels from being classified prematurely. Model-driven classification did not suffer from this problem and displayed a better overall performance since isolated mixed pixels were easy to classify.

Compared to the methods based on classification, model-driven decomposition gave a much better estimation of both the proportions per pixel as well as the acreage per crop. The big disadvantage of this method was that only a very limited number of ground cover types could be handled simultaneously, about five for Landsat-TM images. Besides that, the need to consider all classes simultaneously increased the possibility of spectral confusion and diminished the accuracy of the estimation because small fractions were assigned to wrong classes. Data-driven decomposition, on the other hand, did not suffer from any of these weaknesses. Because for each pixel a different combination of classes was tried, the number of classes was unlimited. Furthermore, since only those classes were considered which probably contributed to a pixel, less spectral confusion could occur and leaking of small fractions was prevented. As a result, data-driven decomposition by far achieved the lowest average error in accuracy and summed error in acreage of all four methods. For all pixels, including those belonging to narrow structures, the correct regions could be found. The number of iterations needed for stage 2 was low, and few pixels were processed in this phase. Isolated mixed pixels could well be discriminated from ordinary mixed pixels. A threshold setting of 2^2 times the number of bands was a good choice and was not dependent on the image to be processed. The reasons why some mixed pixels were incorrectly decomposed were mostly implicit to the problem and not caused by the algorithm. The most important cause was the spectral confusion between the combination of soil and barley, and grass, which resulted in pixels with an overestimated proportion of grass. Another cause was the sometimes high deviation of a pixel's component from the mean vector of its class. As a result, a wrong combination of classes seemed a bit more probable, or the unreliability of the estimation was just above the threshold. A third reason was that several pixels consisted of three or four ground cover types, something the algorithm could not handle. Although this could be implemented, the increased number of classes would probably negatively influence the estimations of the other mixed pixels to such an extent that the overall accuracy would decrease. A last cause of error was the occurrence of isolated pixels on mixed pixel locations. Due to the many neighbouring regions, the possibilities for spectral confusion were numerous.

One suggestion for future work is to investigate the use of a different decomposition method. Instead of the shortcut method which was used for this work, the classical or alternative method could be used—see [4]. However, considering the high computational complexity of these methods and the good results already achieved, little gain is to be expected. A more interesting line of research is to use the output of a segmentation program instead of a perfect segmentation. The performance of data-driven decomposition on a segmented image with many small regions and pixels incorrectly marked as mixed pixels will probably be less good but in more accordance with reality. The most obvious extension of this work, however, is the application of data-driven decomposition to a real satellite image. Although a more realistic image could be generated using more edge classes and gradual instead of step-edges, some properties like non-linear mixing are very difficult to simulate. A big problem when using a real satellite image is the availability of supervised geographical data at subpixel level, which is needed for a quantitative evaluation of the algorithm. However, with quality measures specially designed to fit the properties of the supervised data, interesting experiments might be carried out.

References

[1] A.M. Cross, J.J. Settle, N.A. Drake, and R.T.M. Paivinen. Subpixel measurement of tropical forest cover using AVHRR data. *International Journal of Remote Sensing*, 12(5):1119–1129, 1991.

[2] H.M. Horwitz, R.F. Nalepka, P.D. Hyde, and J.P. Morgenstern. Estimating the proportions of objects within a single resolution element of a multispectral scanner. In *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, pages 1307–1320, 1971. Ann Arbor, Michigan.

[3] L.L.F. Janssen. *Methodology for updating terrain object data from remote sensing data.* PhD thesis, Wageningen Agricultural University, The Netherlands, 1994.

[4] M.S. klein Gebbinck and T.E. Schouten. Decomposition of mixed pixels. In J. Desachy, editor, *Image and Signal Processing for Remote Sensing II*, pages 104–115. SPIE 2579, 1995. Paris.

[5] P.M. Mather. Theoretical problems in image classification. In JA Clark and MD Stevens, editors, *Application of RS in Agriculture*, pages 127–135. Butterworths, London, 1990.

[6] R.P. Pech, A.W. Davies, R.R. Lamcroft, and R.D. Graetz. Calibration of LANDSAT data for sparsely vegetated semi-arid rangelands. *International Journal of Remote Sensing*, 7(12):1729–1750, 1986.

[7] N.A. Quarmby, J.R.G. Townshend, J.J. Settle, K.H. White, M. Milnes, T.L. Hindle, and N. Silleos. Linear mixture modelling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing*, 13(3):415–425, 1992.

[8] R.P.H.M. Schoenmakers. *Integrated methodology for segmentation of large optical satellite images in land applications of remote sensing.* PhD thesis, Department of Informatics, University of Nijmegen, The Netherlands, 1995.

[9] T.E. Schouten and M.S. klein Gebbinck. Quality measures for image segmentation using generated images. In J. Desachy, editor, *Image and Signal Processing for Remote Sensing II*, pages 411–422. SPIE 2579, 1995. Paris.

[10] J.J. Settle and N.A. Drake. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, 14(6):1159–1177, 1993.

A The REGIN3 format

A segmentation of an image corresponds to a partitioning of that image in regions. Therefore, the file to which a segmentation is saved is called a region information file. A first proposal in [8] to standardise these files has led to the REGINF format. However, this format could only deal with pure pixels, hence a new standard had to be adapted: the REGIN2 format. Although this format is sufficient to handle mixed pixels, during research it appeared that extension of the format with ancillary information was useful. As a result, the REGIN3 format was created.

The REGIN3 format incorporates the REGINF and REGIN2 standards. The header structure of all formats is equal to:

```
struct
{
    char hdword[6];               /* magic number */
    unsigned short byteorder;     /* sun or vax?  */
    unsigned short nrows;         /* number of rows */
    unsigned short ncols;         /* number of columns */
    unsigned short nbands;        /* number of bands */
    unsigned short accuracy;      /* size of mean */
    int nregs;                    /* number of regions */
    unsigned short spare[6];      /* future expansions */
} header;
```

For the REGIN3 format, the head word or magic number is REGIN3. The `spare[0]` field is used to contain the number of global regions—see Section 3.1—which are saved immediately after the header. Field `spare[1]` flags whether the last region contains pixels that are part of the background; on background pixels no information is available, but they do not have to be processed either.

The first part of the description of a region starts with its class using `sizeof(unsigned short)` bytes. The next `sizeof(int)` bytes represent the number of pure pixels belonging to the region. Hereafter, the region's mean spectral vector multiplied by 100 is given using `sizeof(unsigned short)` or `sizeof(int)` bytes per band, depending on the accuracy. The corresponding variance-covariance matrix is saved as well in a row-by-row fashion; each entry is multiplied by 100 and occupies `sizeof(int)` bytes. Next, the pure pixels of the region are listed by run-length encoding of the pixel sequences on a row:

```
struct
{
    unsigned short y;  /* start at row */
    unsigned short x;  /* start at column */
    unsigned short n;  /* length of segment */
} segment;
```

The end of this list is marked by `(unsigned short)-1`.

In the second part the mixed pixels of the region are described. Because no easy pattern is present in the positions of the mixed pixels, the description is kept as simple as possible. The result is a list of mixed pixels which holds for each pixel its position and the fraction belonging to the current region:

```
struct
{
    unsigned short y;  /* y-position */
    unsigned short x;  /* x-position */
    unsigned short f;  /* fraction of region */
} mixpix;
```

Field `f` contains the fraction multiplied by 1000, so fractions as small as 0.1% can be distinguished. The end of the list of mixed pixels is marked by `(unsigned short)-2`.

Similar to the earlier formats, the user is responsible for the integrity of the data. The property that a region is contiguous can easily be violated, as is done for global regions. A more serious violation is the creation of a mixed pixel whose fractions do not sum to 1.000. Whether or not the first case is an error is determined by the definition of a region, but the second case is impossible in our universe and should be avoided. Furthermore, there are a number of conventions the user should be aware of. One of them is that a class number of 0 or a number of pure pixels of -1 means that the field is undefined. Similar, a mean vector or variance-covariance matrix containing only zeroes indicates that these statistics have not been determined yet.