Decomposition of mixed pixels based on a physical linear mixture model

M. S. KLEIN GEBBINCK and T. E. SCHOUTEN
Computing Science Institute, Faculty of Mathematics and Informatics,
University of Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, the Netherlands

J. J. SETTLE
Environmental Systems Science Centre, University of Reading, Whiteknights,
Reading RG6 6AL, United Kingdom

Abstract

Classically, decomposition of mixed pixels is based on a linear mixture model that attributes deviations to random statistical noise. In order to utilise the existing correlations between neighbouring pixels, which cannot be described in this statistical model, we developed a model in which deviations are attributed to physical factors with a long length scale. Experiments on semi-real data showed that the proportions of mixed pixels can be estimated more accurately by solving the physical instead of the statistical linear mixture model. In addition, solving the physical model provides an estimate of the pixel's component spectra not supplied by the statistical model, which may be an important feature for some applications.

1 Introduction

In many application areas like environment quality assessment and agricultural resource management, remotely-sensed data is used to compose thematic maps. A concrete example of the latter area that we focus on is the verification of claims made on the agricultural subsidy system of the European Union. Since the height of the subsidy depends on the type and acreage of the crops grown, up-to-date and accurate ground cover maps are needed to check the claim made by a farmer. With satellite remote sensing, the first requirement is easily satisfied because the temporal resolution of most satellites is sufficiently high. Unfortunately, the spatial resolution of some widely used scanners is relatively low (Landsat-TM $30x30$ m^2, SPOT-XS $20x20$ m^2), which may violate the second requirement. Therefore, the methods that are used to convert remotely-sensed images to ground cover maps have to be highly accurate.

The usual method to produce a ground cover map from remotely-sensed data is pixel-based classification. This technique allocates each pixel to only one of a small number of preselected classes. However, many pixels, for instance those representing field boundaries, comprise more than one ground cover type. Classification of these so-called mixed pixels leads to an inaccurate ground cover map because at least a fraction of each pixel is allocated to an incorrect class. Furthermore, the spectral response of a mixed pixel may be so confusing that the pixel is allocated to an incorrect class altogether. Therefore, classification is not very suitable if an accurate ground cover map is to be constructed of a scene containing a considerable number of mixed pixels.

A better way to handle mixed pixels is to try to estimate how the area of each pixel is divided between the different ground cover types. After this unmixing step, each pixel can be allocated to several classes proportional to their contribution to the pixel's spectral response. Instead of a single ground cover map, this approach results in a series of maps giving the concentration per ground cover type. The classical method to estimate the ground cover proportions of a pixel is to solve the linear mixture model introduced by Horwitz *et al.* in 1971. In this model the deviations of pixels that do not fit the basic linear mixing equations are attributed to sensor noise having certain statistical properties. However, a close look at the pixels inside an agricultural field shows that their deviations are far less independent than expected. Although the results of the classical decomposition method usually are quite good, the inadequacy of the statistical linear mixture model may lead to sub-optimal estimates.

In this paper we propose to estimate the proportions of a mixed pixel based on a physical linear mixture model. In this model, deviations from the basic linear mixing equations are attributed to changes in physical factors such as humidity, soil type, elevation, etc. Since these changes usually have a long length scale relative to the dimensions of a pixel, correlations between the deviations of neighbouring pixels, which can be observed in most agricultural fields, can be explained. Because the physical model apparently is more realistic than the statistical model, an estimation of the ground cover proportions based on the physical model is probably more accurate as well. Furthermore, since solving the physical model entails estimating the actual components of a mixed pixel before their proportions can be calculated, additional information that may be useful for some applications is extracted. It is for these reasons that we think that decomposition of mixed pixels based on a physical model can be a viable alternative to the classical statistical unmixing approach.

The primary goal of our research was to determine whether the ground cover proportions of a mixed pixel could be estimated more accurately by solving the statistical model (see Section 2) or the physical model (see Section 3). For this purpose we conducted an experiment with semi-real data in order to make a quantitative comparison between the accuracy of both methods. Section 4 describes the set-up of this experiment and discusses the results; Section 5 presents the conclusions.

2 The statistical linear mixture model

The linear mixture model describes the relationship between the spectral response represented by a pixel and the proportions and characteristics of its components. This model assumes that each photon reaching the scanner interacted with only one ground cover type, which has proved to be an acceptable abstraction for remotely-sensed data. If we suppose that the scanner collects data on n spectral bands and that the number of ground cover types equals c, the basic linear mixture model is given by

$$\mathbf{x} = \mathbf{Mf}. \tag{1}$$

The $(nx1)$ vector $\mathbf{x}$ denotes the multi-spectral observation or pixel, while the proportions of the ground cover types are represented by the $(cx1)$ vector $\mathbf{f}$. The (nxc) matrix $\mathbf{M}$ contains the spectra of the c so-called endmembers that take part in the mixed pixel. An endmember spectrum is the response that would be received from a ground resolution element containing nothing but the cover type of interest. Such a spectrum can be derived from the image itself, for instance by averaging the pure pixels belonging to an agricultural field. With the mixing equations of Equation (1), the basic linear mixture model is largely defined.

In order to be physically sensible, the linear mixture model has to satisfy two constraints. The first constraint is that the pixel is well-defined by its components, whose proportions should therefore add to unity:

$$\sum_{i=1}^{c} f_i = 1. \tag{2}$$

This equation is called the sum-to-one constraint and can easily be satisfied. The second restriction is that each component of a mixed pixel cannot make a negative contribution, which is expressed by the positivity constraint:

$$f_i \geq 0 \quad \text{for} \quad i = 1, \ldots, c. \tag{3}$$

Because satisfaction of this last constraint is not straightforward, it is often ignored during the initial estimation of the proportions vector and satisfied in a separate post-processing step.

Classically, the basic linear mixture model is extended with a statistical noise term to facilitate inverse estimation of the proportions vector. If all the endmember spectra are linearly independent of each other, the system of equations defined by Equations (1) and (2) has infinitely many solutions for $\mathbf{f}$ if $c>n+1$, exactly one solution if $c=n+1$, and at most one solution if $c<n+1$. To provide a solution in the last case, Horwitz *et al.* (1971) added an error term to the basic mixing equations, which resulted in the statistical linear mixture model:

$$\mathbf{x} = \mathbf{M}\mathbf{f} + \mathbf{e}. \tag{4}$$

The (nx1) error vector $\mathbf{e}$, which can be regarded as random noise received by the sensor (Settle and Drake 1993), is assumed to have a multivariate normal distribution with a zero mean and a variance-covariance matrix $\mathbf{N_f}$ that depends on the pixel's composition. Horwitz *et al.* (1971) argued that each component makes a contribution to $\mathbf{N_f}$ that is proportional to the fraction of the pixel area it covers:

$$\mathbf{N_f} = \sum_{i=1}^{c} f_i \mathbf{N}_i. \tag{5}$$

In this equation, the (nxn) matrices $\mathbf{N}_i$ represent the statistical fluctuations that would be received for pure pixels of cover type i. One way to derive such a matrix from the image itself is to calculate the distribution of the pure pixels that belong to an agricultural field of the ground cover type of interest. Because of this error term, the statistical linear mixture model can always be solved.

In fact, for every pixel vector $\mathbf{x}$ there exist infinitely many solutions for the proportions vector $\mathbf{f}$ that all satisfy the statistical linear mixture model. The solution that is optimal according to the maximum likelihood approach minimises

$$\mathbf{e}^T \mathbf{N_f}^{-1} \mathbf{e} + \ln|\mathbf{N_f}| = (\mathbf{x} - \mathbf{M}\mathbf{f})^T \mathbf{N_f}^{-1} (\mathbf{x} - \mathbf{M}\mathbf{f}) + \ln|\mathbf{N_f}|, \tag{6}$$

where $\ln|\mathbf{N_f}|$ denotes the natural logarithm of the determinant of $\mathbf{N_f}$. Because $\mathbf{N_f}$ depends on the proportions vector that is to be estimated, a general analytic solution is hard to come by. However, the best estimate of $\mathbf{f}$ can always be approximated by testing all proportions vectors within a certain range. With this brute force method, a near-optimal solution of the statistical linear mixture model is found that satisfies both the sum-to-one and the positivity constraint.

If the variance-covariance matrix $\mathbf{N_f}$ would not depend on the proportions vector, a fast analytic solution for $\mathbf{f}$ could be derived. Therefore, the matrix is often approximated by

$$\mathbf{N} = \frac{1}{c} \sum_{i=1}^{c} \mathbf{N}_i, \tag{7}$$

which does not seriously affect the estimation as long as the dispersion of the endmembers is relatively large compared to the fluctuations represented by the $\mathbf{N}_i$. Now the minimum of Equation (6) can be calculated by setting its partial derivatives with respect to each f_i to zero, which gives the unconstrained estimator

$$\hat{\mathbf{f}}_0 = \mathbf{U}\mathbf{M}^T \mathbf{N}^{-1} \mathbf{x} \quad \text{with} \quad \mathbf{U} = (\mathbf{M}^T \mathbf{N}^{-1} \mathbf{M})^{-1}. \tag{8}$$

If the sum-to-one constraint is imposed, the estimator changes to

$$\hat{\mathbf{f}} = \hat{\mathbf{f}}_0 + \alpha \left(1 - \mathbf{1}^T \hat{\mathbf{f}}_0\right) \mathbf{U}\mathbf{1} \quad \text{with} \quad \alpha = \left(\mathbf{1}^T \mathbf{U}\mathbf{1}\right)^{-1} \tag{9}$$

and $\mathbf{1}$ being the (cx1) vector consisting of 1's. Satisfaction of the positivity constraint is much harder to enforce, and requires either an entirely different method of solution or a separate post-processing step. A simple post-processing approach is to set any negative fractions to zero and rescale the proportions vector (Settle and Drake 1993). Thus, by approximating $\mathbf{N_f}$ and correcting the resulting analytic solution afterwards, a fast shortcut for the classical estimator based on the statistical linear mixture model is derived.

3 The physical linear mixture model

The statistical model described in the previous section assumes that the deviations from the linear model $\mathbf{x}$=$\mathbf{Mf}$ are caused by independent random fluctuations. In practice these are likely to arise from variations in physical factors such as humidity, soil type, elevation, etc. This view is supported by the fact that often correlations are found between the neighbouring pixels in an agricultural field, which cannot be explained by the statistical noise term described in the previous section. Since the scale of the variations in the physical factors usually is larger than a single pixel, correlations between the spectra of adjacent pixels are to be expected. Therefore we propose the following form of mixture model, which recognises that component spectra vary from pixel to pixel:

$$\mathbf{x} = (\mathbf{M} + \mathbf{E})\mathbf{f}. \tag{10}$$

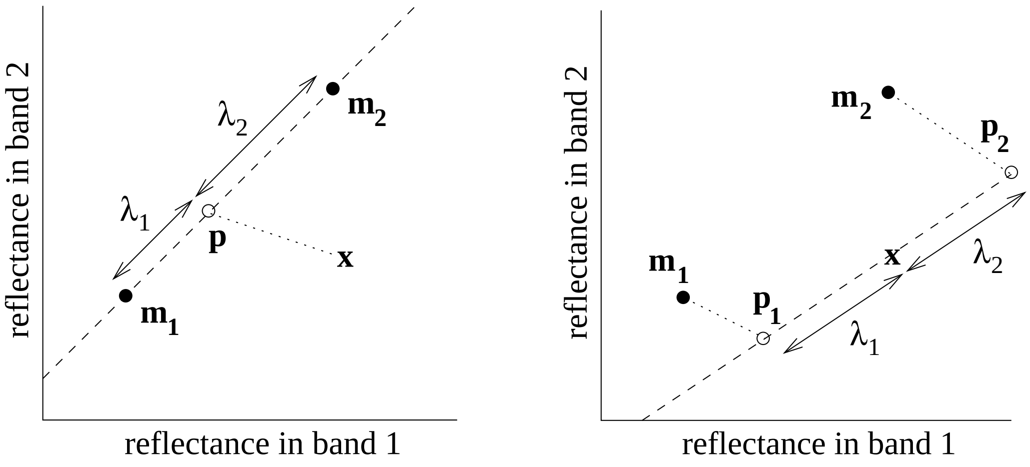
As before, $\mathbf{x}$ represents the n-dimensional pixel, $\mathbf{f}$ the $cx1$ proportions vector, and $\mathbf{M}$ the c endmember spectra averaged over the image. For a given pixel, the (nxc) matrix $\mathbf{E}$ contains the deviations from the mean endmember spectra that are due to physical variations, whose properties will be described further in this section. In addition, the sum-to-one and the positivity constraint, which are given by Equations (2) and (3) respectively, should also be satisfied. For this physical model, a solution can always be found because of the many degrees of freedom associated with $\mathbf{E}$.

For every pixel $\mathbf{x}$ there exist infinitely many combinations of an error matrix and a proportions vector that solve the physical linear mixture model. A criterion that could be used to determine which solution is optimal is that the deviations from the endmember spectra should be as small as possible. If we assume that the deviation from the i-th endmember, which is given by the i-th column $\mathbf{e}_i$ of $\mathbf{E}$, has a multivariate normal distribution with a zero mean and a variance-covariance matrix $\mathbf{N}_i$, the maximum likelihood approach favours the solution of Equation (10) that minimises

$$\sum_{i=1}^{c} \mathbf{e}_i^T \mathbf{N}_i^{-1} \mathbf{e}_i. \tag{11}$$

However, this strategy still suffers from the drawback observed earlier that correlations between the spectra of adjacent pixels are not modelled and, consequently, cannot be exploited. Therefore, a better criterion to determine the optimal solution is to minimise the difference between the error matrices of neighbouring pixels instead of the error matrices themselves. For this purpose, the deviation from the i-th endmember is split into a deterministic term $\mathbf{e}_{i,1}$ that is typical for the local environment of pixel $\mathbf{x}$, and a statistical term $\mathbf{e}_{i,2}$ that should be minimised. When dealing with agricultural scenes, we can easily determine the environmental component by looking at the pure pixels that surround the mixed pixel under observation—see Section 4.1 for an example. The noise term $\mathbf{e}_{i,2}$ is once again assumed to have a multivariate normal distribution with a zero mean and a variance-covariance matrix $\mathbf{N}_i$. It must be noted, however, that the definition of this matrix differs from the one used before: while $\mathbf{N}_i$ was first determined by calculating the distribution of the pure pixels belonging to an agricultural field, it is now estimated by calculating the distribution of the difference between neighbouring pixels of the same field. By substituting $\mathbf{m}_i$ with $\mathbf{m}'_i$=$\mathbf{m}_i$+$\mathbf{e}_{i,1}$—so $\mathbf{m}'_i$ depends on the pixel under observation but is still deterministic—prior to solving the physical model, we can unify the two approaches.

Both criteria presented in the last paragraph can be used to select a solution from the infinity of possibilities, but they do not suggest how this solution can be found. Among other methods, Klein Gebbinck (forthcoming) discusses a brute force approach that is based on the theory of convex geometry. Although this method is not described here, a look at the n-dimensional feature space can still be informative. According to the basic linear mixture model, pixels that are composed of c linearly independent endmembers lie in the c-dimensional linear subspace spanned by the c vectors $\mathbf{m}_i$ that represent the endmember spectra. If the proportions of the mixed pixel sum to one, this subspace is restricted to the (c-1)-dimensional subspace containing the $\mathbf{m}_i$ but not (necessarily) the origin of the coordinate system; the latter subspace is further limited to the



(a) The statistical model. (b) The physical model.

Figure 1: Maximum likelihood solution of two linear mixture models. In each two-dimensional feature space, linear combinations of endmembers $\mathbf{m}_1$ and $\mathbf{m}_2$ satisfying the sum-to-one constraint occupy the dashed line; those that satisfy the positivity constraint as well lie on $\mathbf{m}_1\mathbf{m}_2$. The corresponding fractions vector $\mathbf{f}$ equals $(\lambda_2/(\lambda_1+\lambda_2), \lambda_1/(\lambda_1+\lambda_2)^T$. (a) Solving the statistical model equals to finding the point $\mathbf{p}$ that has the smallest Mahalanobis distance to $\mathbf{x}$. (b) To solve the physical model, one point for each endmember has to be found such that the line through these points contains $\mathbf{x}$. The optimal solution is defined by the line that leads to the smallest sum of Mahalanobis distances between $\mathbf{m}_i$ and $\mathbf{p}_i$.

simplex defined by the $\mathbf{m}_i$ if all fractions are greater than (or equal to) zero. The statistical linear mixture model attributes the occurrence of pixels outside this simplex to statistical noise, and solving this model by minimising Equation (6) under the above constraints corresponds to finding a point $\mathbf{p}$ on the simplex that has a minimal Mahalanobis distance (M-distance) to $\mathbf{x}$—see Figure 1(a). The physical linear mixture model, on the other hand, ascribes the occurrence of pixels outside the simplex to local variations in the endmember spectra; solving this model, therefore, requires that for each endmember a point $\mathbf{p}_i$ is found such that these vectors span a c-dimensional linear subspace containing $\mathbf{x}$. If the sum-to-one is to be satisfied, $\mathbf{x}$ must also be contained in the (c-1)-dimensional subspace through the $\mathbf{p}_i$ (but not necessarily through the origin), and satisfaction of the positivity constraint demands that $\mathbf{x}$ is contained in the simplex made up of the $\mathbf{p}_i$. Even with these constraints, the number of possible solutions still is infinite. The solution that is favoured by Equation (11) has a minimal sum of M-distances between the $\mathbf{p}_i$'s and their corresponding $\mathbf{m}_i$'s (or $\mathbf{m}_i'$'s)—see Figure 1(b). Based on these insights, the brute force approach mentioned earlier was developed.

As could be done for the statistical model, the optimal solution of the physical model can be approximated by a closed formula as well. Once again it is required that all variance-covariance matrices $\mathbf{N}_i$ are identical, which can be accomplished by taking the mean matrix $\mathbf{N} = 1/c\sum_{i=1}^{c}\mathbf{N}_i$. Following a similar reasoning as Horwitz *et al.* (1971) did for the statistical model, one can make plausible that the resulting loss of accuracy probably is not very large as long as the $\mathbf{N}_i$ are small relative to the dispersion of the endmembers. Under the assumption of equal variance-covariance matrices, the expression to be minimised—$\sum_{i=1}^{c}\mathbf{e}_i^T\mathbf{N}^{-1}\mathbf{e}_i$, subject to the conditions $\mathbf{x} = (\mathbf{M}+\mathbf{E})\mathbf{f}$ and $\sum_{i=1}^{c} f_i = 1$—can be rewritten to the following Lagrangian:

$$\mathrm{Trace}(\mathbf{E}^T\mathbf{N}^{-1}\mathbf{E}) + 2\boldsymbol{\alpha}^T(\mathbf{x} - (\mathbf{M}+\mathbf{E})\mathbf{f}) + 2\beta(\mathbf{1}^T\mathbf{f} - 1). \tag{12}$$

In this equation, Trace($\mathbf{A}$) represents the sum of the diagonal entries of matrix $\mathbf{A}$, $\boldsymbol{\alpha}$ stands for an

$nx1$ vector of Lagrangian constants, and β denotes another Lagrangian constant. The minimum of Equation (12) is found by differentiation and setting the result to zero. Differentiating with respect to $\mathbf{E}$ gives

$$2\mathbf{N}^{-1}\mathbf{E} - 2\boldsymbol{\alpha}\mathbf{f}^T = \mathbf{0}, \quad \text{or, equivalently,} \quad \mathbf{E} = \mathbf{N}\boldsymbol{\alpha}\mathbf{f}^T, \tag{13}$$

where the (nxc) matrix $\mathbf{0}$ is filled with zeroes; differentiating with respect to $\mathbf{f}$ results in

$$(\mathbf{M}+\mathbf{E})^T\boldsymbol{\alpha} = \beta\mathbf{1}, \tag{14}$$

where $\mathbf{1}$ is a $(cx1)$ vector of 1's only. From this last equation, a relation between β and the other Lagrangian constants can be derived:

$$\begin{array}{lll} (\mathbf{M}+\mathbf{E})^T\boldsymbol{\alpha} = \mathbf{1}\beta & \Longrightarrow & \{*\mathbf{f}^T\} \\ \mathbf{f}^T(\mathbf{M}+\mathbf{E})^T\boldsymbol{\alpha} = \mathbf{f}^T\mathbf{1}\beta & \Longrightarrow & \{(10),\,(2)\} \\ \mathbf{x}^T\boldsymbol{\alpha} = \beta. & & \end{array} \tag{15}$$

Based on Equations (13) to (15), a fast method for the simultaneous retrieval of $\mathbf{E}$ and $\mathbf{f}$ can be developed.

The general principle followed to derive an estimate of the proportions was to combine the above equations in trying to eliminate the Lagrangian constants until an expression containing only the unknown fractions vector $\mathbf{f}$ was left. Starting with Equation (14), one can make the following derivation:

$$\begin{array}{llll} (\mathbf{M}+\mathbf{E})^T\boldsymbol{\alpha} & = \mathbf{1}\beta & \Longrightarrow & \{(13),\,(15)\} \\ (\mathbf{M}+\mathbf{N}\boldsymbol{\alpha}\mathbf{f}^T)^T\boldsymbol{\alpha} & = \mathbf{1}(\mathbf{x}^T\boldsymbol{\alpha}) & \Longrightarrow & \{\text{unfold, } \mathbf{N}^T\text{=}\mathbf{N}\} \\ \mathbf{M}^T\boldsymbol{\alpha} + (\mathbf{f}\boldsymbol{\alpha}^T\mathbf{N})\boldsymbol{\alpha} & = \mathbf{1}\mathbf{x}^T\boldsymbol{\alpha} & \Longrightarrow & \{\text{rearrange}\} \\ \mathbf{M}^T\boldsymbol{\alpha} - \mathbf{1}\mathbf{x}^T\boldsymbol{\alpha} & = -\mathbf{f}\boldsymbol{\alpha}^T\mathbf{N}\boldsymbol{\alpha} & \Longrightarrow & \{\text{fold}\} \\ (\mathbf{M}-\mathbf{x}\mathbf{1}^T)^T\boldsymbol{\alpha} & = -\mathbf{f}(\boldsymbol{\alpha}^T\mathbf{N}\boldsymbol{\alpha}). & & \end{array} \tag{16}$$

The mixing equations described by Equation (10) can be transformed as follows:

$$\begin{array}{llll} (\mathbf{M}+\mathbf{E})\mathbf{f} & = \mathbf{x} & \Longrightarrow & \{(13)\} \\ (\mathbf{M}+\mathbf{N}\boldsymbol{\alpha}\mathbf{f}^T)\mathbf{f} & = \mathbf{x} & \Longrightarrow & \{\text{unfold, rearrange}\} \\ \mathbf{M}\mathbf{f} - \mathbf{x} & = -(\mathbf{N}\boldsymbol{\alpha}\mathbf{f}^T)\mathbf{f} & \Longrightarrow & \{(2)\} \\ \mathbf{M}\mathbf{f} - \mathbf{x}\mathbf{1}^T\mathbf{f} & = -\mathbf{N}\boldsymbol{\alpha}\mathbf{f}^T\mathbf{f} & \Longrightarrow & \{\text{fold}\} \\ (\mathbf{M}-\mathbf{x}\mathbf{1}^T)\mathbf{f} & = -\mathbf{N}\boldsymbol{\alpha}(\mathbf{f}^T\mathbf{f}). & & \end{array} \tag{17}$$

Finally, we multiply the above result with $(\mathbf{M}-\mathbf{x}\mathbf{1}^T)^T\mathbf{N}^{-1}$ to combine it with Equation (16):

$$\begin{array}{rcl} (\mathbf{M}-\mathbf{x}\mathbf{1}^T)^T\mathbf{N}^{-1}(\mathbf{M}-\mathbf{x}\mathbf{1}^T)\mathbf{f} & = & (\mathbf{M}-\mathbf{x}\mathbf{1}^T)^T\mathbf{N}^{-1}(-\mathbf{N}\boldsymbol{\alpha}(\mathbf{f}^T\mathbf{f})) \\ & = & -(\mathbf{M}-\mathbf{x}\mathbf{1}^T)^T\boldsymbol{\alpha}(\mathbf{f}^T\mathbf{f}) \\ & = & \mathbf{f}(\boldsymbol{\alpha}^T\mathbf{N}\boldsymbol{\alpha})(\mathbf{f}^T\mathbf{f}). \end{array} \tag{18}$$

In this last expression the well-known system $\mathbf{A}\mathbf{f} = \lambda\mathbf{f}$ can be recognised, given that $\mathbf{A} = (\mathbf{M}-\mathbf{x}\mathbf{1}^T)^T\mathbf{N}^{-1}(\mathbf{M}-\mathbf{x}\mathbf{1}^T)$ and $\lambda = (\boldsymbol{\alpha}^T\mathbf{N}\boldsymbol{\alpha})(\mathbf{f}^T\mathbf{f})$. The $\mathbf{f}$ that solve this system are the eigenvectors of matrix $\mathbf{A}$, which can be calculated in a straightforward fashion (see Strang 1980). From these alternatives, the $\mathbf{f}$ that minimises $\mathrm{Trace}(\mathbf{E}^T\mathbf{N}^{-1}\mathbf{E})$ is selected; since $\mathrm{Trace}(\mathbf{E}^T\mathbf{N}^{-1}\mathbf{E}) = \mathrm{Trace}(\mathbf{f}\boldsymbol{\alpha}^T\mathbf{N}\mathbf{N}^{-1}\mathbf{N}\boldsymbol{\alpha}\mathbf{f}^T) = (\boldsymbol{\alpha}^T\mathbf{N}\boldsymbol{\alpha})(\mathbf{f}^T\mathbf{f})$, the eigenvector corresponding to the smallest eigenvalue λ is considered to be the optimal solution. Subsequently, the error matrix $\mathbf{E}$ is calculated by combining Equations (13) and (17):

$$\mathbf{E} = -\frac{(\mathbf{M}-\mathbf{x}\mathbf{1}^T)\mathbf{f}\mathbf{f}^T}{\mathbf{f}^T\mathbf{f}}. \tag{19}$$

Note—expand $(\mathbf{M}+\mathbf{E})\mathbf{f}$—that this expression requires that the sum-to-one constraint is satisfied. Since some standard packages supply eigenvectors that are normalised according to Minkowski's L2-norm, the proportions vector may have to be rescaled. In order to satisfy the positivity constraint, some post-processing is necessary as well. One possible method already suggested in Section 2 is to set negative proportions to zero and rescale the fractions vector. Although this approach introduces some further inaccuracy, its computational costs are very low. As a result, the overall method to approximate the optimal solution of the physical linear mixture model is relatively fast.

4 Experiments with semi-real data

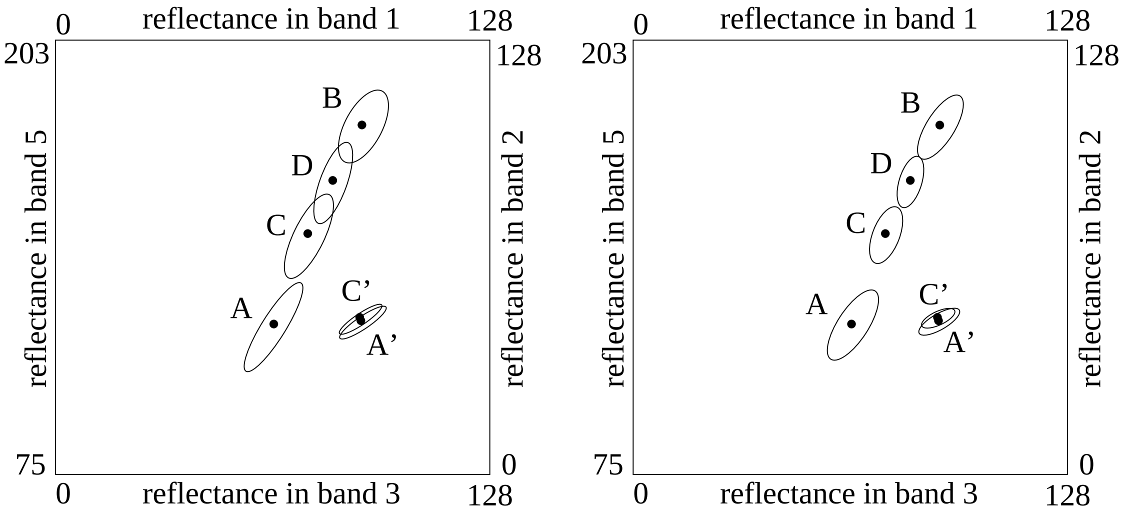
In order to determine the suitability of the statistical and the physical linear mixture model for mixed pixel decomposition, several experiments were carried out. In these experiments, the crop area estimation accuracies of six methods were evaluated using artificial images, which was necessary to make a quantitative comparison possible. Section 4.1 describes the set-up of the experiments, while Section 4.2 discusses the results.

4.1 *The experimental set-up*

Each experiment consisted of three steps. In the first step, an image of an agricultural scene was generated with fields having certain statistical properties. The second step comprised the processing of the image by two methods based on the statistical model and two methods per minimisation criterion based on the physical model. In the third and last step, the results of the various methods were quantified using several accuracy measures.

To simulate remotely-sensed images, a modified version of the method described by Schouten and Klein Gebbinck (1995) was used. The basis of the method was a high resolution image containing closed polygons that represented field boundaries. After assigning a class to each field such that neighbouring fields had different classes, the proportion vectors of the target image pixels were determined by calculating the composition of $4x4$ pixel blocks of the original image. The pixel values of the resulting low resolution image were obtained by overlaying the image with class templates multiplied by the corresponding proportion vectors: each pure pixel got a value equal to the template pixel to which it had been mapped, while the value of a mixed pixel was set equal to the summed contributions of several class templates weighed by the relative number of subpixels per class in the pixel's $4x4$ block. Each of these templates was derived from a real Landsat-TM image of Portugal by clipping a square block of pixels out of the middle of a large agricultural field belonging to a particular class. Since these clippings comprised only $15x15$ pixels, several copies had to be mirrored and stitched together to get sufficiently large templates in which the correlations between adjacent pixels were preserved. This way, we derived four templates labelled A–D, whose statistical distributions are shown in Figure 2. All artificial images were made using only two templates at the same time, while the number of bands was reduced from six to two or three. In most cases bands 3 and 5 (plus band 7 for n=3) were selected, because they provided the best discrimination between vegetation types (Sabins 1987). The only exceptions were two images created with templates A$'$ and C$'$, which were derived from the six band templates A and C by selecting bands 1 and 2 (plus band 4 in case n=3). With this method, 12 images were generated that each contained 297 mixed pixels to be decomposed.

Each simulated image was processed by six different decomposition methods. Two of these methods were based on the statistical model, while the other four methods were based on the physical model, minimising either the magnitude of the error matrix $\mathbf{E}$—the basic approach—or the differences between the error matrices of neighbouring pixels—the environmental approach. Besides this tripartite division, the methods can also be grouped three by three according to the method of solution applied: brute force or mathematical analysis. Whereas the brute force approaches approximate the optimal solution by searching the entire solution space, the analytic methods approximate the variance-covariance matrices of all classes by their average such that the



(a) Distributions for the basic physical approach.

(b) Distributions for the environmental approach.

Figure 2: Two-dimensional feature space showing six class distributions. A–D are extracted from bands 3 and 5, while A$'$ and C$'$ are extracted from bands 1 and 2. (a) Class mean $\mathbf{m}_i$ surrounded by the ellipse $\mathbf{x}^T\mathbf{N}_i^{-1}\mathbf{x} = 4$, where $\mathbf{N}_i$ represents the variation in the pixels themselves. (b) Similar set-up, except that the ellipses now represent the variation in the differences between neighbouring pixels.

optimal solution can be expressed as a closed formula. The endmember distributions used for the statistical and the basic physical approaches were estimated by calculating the mean vector and variance-covariance matrix of the pixels of each class template. For the environmental approach, the endmember distributions had to be calculated totally differently, which will be described in the remainder of this paragraph. Since the simulated images represented agricultural scenes, each mixed pixel was surrounded by pure pixels of the fields it was composed of. The spectral response of such a pure pixel q is approximately equal to one of the local endmember spectra (say i) of the neighbouring mixed pixel p, albeit that it also includes a small statistical term $\mathbf{e}_{i,2,q}$. The only consequence of this minor inconvenience is that, after substituting $\mathbf{x}_q$ for $\mathbf{m}_i + \mathbf{e}_{i,1,p}$, the term $\mathbf{e}_{i,2,p} - \mathbf{e}_{i,2,q}$ instead of $\mathbf{e}_{i,2,p}$ has to be determined in order to solve the model. The distribution of this subtraction sum was estimated based on the differences calculated between all pairs of neighbouring pixels in the corresponding template. As expected, the mean vector of the subtraction sum turned out to be $\mathbf{0}$ for all templates. The only issue that remained was which neighbouring pixels were to be used to approximate the local endmember spectra. This problem was solved by trying all combinations of neighbouring pure pixels—one pixel per class—and selecting the combination that minimised Equation (11) (note that $\mathbf{e}_i$ now represents the subtraction sum). Thus, by defining different endmember distributions for all mixed pixels, the environmental approach was unified with the basic physical approach.

The performance of each method was expressed by three accuracy measures. To measure how accurately the proportions of a mixed pixel could be estimated, the pixel's true proportions f_i were compared with their estimates $\hat{f}_i$:

$$\varepsilon_f = \left\langle \frac{1}{2} \sum_{i=1}^{c} \left| \hat{f}_i - f_i \right| \right\rangle . \tag{20}$$

In this equation, $|\ |$ stands for taking the absolute value, and $< >$ denotes that the enclosed expression is averaged over all mixed pixels. The scalar $\frac{1}{2}$ was introduced because overestimation of one class automatically leads to underestimation of the other; due to this scalar the value of ε_f was restricted to the interval [0,1]. The second accuracy measure expressed how large the statistical deviations had to be in order to reconstruct the mixed pixel and is defined by

$$\varepsilon_x = \left\langle \frac{1}{n} \mathbf{e}^T \mathbf{N_f}^{-1} \mathbf{e} \right\rangle \quad \text{or} \quad \varepsilon_x = \left\langle \frac{1}{nc} \sum_{i=1}^{c} \mathbf{e}_i^T \mathbf{N}_i^{-1} \mathbf{e}_i \right\rangle , \tag{21}$$

depending on whether the statistical or one of the physical linear mixture models was assumed. Although the magnitude of the deviations is of no direct importance, it may be an indication of the reliability of the reconstruction. Obviously, ε_x cannot be used to compare the methods based on different mixture models. The third and last measure was determined for the methods based on the physical models only and indicated how well the component spectra of a mixed pixel could be estimated by calculating the difference between the true value of the i-th component $\mathbf{p}_i$ and its estimate $\hat{\mathbf{p}}_i$:

$$\varepsilon_{p_i} = \left\langle \frac{1}{n} (\hat{\mathbf{p}}_i - \mathbf{p}_i)^T \mathbf{N}_i^{-1} (\hat{\mathbf{p}}_i - \mathbf{p}_i) \right\rangle . \tag{22}$$

Note that in this equation as well as in Equation (21) different $\mathbf{N}_i$ were used for the basic physical approach and the environmental approach. With these three measures, a quantitative comparison of the performance of the different methods was made in a straightforward way.

4.2 *Results and discussion*

Two important factors that influence the accuracy of a decomposition method are the relative positions of the endmembers in the feature space and the statistical distribution of the error components. Figure 2(a) shows these class characteristics as computed for both the statistical and the basic physical approach, while Figure 2(b) displays the characteristics as determined for the environmental approach. In both figures it can be seen that the dispersion of the endmembers A–D is relatively large, whereas the distance between the endmembers A′ and C′ is relatively small. The latter two classes were defined by selecting bands 1 and 2 instead of the usual bands 3 and 5, which was done to make in order to simulate an image containing mixed pixels that were difficult to decompose. Addition of a third band—band 7 for A–D, band 4 for A′ and C′—to simulate images of dimension three did not change the relative positions of the endmembers to any great degree. A comparison of Figures 2(a) and 2(b) clearly reveals that neighbouring pixels are correlated. If two adjacent pixels p and q had been independent, then the environmental error component they have in common would be zero, and the distribution of the statistical error components $\mathbf{e}_{i,2,p}$ and $\mathbf{e}_{i,2,q}$ would be equal to that of $\mathbf{e}_i$. As a result, the variance-covariance of the subtraction sum $\mathbf{e}_{i,2,p} - \mathbf{e}_{i,2,q}$ calculated for the environmental approach would be twice as large as that of the noise vector $\mathbf{e}_i$ used by both the statistical and the basic physical approach, which would translate to ellipses that were $\sqrt{2}$ times as large. Although this effect indeed was observed if pixels p and q were chosen at random, the fact that the ellipses in Figure 2(b) are smaller than those in Figure 2(a) demonstrates the correlation in case p and q are neighbours. Based on this collection of endmembers, it was possible to evaluate the performance of the different decomposition methods under a variety of conditions.

Looking at the results of the experiments presented in Table 1, we can make several observations. In general, the accuracy of the proportion estimates decreased as the distance between the endmembers got smaller, which was to be expected because the influence of the statistical deviations became more important. Another anticipated result is that the accuracy of the estimates improved as the number of bands increased. An explanation for this phenomenon is that the effect of the statistical deviation in one band was compensated for by opposing deviations in the other bands. As a consequence, however, an exact reconstruction of the spectral response of pixel $\mathbf{x}$ became more difficult—ε_x is higher—because the deviations could not all be fitted at the same time. The third and most important observation is that there is a clear order in which the

mix	method	n=2				n=3			
		ε_f(%)	ε_x	ε_{p1}	ε_{p2}	ε_f(%)	ε_x	ε_{p1}	ε_{p2}
	Stat. BF	5.8	0.34	–	–	5.7	0.46	–	–
	Stat. MA	6.0	0.35	–	–	5.8	0.46	–	–
	Phys.BF	5.9	0.27	0.78	0.74	5.7	0.36	0.71	0.66
AB	Phys.MA	6.0	0.29	0.85	0.77	5.8	0.37	0.75	0.67
	Env. BF	5.5	0.06	0.71	0.84	5.0	0.12	0.66	0.71
			(0.07)	(0.73)	(0.61)		(0.16)	(0.84)	(0.65)
	Env. MA	5.4	0.05	0.75	0.82	5.0	0.11	0.68	0.69
			(0.06)	(0.75)	(0.61)		(0.14)	(0.86)	(0.63)
	Stat. BF	13.8	0.32	–	–	11.0	0.45	–	–
	Stat. MA	13.8	0.32	–	–	10.9	0.46	–	–
	Phys.BF	14.7	0.23	0.77	0.79	12.3	0.34	0.69	0.69
AC	Phys.MA	14.5	0.24	0.78	0.77	12.0	0.35	0.70	0.66
	Env. BF	9.3	0.07	0.63	0.75	9.2	0.12	0.63	0.71
			(0.08)	(0.66)	(0.44)		(0.14)	(0.84)	(0.52)
	Env. BF	9.3	0.08	0.65	0.77	9.4	0.13	0.64	0.75
			(0.08)	(0.69)	(0.44)		(0.14)	(0.87)	(0.54)
	Stat. BF	10.4	0.36	–	–	10.3	0.43	–	–
	Stat. MA	10.4	0.36	–	–	10.3	0.43	–	–
	Phys.BF	10.7	0.27	0.64	0.78	10.8	0.32	0.61	0.70
BC	Phys.MA	10.6	0.27	0.66	0.77	10.8	0.32	0.62	0.70
	Env. BF	8.4	0.05	0.75	0.86	8.2	0.11	0.64	0.74
			(0.03)	(0.56)	(0.50)		(0.11)	(0.57)	(0.52)
	Env. BF	8.4	0.05	0.75	0.85	8.0	0.10	0.62	0.73
			(0.03)	(0.56)	(0.49)		(0.10)	(0.56)	(0.51)
	Stat. BF	19.3	0.31	–	–	19.3	0.43	–	–
	Stat. MA	19.4	0.30	–	–	19.6	0.43	–	–
	Phys.BF	20.5	0.20	0.83	0.96	23.9	0.27	0.80	0.89
BD	Phys.MA	20.5	0.21	0.83	0.98	24.0	0.30	0.85	0.92
	Env. BF	13.3	0.05	0.62	0.68	12.7	0.10	0.64	0.65
			(0.03)	(0.46)	(0.37)		(0.08)	(0.61)	(0.46)
	Env. BF	13.2	0.05	0.62	0.71	12.9	0.10	0.63	0.68
			(0.03)	(0.46)	(0.38)		(0.09)	(0.60)	(0.48)
	Stat. BF	23.4	0.33	–	–	14.9	0.47	–	–
	Stat. MA	23.4	0.33	–	–	15.0	0.47	–	–
	Phys.BF	25.9	0.21	0.84	0.89	17.0	0.33	0.75	0.69
CD	Phys.MA	26.2	0.21	0.85	0.91	18.1	0.36	0.77	0.77
	Env. BF	13.8	0.05	0.78	0.71	11.5	0.10	0.71	0.61
			(0.03)	(0.44)	(0.36)		(0.08)	(0.50)	(0.41)
	Env. BF	13.9	0.05	0.81	0.73	11.7	0.10	0.73	0.60
			(0.03)	(0.46)	(0.37)		(0.09)	(0.50)	(0.41)

continued on next page

continued from previous page

		$\varepsilon_f(\%)$	ε_x	ε_{p1}	ε_{p2}	$\varepsilon_f(\%)$	ε_x	ε_{p1}	ε_{p2}
	Stat. BF	39.4	0.49	–	–	36.4	0.53	–	–
	Stat. MA	39.8	0.49	–	–	39.5	0.51	–	–
	Phys.BF	45.0	0.25	0.70	0.86	42.5	0.25	0.57	0.89
A′C′	Phys.MA	45.6	0.25	0.73	0.86	47.0	0.28	0.70	0.92
	Env. BF	27.9	0.06	0.67	0.73	23.7	0.10	0.50	0.73
			(0.08)	(0.77)	(0.75)		(0.12)	(0.58)	(0.73)
	Env. BF	28.2	0.06	0.66	0.73	23.7	0.10	0.51	0.73
			(0.08)	(0.76)	(0.75)		(0.12)	(0.60)	(0.73)

Table 1: Performance of six decomposition methods for different mixtures using two or three spectral bands. Methods: Stat=statistical, Phys=basic physical, Env=environmental, BF=brute force, MA=mathematical analysis. Step sizes of the brute force methods were 0.01 (statistical) and 0.01π (other). Values in parentheses are calculated using the $\mathbf{N}_i$ of the basic physical approach.

methods were best able to estimate the proportions of a mixed pixel. In all experiments, the environmental approaches to solve the physical model were the most accurate, followed at some distance by the methods that assume the statistical model, which on their turn performed slightly better than the basic approaches towards the physical model. This order is apparent especially in the more difficult cases, like the three-dimensional A′C′ experiment for instance. Methods based on the same paradigm performed approximately equally well, which is explained by the fact that only simple mixtures of two classes having similar variance-covariance matrices were considered. As a result, the average matrix $\mathbf{N}$ was nearly identical to the $\mathbf{N}_i$ (and thus also to $\mathbf{N}(\mathbf{f})$), which led to similar proportions, similar reconstruction errors ε_x, and similar component estimates.

A further comparison of the performance of the methods based on the physical model once again demonstrates that neighbouring pixels are correlated. The big difference in values of ε_x between the basic physical and the environmental approach indicates that in the latter model smaller displacements were needed to reconstruct pixel $\mathbf{x}$; apparently, the true spectra of a pixel's components were approximated well by the spectra of the components of neighbouring pixels. This hypothesis is confirmed by the fact that the accuracy measure ε_p on average shows lower values for the environmental approach than for the basic physical approach, even when the values for the environmental approach are calculated using the variance-covariance matrices of the basic physical approach (see Tables 1, values given in parentheses). Although the accuracy of the estimated components seems to be a bit disappointing in an absolute sense, it should be noted that the estimates were considerably more accurate than samples drawn at random: the expected value of ε_{p_i} for a random set of samples drawn according to class i's distribution is equal to 1.0.

As can be seen in Table 2, the time needed to process one image differed significantly per method. For each of the three paradigms, the analytic solution was calculated much faster than the brute force approximation, which was not surprising since the computational complexity of the analytic methods was polynomial in n and c, while that of the brute force methods was exponential. When the number of spectral bands increased from two to three, the analytic methods hardly needed more processing time. The processing time of the brute force methods based on the physical model, on the other hand, increased by a factor of about 100, which matched the sampling rate chosen. The brute force approximation of the statistical model did not require that much extra time, since its complexity was exponential in c only. Compared to the basic physical approach, the brute force solution of the environmental approach cost approximately eight times more; the analytic variants differed about factor of only 5.5 due to the relatively greater impact of the evaluation routine which required a constant amount of time. The computational costs of the brute force methods were the reason that the experiments were limited to at most three dimensions and two classes. The latter restriction, however, also was more convenient for the endmember selection mechanism applied by the environmental approaches.

	statistical brute force	statistical analytic	physical brute force	physical analytic	environmental brute force	environmental analytic
$n = 2$	40.9	1.8	97.3	3.4	771.3	18.7
$n = 3$	42.0	1.8	9386.6	3.6	76523.0	19.8

Table 2: Average processing time (in seconds) for one image made up of two classes. The methods were implemented using MATLAB 5.0 and were executed on a SUN SPARCstation-20 running Solaris 2.5.1.

5 Summary and conclusions

In this chapter a new linear mixture model was introduced, which attributes the variations in the spectral response of similarly composed pixels to physical factors instead of random, statistical noise. In this physical model, the correlations that exist between the endmember spectra of neighbouring pixels can be described and exploited to acquire a more accurate estimation of the proportions of a mixed pixel. Solving this model is rather complicated, because it requires estimating the spectral response of each component of the mixed pixel. However, this additional information may be a useful feature for some applications. Based on the characteristics of the component spectra, the proportion of the pixel area covered by each class can be calculated.

To determine the performance of two methods based on the statistical models and four methods based on the physical model, a number of experiments were carried out. In these experiments, artificial images of agricultural scenes were generated using real Landsat-TM data and processed in order to make a quantitative comparison of the accuracy of the decomposition methods possible. It appeared that, independent of the approach followed, the proportions were estimated more accurately as the dispersion of the endmembers got larger; increasing the number of bands also led to a higher accuracy. However, the choice of the mixture model had a considerable influence on the accuracy as well. The environmental methods based on the physical model, which utilise the fact that neighbouring pixels are correlated, gave the most accurate estimates. The results of the decomposition methods based on the classic statistical model were less good, while those following the basic physical approach, which ignored the pixel correlations altogether, had an accuracy that was even slightly lower. Based on these results, it can be concluded that the physical linear mixture model is better suited to decompose mixed pixels than the classic statistical model, but only if the correlations between neighbouring pixels are exploited.

In addition to the accuracy of the proportion estimates, the accuracy of the estimated component spectra and the required processing time may be of importance. The components of a mixed pixel were determined more accurately using the environmental approach than using the basic physical approach; the methods based on the statistical linear mixture model did not provide an estimate of the components at all. However, even the components estimated by the environmental approaches did not seem to be very accurate, although this can only be decided in relation to the purpose they are used for. The processing speed of the decomposition methods depended primarily on the method of solution—brute force or mathematical analysis. Whereas the analytic solutions of all models could be calculated very fast due to their polynomial complexity, the methods utilising brute force had an exponential complexity and were much slower. If the methods are ordered according to their underlying paradigms, those based on the statistical model were the fastest, followed by the basic physical approach, and the environmental approach. Therefore, it can be concluded that the analytic solution of the statistical model should be selected if processing speed is of the utmost importance. In case the accuracy of the estimated proportions is the most important, or if the spectral response of the components of a mixed pixel must be extracted, solving the environmental variant of physical linear mixture model is the best choice.

References

HORWITZ, H. M., NALEPKA, R. F., HYDE, P. D., and MORGENSTERN, J. P., 1971, Estimating the proportions of objects within a single resolution element of a multispectral scanner. *Proceedings of the 7th International Symposium on Remote Sensing of Environment* (Ann Arbor, Michigan: Environmental Research Institute of Michigan), pp. 1307–1320.

KLEIN GEBBINCK, M. S., (forthcoming), *Decomposition of mixed pixels in remote sensing images to improve the area estimation of agricultural fields.* PhD thesis Faculty of Mathematics and Informatics, University of Nijmegen, Nijmegen, The Netherlands.

SABINS, F. F., 1987, *Remote Sensing: principles and interpretation*, Second edition (New York: Freeman and Company).

SCHOUTEN, T. E., and KLEIN GEBBINCK, M. S., 1995, Quality measures for image segmentation using generated images. *Proceedings Image and Signal Processing for Remote Sensing II* (Washington: S.P.I.E.), pp. 411–422.

SETTLE, J. J., and DRAKE, N. A., 1993, Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, 14, 1159–1177.

STRANG, G., 1980, *Linear algebra and its applications*, Second edition (New York: Academic Press).