Decomposition of mixed pixels

Maurice S. klein Gebbinck, Theo E. Schouten

University of Nijmegen, Computing Science Institute,
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

ABSTRACT

Within the instantaneous field of view of a scanning device often more than one object is included, resulting in a pixel in which several characteristics are mixed. Classically the proportions of the components of such a mixed pixel are estimated using a linear mixture model. In this paper a new method is introduced for estimating the characteristics of these components, from which their proportions can be derived. Experiments with simulated data sets are conducted to compare the methods regarding their accuracy on estimating the proportions. In addition it is determined how well the proposed method can estimate the characteristics of each component.

Keywords: ground cover proportions, linear mixture model, mixed pixel, remote sensing, subpixel resolution.

1 INTRODUCTION

Thematic maps such as ground cover maps are used for many purposes including agricultural resource management, environment quality assessment and geological studies. Usually these maps are composed using remotely-sensed data, which is acquired through airborne or space-borne scanners. Each pixel in a ground cover map is assigned to only one of a small number of categories. However, because of the limited spatial resolution of the scanners used (e.g. Landsat Thematic Mapper 30x30 m^2, Advanced Very High Resolution Radiometer 1x1 km^2), often more than one ground cover type is present within the instantaneous field of view. The result is a pixel in which the characteristics of the different ground cover types are mixed. These so-called mixed pixels are not only difficult to classify, but they for instance also decrease the accuracy of a crop area estimation.

A more suitable representation of mixed pixels is the mixture map, which is a series of maps giving the concentration per ground cover type. In such a map each pixel is proportionally assigned to several categories. The classical technique to estimate these proportions is based on the linear mixture model and was first described by Horwitz[4] in 1971. Especially in recent years this method has been used in a great many case-studies with different ground cover types, e.g. sparsely vegetated semi-arid rangelands,[5] urban land cover,[3] tropical forests[1] and agricultural crops.[6] Although other approaches have been taken,[7,2] the basic technique is still the same.

In this paper an alternative solution to the mixed pixel problem is presented. Using a bit more complicated model based on linear mixing than Horwitz, a mixed pixel is decomposed into a number of distinct components. Each component represents the contribution made by a particular ground cover type present in the scanner's instantaneous field of view associated with the mixed pixel, and can be modelled as a pure sample drawn from the probability distribution underlying that ground cover type. Once the individual components have been estimated, their proportions can easily be determined.

Section 2 briefly describes the classical method of Horwitz as well as one of its fast variants. In section 3 the alternative approach mentioned above is explained in more detail. The experiments which have been conducted to determine the capabilities and limitations of the methods are discussed in section 4. In section 5 finally conclusions are drawn and possible directions for future work are given.

2 CLASSICAL LINEAR MIXTURE MODELLING

A mixture model predicts the response of a system based on the proportions and characteristics of its components. When dealing with remotely-sensed data, the somewhat restricting linear mixture model has proven to be a acceptable abstraction. For this model the assumption is made that each photon reaching the scanner has interacted with only one cover type. Suppose the scanner collects data on n spectral bands and the maximum number of ground cover components equals c. Following the nomenclature of Settle[7] the linear mixture model in this case is defined by:

$$\mathbf{x} = \mathbf{M}\mathbf{f} \tag{1}$$

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 $\mathbf{f}$. $\mathbf{M}$ is a (nxc) matrix with each column $\mathbf{m_j}$ containing the end-member spectrum of cover type j. An end-member spectrum is the response that is observed for a pure pixel, i.e. consisting of a single ground cover type. Determination of the end-member spectra can be done through laboratory measurements or by inference from supervised data, both of which are not considered in this paper. The n linear equations defined by Eq. (1) are referred to as the mixing equations.

In order for the mixture model to be compliant with the laws of nature a few constraints have to be satisfied. The first constraint is that the pixel is well-defined by its components, whose proportions should therefore add to unity:

$$\sum_{j=1}^{c} \mathrm{f}_j = 1 \tag{2}$$

Equation (2) is called the sum-to-one constraint and can easily be satisfied. A second restriction is that the contribution each component makes to the mixed pixel cannot be negative:

$$\mathrm{f}_j \geq 0 \quad \text{for} \quad j = 1, \ldots, c. \tag{3}$$

This last condition is referred to as the positivity constraint. Unlike the sum-to-one constraint, satisfaction of the positivity constraint is not straightforward. Sometimes this constraint is ignored during the estimation of the proportions vector, but is satisfied by a separate post-processing step—see further in this section.

If all the end-member spectra are linearly independent of each other, the system of n+1 equations defined by Eqs. (1) and (2) will always have a unique solution when c=n+1. If the number of components is larger, the proportions vector can be solved in infinitely many ways among which no best solution can be identified. In case less components are mixed at most one solution exists, but often there is none. One possible explanation for the absence of an exact solution in the last case is the occurrence of sensor noise (Settle[7]), which can be modelled by modifying the mixing equations to:

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

The (nx1) error vector $\mathbf{e}$ is assumed to have a multivariate normal distribution with a zero mean and covariance matrix $\mathbf{N_f}$, which depends on the composition of the pixel. The contribution each component makes to $\mathbf{N_f}$ is relative to its proportion and is dependent on the covariance matrix $\mathbf{N_j}$ of its corresponding ground cover type:

$$\mathbf{N_f} = \sum_{j=1}^{c} \mathrm{f}_j \mathbf{N_j} \tag{5}$$

A detailed derivation of Eq. (5) based on the theory of statistics can be found in Horwitz.[4]

The estimation of the proportions vector underlying pixel $\mathbf{x}$ based on Eq. (4) yields an infinity of solutions. Following the maximum likelihood approach the $\mathbf{f}$ is selected which minimizes:

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

Here $\ln |\mathbf{N_f}|$ denotes the natural logarithm of the determinant of $\mathbf{N_f}$. The dependence of matrix $\mathbf{N_f}$ on the proportions vector to be estimated makes an analytical solution of $\mathbf{f}$ very complicated. So far, the only known strategy leading to a solution in general is brute force, which is computationally very expensive. However, with this approach satisfaction of the sum-to-one and positivity constraints is obtained simply by choosing suitable candidate proportions vectors.

If the covariance matrices $\mathbf{N_j}$ are all equal to say $\mathbf{N}$, a fast analytic solution for $\mathbf{f}$ can be derived. Setting the partial derivatives with respect to each $\mathbf{f}_j$ to zero gives the unconstrained estimator

$$\hat{\mathbf{f}}_0(\mathbf{x}) = \mathbf{U}\mathbf{M}^T\mathbf{N}^{-1}\mathbf{x} \quad \text{with} \quad \mathbf{U}^{-1} = \mathbf{M}^T\mathbf{N}^{-1}\mathbf{M}. \tag{7}$$

Imposing the sum-to-one constraint changes the estimator to

$$\hat{\mathbf{f}}(\mathbf{x}) = \hat{\mathbf{f}}_0(\mathbf{x}) + \alpha \left(1 - \mathbf{1}^T\hat{\mathbf{f}}_0(\mathbf{x})\right) \mathbf{U1} \quad \text{with} \quad \alpha = \left(\mathbf{1}^T\mathbf{U1}\right)^{-1} \tag{8}$$

where $\mathbf{1}$ is the (cx1) vector consisting entirely of 1's. Unfortunately the proportions estimated using Eq. (8) are not always non-negative, and no simple transformation exists which guarantees this property. If the positivity constraint must always be satisfied alternative methods of solution should be used, or the estimation must be corrected afterwards, for instance by a subsequent exhaustive search of its neighbourhood. However, a much faster method is setting negative proportions to zero and renormalizing the proportions vector. Although this strategy gives less accurate results, it is heavily used because of its simplicity and speed.

The analytic method gives a huge increase in speed compared to the brute force method, especially when c is large. This is why the matrices $\mathbf{N_j}$ are often approximated by their mean:

$$\overline{\mathbf{N}} = 1/c \sum_{j=1}^{c} \mathbf{N_j} \tag{9}$$

As long as the covariances in $\mathbf{N_j}$ are relatively small compared to the dispersion of the end-members $\mathbf{m_j}$, this shortcut does not seriously affect the estimation.

A clarifying view on the linear mixture model can be obtained by looking at the geometry of the n-dimensional feature space. In general, pixels that are a linear combination of c components and satisfy the sum-to-one constraint lie in the (c-1)-dimensional linear subspace joining the end-member spectra. The simplex defined by these end-members contains all pixels that also satisfy the positivity constraint. As sensor noise causes pixels to be positioned somewhat outside the subspace, minimizing Eq. (6) with a fixated $\mathbf{N_f}$ corresponds to finding a point $\mathbf{p}$ inside the subspace for which the scaled quadratic distance to $\mathbf{x}$ is minimal. For example (see Fig. 1), in the absence of noise mixtures of two components lie on the line through $\mathbf{m_1}$ and $\mathbf{m_2}$, or, if the positivity constraint is satisfied, on the line between $\mathbf{m_1}$ and $\mathbf{m_2}$. If a noisy pixel $\mathbf{x}$ is encountered, a projection $\mathbf{p}$ of $\mathbf{x}$ on the line is looked for such that the scaled quadratic distance between $\mathbf{p}$ and $\mathbf{x}$ is minimal.

3 MINIMIZING AN ALTERNATIVE COST FUNCTION

Where classical linear mixture modelling ascribes the occurrence of imperfect mixed pixels to sensor noise, it can also be explained by the natural variation present within each ground cover type. These variations are

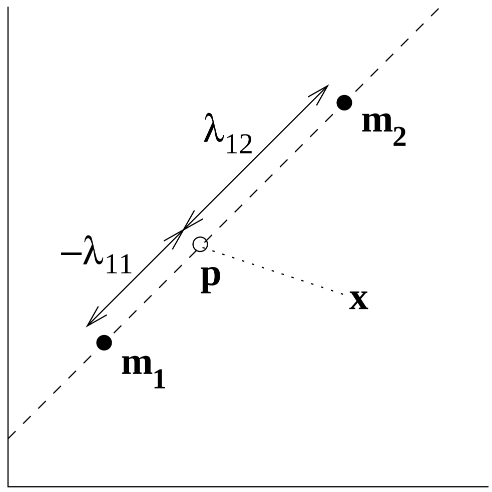


Figure 1: Two-dimensional feature space with end-members $\mathbf{m_1}$ and $\mathbf{m_2}$. Linear combinations of $\mathbf{m_1}$ and $\mathbf{m_2}$ satisfying the sum-to-one constraint lie on the dashed line, the positivity constraint results in pixels positioned on $\mathbf{m_1 m_2}$. Point $\mathbf{p}$ has the smallest scaled quadratic distance to $\mathbf{x}$.

caused by local differences in humidity, soil type, elevation, etcetera. Formalizing this model leads to the following mixing equations:

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

As before $\mathbf{x}$ denotes a pixel and $\mathbf{f}$ its proportions vector. Each column $\mathbf{m_j}$ of matrix $\mathbf{M}$ now contains the mean of ground cover class j, a concept very similar to that of end-member spectrum. The (nxc) matrix $\mathbf{E}$ consists of columns $\mathbf{e_j}$ describing the deviation from class j's mean. Assuming ground cover class j has a multivariate normal distribution with mean $\mathbf{m_j}$ and covariance matrix $\mathbf{N_j}$, $\mathbf{e_j}$ is distributed multivariate normal with a zero mean and covariance matrix $\mathbf{N_j}$. Together $\mathbf{m_j}$ and $\mathbf{e_j}$ define the characteristics of the j-th component of the mixed pixel.

When estimating the individual components of a mixed pixel, Eq. (10) can be solved in infinitely many ways. Analogue to Eq. (6) the best solution is defined to be the one which minimizes:

$$\sum_{j=1}^{c} \mathrm{f}_j \left(\mathbf{e_j}^T \mathbf{N_j}^{-1} \mathbf{e_j} + \ln |\mathbf{N_j}| \right) \tag{11}$$

But while for classical linear mixture modelling $\mathbf{e}$ can easily be expressed in $\mathbf{x}$, $\mathbf{M}$ and $\mathbf{f}$, this cannot be done for the $\mathbf{e_j}$. Not only are all $\mathbf{e_j}$ related, $\mathbf{f}$ depends on $\mathbf{E}$ as well. However, from the geometry of the feature space a method can be derived. As before, mixtures of c components satisfying the sum-to-one constraint lie in the (c-1)-dimensional linear subspace containing the c class means, with the positivity constraint restricting the subspace to the simplex with vertices $\mathbf{m_j}$. Where classical linear mixture modelling searches the original subspace for a point with a minimal scaled quadratic distance to pixel $\mathbf{x}$, minimizing Eq. (11) corresponds to finding a different (c-1)-dimensional linear subspace containing $\mathbf{x}$ and a projection $\mathbf{p_j}$ for each $\mathbf{m_j}$ such that the Mahalanobis distance (M-distance) of $\mathbf{p_j}$ to $\mathbf{m_j}$ weighted by the corresponding fraction is minimal. For example, when dealing with two classes, a line through $\mathbf{x}$, $\mathbf{p_1}$ and $\mathbf{p_2}$ is looked for (see Fig. 2), while in case of c=3 the objective is to find a plane containing $\mathbf{x}$, $\mathbf{p_1}$, $\mathbf{p_2}$ and $\mathbf{p_3}$ (Fig. 3). For both the line as well as the plane holds that the weighted sum of the M-distances between $\mathbf{m_j}$ and $\mathbf{p_j}$ is minimal.

In n-space, a linear subspace of dimension c-1 is spanned by c-1 linearly independent (nx1) vectors $\mathbf{b_j}$. So the points $\mathbf{p_j}$ in the linear subspace through $\mathbf{x}$ can can be described as a linear combination of these vectors:

$$\mathbf{p_j} = \mathbf{x} + \lambda_{1j}\mathbf{b_1} + \ldots + \lambda_{(c-1)j}\mathbf{b_{c-1}} = \mathbf{x} + \mathbf{B}\lambda_\mathbf{j} \tag{12}$$

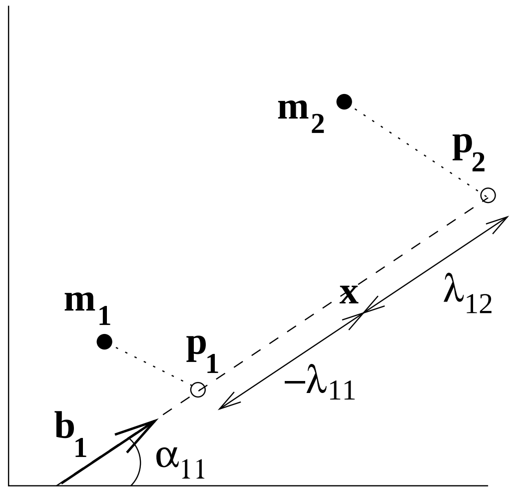


Figure 2: Two-dimensional feature space with class means $\mathbf{m_1}$ and $\mathbf{m_2}$. Given the dashed line defined by $\mathbf{b_1}$, the weighted Mahalanobis distance between $\mathbf{p_1}$ and $\mathbf{m_1}$ as well as $\mathbf{p_2}$ and $\mathbf{m_2}$ is minimal. The objective is to find that line through $\mathbf{x}$ resulting in a minimal weighted sum of M-distances.

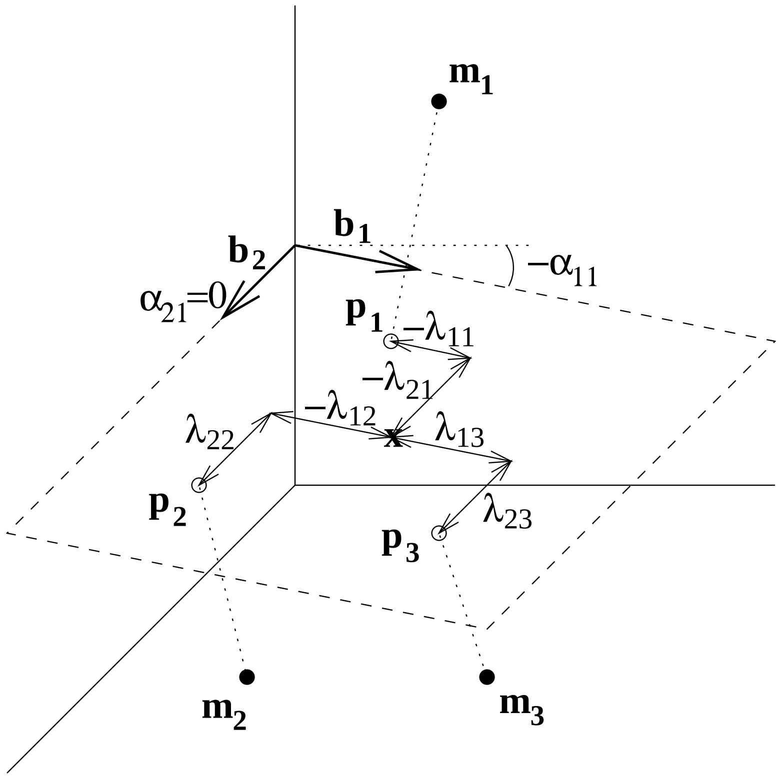


Figure 3: Three-dimensional feature space with class means $\mathbf{m_1}$, $\mathbf{m_2}$ and $\mathbf{m_3}$. For the plane spanned by $\mathbf{b_1}$ and $\mathbf{b_2}$, the weighted M-distance between each pair $\mathbf{p_j}$ and $\mathbf{m_j}$ is minimal. The goal is to find the plane through $\mathbf{x}$ for which the weighted sum of M-distances is minimal.

Here $\boldsymbol{\lambda}_\mathbf{j}$ represents the ((c-1)x1) column vector $[\lambda_{1j}, \ldots, \lambda_{(c-1)j}]^T$ and $\mathbf{B}$ is the (nx(c-1)) matrix with columns $\mathbf{b_j}$. Since $\mathbf{p_j} = \mathbf{m_j} + \mathbf{e_j}$, substitution of Eq. (12) in Eq. (10) results in a system of equations which is guaranteed to contain a solution for $\mathbf{f}$, provided all $\mathbf{m_j}$ are linearly independent of each other. Expressed in the coordinate-system defined by the set of basis vectors $\{\mathbf{b_1}, \ldots, \mathbf{b_{c-1}}\}$ and origin $\mathbf{x}$, the mixing equations are:

$$\boldsymbol{\Lambda}\mathbf{f} = \mathbf{0} \tag{13}$$

The ((c-1)xc) matrix $\boldsymbol{\Lambda}$ consists of columns $\boldsymbol{\lambda}_\mathbf{j}$, and $\mathbf{0}$ denotes the null vector, in this case of dimension c-1. If the sum-to-one constraint is explicitly satisfied by taking $\mathrm{f}_c = 1 - \sum_{j=1}^{c-1} \mathrm{f}_j$, Eq. (13) can be transformed into

$$\boldsymbol{\Lambda}'\mathbf{f}' = -\boldsymbol{\lambda}_\mathbf{c} \quad \text{with} \quad \boldsymbol{\Lambda}' = [\boldsymbol{\lambda}_\mathbf{1} - \boldsymbol{\lambda}_\mathbf{c}, \ldots, \boldsymbol{\lambda}_\mathbf{c-1} - \boldsymbol{\lambda}_\mathbf{c}] \tag{14}$$

where $\mathbf{f}'$ is equal to proportions vector $\mathbf{f}$ without its last element f_c. This last system of equations is well-known and has a solution for $\mathbf{f}'$ that is normally given using Cramer's rule[8]:

$$\mathrm{f}'_j = \frac{|\boldsymbol{\Lambda}'_j(-\boldsymbol{\lambda}_\mathbf{c})|}{|\boldsymbol{\Lambda}'|} \tag{15}$$

As before $|\cdot|$ stands for taking the determinant; $\boldsymbol{\Lambda}'_j(-\boldsymbol{\lambda}_\mathbf{c})$ is the ((c-1)x(c-1)) matrix $\boldsymbol{\Lambda}'$ with the j-th column replaced by $-\boldsymbol{\lambda}_\mathbf{c}$. With Eq. (15) $\mathbf{f}$ can be given using the λ_{ij} only, for example when dealing with two components $\mathbf{f}^T = (\lambda_{12}/(\lambda_{12} - \lambda_{11}), -\lambda_{11}/(\lambda_{12} - \lambda_{11}))$—see also Fig. 2.

Substitution of the equations found for $\mathbf{p_j}$ and f_j—Eqs. (12) and 15—in Eq. (11) results in an expression in which the λ_{ij} and $\mathbf{b_j}$ are the only variables. The next step in minimizing this error term is expressing the scalars λ_{ij} in terms of the basis vectors $\mathbf{b_j}$, which corresponds to finding the best solution for a fixated though undefined linear subspace. Unfortunately the error term is still too difficult for the λ_{ij} to be extracted analytically, due to their mutual dependencies. Therefore all f_j are initially approximated by $1/c$, which boils down to independently minimizing $\mathbf{e_j}^T\mathbf{N_j}^{-1}\mathbf{e_j}$ for each component. Analogue to the derivation of Eq. (7), successively setting the partial derivatives with respect to $\lambda_{1j}, \ldots, \lambda_{(c-1)j}$ to zero gives:

$$\boldsymbol{\lambda}_\mathbf{j} = (\mathbf{B}^T\mathbf{N_j}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{N_j}^{-1}\mathbf{m}'_\mathbf{j} \quad \text{with} \quad \mathbf{m}'_\mathbf{j} = \mathbf{m_j} - \mathbf{x}. \tag{16}$$

Although this initial guess is usually quite good, further optimization can be achieved by using conjugate gradient methods for instance.

At this stage, the only unknowns in Eq. (11) are the c-1 basis vectors $\mathbf{b_j}$. Considering all elements b_{ij} to be independent variables is incorrect, because the basis vectors have to be linearly independent of each other at all times. Furthermore it results in (c-1)n degrees of freedom (DF), which is far too much according to the following intuitive argument. To define a k-dimensional linear subspace of n-space, k linearly independent vectors have to be chosen, giving DF=n+(n-1)+...+(n-k+1). However, many combinations result in the same subspace since a k-space can be defined with k+(k-1)+...+1=k(k+1)/2 degrees of freedom. In order to cover all different subspaces no more than DF=(n+...+(n-k+1))-(k+...+1)=k(n-k) is needed; in case k=c-1, varying all b_{ij} wastes $(c\text{-}1)^2$ degrees of freedom. A more efficient construction of the k basis vectors in an n-dimensional space is based on the usual definition of a k-dimensional space with k orthonormal vectors. For each dimension i higher than k, basis vector $\mathbf{b_j}$ is multiplied with $\cos\alpha_{ji}$ and expanded with a term $\sin\alpha_{ji}$. For example, if k=1 the construction of $\mathbf{b_1}$ is given by

$$[1] \;\rightarrow\; \begin{bmatrix} \cos\alpha_{11} \\ \sin\alpha_{11} \end{bmatrix} \;\rightarrow\; \begin{bmatrix} \cos\alpha_{11}\cos\alpha_{12} \\ \sin\alpha_{11}\cos\alpha_{12} \\ \sin\alpha_{12} \end{bmatrix} \;\rightarrow\; \cdots \quad \text{for } n\text{=1, 2, 3, } \ldots \tag{17}$$

while in case k=2, $(\mathbf{b_1}, \mathbf{b_2})$ is formed following the path

$$\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \;\rightarrow\; \left(\begin{bmatrix} \cos\alpha_{11} \\ 0 \\ \sin\alpha_{11} \end{bmatrix}, \begin{bmatrix} 0 \\ \cos\alpha_{21} \\ \sin\alpha_{21} \end{bmatrix} \right) \;\rightarrow\; \cdots \quad \text{for } n\text{=2, 3, } \ldots \tag{18}$$

Unfortunately, except for k=1, some subspaces are not covered, while other subspaces can be formed in multiple ways. For instance in case k=2 and n=3, if $\alpha_{11}=\frac{\pi}{2}$ all values of α_{21} result in the same subspace (YZ-plane). On the other hand, apart from the YZ- and XZ-plane ($\alpha_{21} = \frac{\pi}{2}$), no other subspace perpendicular to the XY-plane can be formed. Therefore, for certain values of the angles α_{ji} the definition of the basis vectors has to be modified. Again for k=2 and n=3, Eq. (18) is valid only if $\alpha_{11}, \alpha_{21} \neq \frac{\pi}{2}$; the missing planes are defined by:

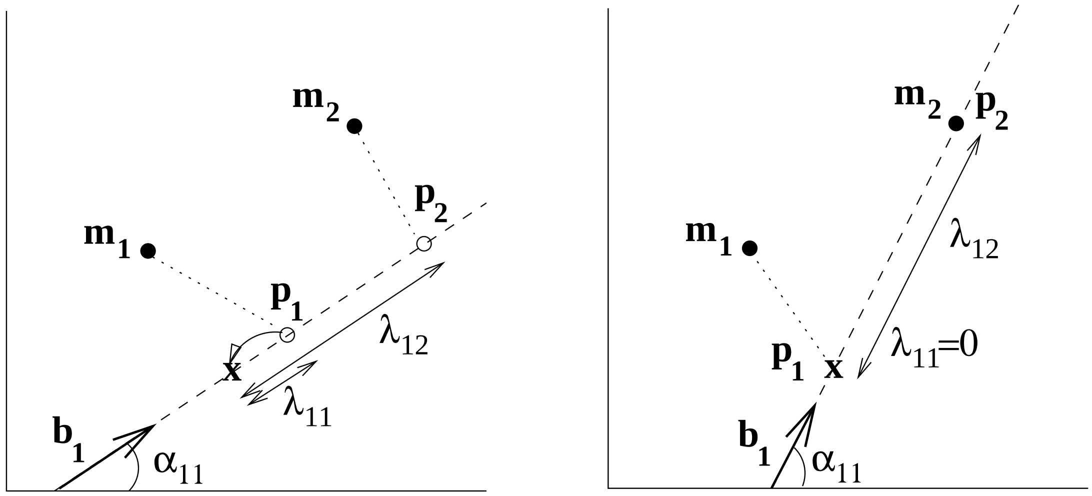
$$(\mathbf{b_1}, \mathbf{b_2}) = \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \cos\alpha_{21} \\ \sin\alpha_{21} \\ 0 \end{bmatrix} \right) \tag{19}$$

For higher dimensional k- and n-spaces similar modifications must be made. Although the above construction of the basis vectors is rather complex, it does produce a set of linearly independent $\mathbf{b_j}$ with only DF=$k(n$-$k)$.

The conditional definition of the basis vectors in the previous paragraph makes finding a general analytic solution for the problem of minimizing Eq. (11) impossible. But even a partial solution using only the basic definition of the $\mathbf{b_j}$ cannot be found, due to the complicated relations between the multiple variables. On top of that, an educated guess which can be optimized afterwards—a strategy that has been followed before—is currently not available. Therefore the only option is to check all k-dimensional subspaces using brute force. But since only a finite number of subspaces can be evaluated, the minimum found will always be an approximation. The best set of subspaces to be checked is the one whose elements are uniformly distributed over the entire n-space, but this requires a more complex definition of the basis vectors. A more obvious set contains subspaces that are spread evenly over the $k(n$-$k)$-dimensional parameter space belonging to the α_{ji}. This set can easily be created by determining the $\mathbf{b_j}$ at regular intervals between successive values of each α_{ji}. Although the domain of each parameter is $(-\pi, \pi]$, it can be restricted to $(-\frac{\pi}{2}, \frac{\pi}{2}]$ because not both $\mathbf{b_j}$ and -$\mathbf{b_j}$ have to be examined. By varying the interval size, the accuracy of the approximation can be traded off against computational costs.

Summarizing the method so far, for each (c-1)-dimensional linear subspace defined by the selected values for the α_{ji}, the following steps are taken. First the optimal set of λ_{ij} is estimated with Eq. (16). Next, from this set the corresponding proportions f_j are derived using Cramer's rule (Eq. (15)). Finally the weighted sum of M-distances given by Eq. (11) is calculated for comparison with other subspaces; the subspace with the smallest weighted sum of M-distances is chosen to produce the matrix $\mathbf{E}$ which best solves Eq. (10). Nowhere in this method is the positivity constraint taken into account, which for some subspaces may result in a negative weighted sum of M-distances. Since correct solutions always have a weighted sum of M-distances greater than zero, one such subspace can lead to a solution that is completely worthless. However, this subspace cannot be neglected either, because one of its sub-optimal solutions satisfying the positivity constraint may still be the best solution possible. A way to solve this problem is best illustrated by considering the case where only two components have to be unmixed, see Fig. 4. Expanding Eq. (15) results in $\mathbf{f}^T = (\lambda_{12}/(\lambda_{12} - \lambda_{11}), -\lambda_{11}/(\lambda_{12} - \lambda_{11}))$, so correct proportions are computed only if λ_{11} and λ_{12} have opposite signs—$\mathbf{x}$ lies between $\mathbf{p_1}$ and $\mathbf{p_2}$. If for the subspace under observation this condition is not met, either $\mathbf{p_1}$ or $\mathbf{p_2}$ should be moved. Assuming again that f_j=$1/c$, analysis of the partial derivative of Eq. (11) with respect to a single λ_{1j} suggests that $\mathbf{p_j}$ should be moved away from its optimal position as little as possible. So, to satisfy the positivity constraint, one of the $\mathbf{m_j}$ must be projected onto $\mathbf{x}$ (Fig. 4(a)), which corresponds to the situation where the pixel is a mix of only one component. Although this correction should be made for all subspaces giving negative proportions, none will result in a smaller weighted sum of M-distances than the subspace containing $\mathbf{x}$ and the other $\mathbf{m_j}$ (Fig. 4(b)), being a valid solution itself. Mixtures with more than two components can be dealt with in a similar manner, resulting in the following recursive algorithm to decompose mixed pixel $\mathbf{x}$ into c components, while satisfying the positivity constraint:

1. determine which (c-1)-subspace has the smallest weighted sum of M-distances and (coincidentally) satisfies the positivity constraint
2. FOR all c combinations of c-1 components DO decompose $\mathbf{x}$ into c-1 components OD
3. select of the c+1 solutions above the one with the smallest weighted sum of M-distances



(a) The solution does not satisfy the positivity constraint; $\mathbf{p_1}$ has to be moved to $\mathbf{x}$.

(b) The subspace through $\mathbf{x}$ and $\mathbf{m_2}$ results in a smaller weighted sum of M-distances.

Figure 4: For the positivity constraint to be satisfied, $\mathbf{x}$ must lie between $\mathbf{p_1}$ and $\mathbf{p_2}$. If this is not the case, one of the $\mathbf{p_j}$ has to be moved as little as possible. The weighted sum of M-distances after this correction can be underestimated by the subspace containing $\mathbf{x}$ and the other $\mathbf{m_j}$.

The recursion in step 2 terminates as soon as the zero-dimensional subspace is encountered. In this case pixel $\mathbf{x}$ is no longer considered to be mixed, but to consist of a single ground cover type only.

4 SOME EXPERIMENTS

In order to determine the capabilities of the various approaches to estimate the proportions of a pixel, a few experiments have been carried out. Another aim of the experiments is to get an idea how well the alternative method is able to estimate the characteristics of a pixel's components. In each experiment a set of 250 mixed pixels was processed by the classical, shortcut and alternative method, while computing several error measures. To calculate these measures, the true characteristics of the components of a mixed pixel must be known. Since no set of real pixels with known compositions was available, artificial samples had to be generated.

The artificial pixels were generated according to the linear mixture model described by Eq. (10). As before, each ground cover class j is assumed to have a multivariate normal distribution with mean vector $\mathbf{m_j}$ and covariance matrix $\mathbf{N_j}$. To make sure the artificial pixels are as realistic as possible, the distributions of several large agricultural fields in a real Landsat-TM image of Portugal were taken as the distributions of the ground cover classes. Of the six fields that were selected, see Table 1, three fields (A–C) had little overlapping distributions; the other three fields (D–F) had distributions that overlapped much more. The first step in generating a mixed pixel is determining its proportions vector $\mathbf{f}$, which was done by independently drawing each proportion from a uniform distribution and normalizing $\mathbf{f}$ afterwards. Next, for every component a sample was drawn from the corresponding multivariate normal distribution, defining its characteristics. The resulting pixel was calculated by summing the individual components weighted by their proportion. This way fourteen sets of 250 pixels each were created, seven sets with pixels composed of classes A–C, and seven more difficult sets with pixels composed of classes D–F. Six of the first seven sets contained pixels in which only two out of three classes were mixed (AB, AC

class	mean	covariance	class	mean	covariance
A	$\begin{bmatrix} 73.0 \\ 33.0 \\ 29.0 \end{bmatrix}$	$\begin{bmatrix} 2.5 & 1.6 & 1.3 \\ 1.6 & 2.5 & 1.6 \\ 1.3 & 1.6 & 2.1 \end{bmatrix}$	D	$\begin{bmatrix} 92.6 \\ 48.1 \\ 79.9 \end{bmatrix}$	$\begin{bmatrix} 5.4 & 3.1 & 4.7 \\ 3.1 & 2.9 & 4.0 \\ 4.7 & 4.0 & 7.1 \end{bmatrix}$
B	$\begin{bmatrix} 89.0 \\ 46.0 \\ 74.0 \end{bmatrix}$	$\begin{bmatrix} 11.3 & 7.3 & 12.2 \\ 7.3 & 6.0 & 9.4 \\ 12.2 & 9.4 & 17.3 \end{bmatrix}$	E	$\begin{bmatrix} 96.3 \\ 50.5 \\ 82.0 \end{bmatrix}$	$\begin{bmatrix} 12.2 & 6.3 & 5.7 \\ 6.3 & 4.4 & 5.0 \\ 5.7 & 5.0 & 13.2 \end{bmatrix}$
C	$\begin{bmatrix} 109.0 \\ 61.0 \\ 100.0 \end{bmatrix}$	$\begin{bmatrix} 9.6 & 6.1 & 9.0 \\ 6.1 & 5.2 & 7.7 \\ 9.0 & 7.7 & 14.9 \end{bmatrix}$	F	$\begin{bmatrix} 100.2 \\ 53.6 \\ 89.8 \end{bmatrix}$	$\begin{bmatrix} 6.0 & 4.2 & 7.1 \\ 4.2 & 4.2 & 6.5 \\ 7.1 & 6.5 & 12.0 \end{bmatrix}$

Table 1: Distributions of the selected ground cover classes (TM bands 1–3).

or BC), having dimension two or three. The seventh set consisted of pixels that were a mix of all three classes, necessarily being of dimension three. The other seven sets were created similarly, using the remaining classes D–F.

To measure how accurately any method can estimate the proportions of a mixed pixel, the pixel's true proportions f_j are compared with their estimates $\hat{\mathrm{f}}_j$:

$$\varepsilon_f = \left\langle \frac{1}{2} \sum_{j=1}^{c} \left| \mathrm{f}_j - \hat{\mathrm{f}}_j \right| \right\rangle \tag{20}$$

In this equation $|\cdot|$ represents the absolute value, and $< \cdot >$ denotes taking the average over the entire set of pixels. Since both the true as well as the estimated proportions vector satisfies the sum-to-one constraint, the value of ε_f is restricted to the interval [0,1]. The closer ε_f gets to 0, the better the proportions are estimated.

Another interesting question is how reliable the estimation is. It is believed that the better pixel $\mathbf{x}$ can be reconstructed with the estimated proportions, the more reliable the estimation is. For the classical methods this can be measured by:

$$\varepsilon_x = \left\langle \frac{1}{n} \left(\mathbf{x} - \mathbf{M}\hat{\mathbf{f}} \right)^T \mathbf{N_f}^{-1} \left(\mathbf{x} - \mathbf{M}\hat{\mathbf{f}} \right) \right\rangle \tag{21}$$

The alternative method however can always reconstruct $\mathbf{x}$ perfectly. In this case the estimation is considered to be more reliable as the characteristics of the components are more probable, i.e. approach the class mean vector:

$$\varepsilon_x = \left\langle \frac{1}{n} \sum_{j=1}^{c} \mathrm{f}_j \left(\hat{\mathbf{p}}_\mathbf{j} - \mathbf{m_j} \right)^T \mathbf{N_j}^{-1} \left(\hat{\mathbf{p}}_\mathbf{j} - \mathbf{m_j} \right) \right\rangle \tag{22}$$

Although both ε_x are related more closely than expected at first sight, they probably cannot be compared.

A last important measure that can only be calculated for the alternative method expresses how accurately the characteristics of a component are estimated. This measure is computed for every ground cover class j separately, and is given by:

$$\varepsilon_{p_j} = \left\langle \frac{1}{n} \left(\mathbf{p_j} - \hat{\mathbf{p}}_\mathbf{j} \right)^T \mathbf{N_j}^{-1} \left(\mathbf{p_j} - \hat{\mathbf{p}}_\mathbf{j} \right) \right\rangle \tag{23}$$

In Eq. (23) ε_{p_j} is defined to be the Mahalanobis distance between the true characteristics of a component $\mathbf{p_j}$ and its estimate $\hat{\mathbf{p}}_\mathbf{j}$, divided by the dimension of the data, and averaged over the entire set of pixels. Similar to the other measures, as ε_{p_j} approaches 0, the estimation becomes better.

mix	method	n=2		n=3		mix	method	n=2		n=3	
		ε_f	ε_x	ε_f	ε_x			ε_f	ε_x	ε_f	ε_x
	classical	0.0784	0.3269	0.0279	0.3665		classical	0.3027	0.4063	0.3153	0.4727
AB	shortcut	0.0781	0.3263	0.0290	0.3733	DE	shortcut	0.2902	0.3842	0.2940	0.4126
	alternate	0.0775	0.3236	0.0274	0.3728		alternate	0.3004	0.3290	0.2955	0.4074
	classical	0.0349	0.3159	0.0212	0.3681		classical	0.1536	0.3528	0.1447	0.4047
AC	shortcut	0.0349	0.3170	0.0215	0.3675	DF	shortcut	0.1510	0.3525	0.1440	0.4044
	alternate	0.0347	0.3143	0.0211	0.3919		alternate	0.1522	0.3410	0.1439	0.3913
	classical	0.0838	0.3199	0.0819	0.3778		classical	0.3126	0.3615	0.2425	0.4091
BC	shortcut	0.0838	0.3197	0.0818	0.3776	EF	shortcut	0.2880	0.3411	0.2157	0.3711
	alternate	0.0842	0.3294	0.0828	0.3790		alternate	0.2953	0.2913	0.2155	0.3867
	classical	–	–	0.0997	0.1461		classical	–	–	0.3208	0.2296
ABC	shortcut	–	–	0.1022	0.1830	DEF	shortcut	–	–	0.3358	0.1857
	alternate	–	–	0.0995	0.1513		alternate	–	–	0.3304	0.2022

Table 2: Estimation of the proportions by the classical (0.01), shortcut and alternative ($\pi/100$) method.

In the first batch of experiments, the seven data sets created using ground cover classes which overlapped little (A–C) were processed by the classical method, its shortcut variant, and the alternative method. Since the classical and alternative method utilize brute force, an interval size has to be chosen. For the classical method the interval was set at 0.01, while the alternative method used a similar interval of size $\pi/100$. The other sets, being more difficult because the mixed ground cover classes overlapped much more, were processed in the second batch, using the same three methods with the same interval sizes. The results of these experiments are presented in Table 2.

As expected, the results of the first batch are much better than those of the second batch. Similar, within each batch, the set created of the classes whose distributions overlapped the least—AC and DF, see Table 1—gave the best results. Another non-surprising observation is that as the dimension of the data gets higher, the proportions are estimated more accurately. An explanation for this is that the effects of the random fluctuations in one dimension of the pixel vector are compensated for by the other dimensions. As a consequence, from the classical point of view, the pixel can be reconstructed less well, because the fluctuations in the extra dimensions must be reckoned with also. Since ε_x increases as more dimensions are available, but an estimation based on more information is considered to be more reliable, this measure cannot be used to compare the reliabilities of data sets of different dimensions. A last obvious result is that an increased number of components leads to a less accurate estimation. The pixel itself can be reconstructed better, because with more components a larger subspace is spanned. Comparison of the three methods shows that the accuracy of their estimates does not differ much, especially for the easy sets. However, it is remarkable that for some sets (e.g. DE and EF) the shortcut method results in a better accuracy than the classical method. These differences cannot be explained solely by the fact that the shortcut method minimizes its error term exactly, while the classical method only approximates it, because the difference in accuracy is about 2.5% (EF) and the classical method's interval size is set at 1%. Another factor which might contribute to the unexpected difference in accuracy is the presence of the logarithmic term in Eq. (6). To check this conjecture, two more batches of experiments were carried out in which both the classical and the alternative method were relieved of their logarithmic term. The results of these experiments can be found in Table 3.

Without the logarithmic term, the classical method can make a better estimation of the proportions. The increase in accuracy achieved by this modification is enough to outperform the shortcut method, which is especially apparent from the more difficult data sets like DE (n=3), EF (n=3) and DEF. A similar improvement

mix	method	n=2		n=3		mix	method	n=2		n=3	
		ε_f	ε_x	ε_f	ε_x			ε_f	ε_x	ε_f	ε_x
	shortcut	0.0781	0.3263	0.0290	0.3733		shortcut	0.2902	0.3842	0.2940	0.4126
AB	classical	0.0775	0.3227	0.0276	0.3661	DE	classical	0.2913	0.3894	0.2714	0.4379
	alternate	0.0777	0.3208	0.0273	0.3726		alternate	0.2941	0.3109	0.2763	0.3420
	shortcut	0.0349	0.3170	0.0215	0.3675		shortcut	0.1510	0.3525	0.1440	0.4044
AC	classical	0.0344	0.3153	0.0213	0.3678	DF	classical	0.1525	0.3524	0.1438	0.4040
	alternate	0.0347	0.3143	0.0210	0.3919		alternate	0.1522	0.3411	0.1440	0.3910
	shortcut	0.0838	0.3197	0.0818	0.3776		shortcut	0.2880	0.3411	0.2157	0.3711
BC	classical	0.0838	0.3198	0.0822	0.3780	EF	classical	0.2816	0.3591	0.2007	0.3822
	alternate	0.0843	0.3288	0.0829	0.3783		alternate	0.2854	0.2936	0.2018	0.3616
	shortcut	–	–	0.1022	0.1830		shortcut	–	–	0.3358	0.1857
ABC	classical	–	–	0.0986	0.1456	DEF	classical	–	–	0.3140	0.1878
	alternate	–	–	0.0994	0.1581		alternate	–	–	0.3174	0.1489

Table 3: Estimation of the proportions by the modified classical (0.01) and alternative ($\pi/100$) method.

is accomplished by deleting the logarithmic term from Eq. (11), which is minimized by the alternative method. Although the increase in accuracy is a bit less, the modified alternative method is also more accurate than the shortcut method. Comparing the classical and alternative method, only small differences in accuracy are found, mostly in favor of the classical method.

During all experiments with the alternative method and its modified version, it was measured how well the characteristics of the components were estimated. The results of this investigation are given in Table 4. Averaging over the results of the components per set shows that the modified method estimates the characteristics more accurately. Surprising is that the results of the difficult sets are better than those of the easy sets. Currently an explanation for this phenomenon has not been found yet. The question how well the alternative method can estimate the characteristics of the components however can partially be answered. Since the individual components have been generated in such a way that their expected M-distance to the corresponding class mean is 1.0, the estimated components deviate significantly less from the true components than the true components deviate from their class mean.

mix	n=2		n=3			mix	n=2		n=3		
	ε_{p_1}	ε_{p_2}	ε_{p_1}	ε_{p_2}	ε_{p_3}		ε_{p_1}	ε_{p_2}	ε_{p_1}	ε_{p_2}	ε_{p_3}
AB	0.8205	0.6275	0.8084	0.6028	–	DE	0.8352	0.5513	0.7950	0.4690	–
	0.8400	0.6206	0.8213	0.5965	–		0.8389	0.5756	0.8488	0.5218	–
AC	0.8329	0.6314	0.8159	0.6259	–	DF	0.7757	0.5775	0.7107	0.5654	–
	0.8329	0.6314	0.8264	0.6259	–		0.7758	0.5801	0.7111	0.5662	–
BC	0.6001	0.8340	0.6195	0.7014	–	EF	0.5875	0.7563	0.5177	0.7745	–
	0.5965	0.8511	0.6221	0.6934	–		0.5902	0.7672	0.5302	0.8173	–
ABC	–	–	0.8121	0.8639	0.8818	DEF	–	–	0.8437	0.7801	0.8578
	–	–	0.8238	0.8651	0.9078		–	–	0.8879	0.8106	0.8905

Table 4: Estimation of the components by the modified (top) and original (bottom) alternative method.

5 CONCLUSIONS

A new linear mixture model has been presented, explaining imperfect mixed pixels by natural variation in ground cover classes. Based on this model a method has been developed to estimate the characteristics of the components of a mixed pixel. From these characteristics the proportions of the components can be derived. The positivity constraint can be satisfied following a simple procedure.

A number of experiments have been carried out with artificial data generated using distributions taken from a real satellite image. The accuracy of the estimated proportions mainly depends on the relative position of the class means in feature space. Increasing the number of dimensions influences the accuracy positively, increasing the number of components has a negative influence. The classical, shortcut and alternative method give similar results, though estimations by the shortcut method tends to be slightly more accurate. Modifying the classical and alternative method by deleting their logarithmic term improves their accuracy, exceeding the shortcut method.

The estimation of the characteristics of a pixel's components is more accurate if the class means lie closer to each other in feature space. At the moment this surprising observation cannot be explained. The average Mahalanobis distance between the true and estimated characteristics is less than the expected M-distance between the true components and their class means.

If a fast estimation of the proportions of a mixed pixel is needed, only the shortcut method can be chosen. For a more accurate but much slower estimation the classical method can be used. However, if apart from the proportions also the characteristics of the components have to be estimated, the alternative method is the only option. The number of computations required by the alternative method is even larger than for the classical method, but they can well be done in parallel. Another solution is replacing the brute force approach by an optimizing strategy starting from an acceptable subspace, but this requires an educated guess of the initial subspace.

6 REFERENCES

[1] AM Cross, JJ Settle, NA Drake, and RTM Paivinen. Subpixel measurement of tropical forest cover using AVHRR data. *Int. J. Remote Sensing*, 12(5):1119–1129, 1991.

[2] GM Foody and DP Cox. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *Int. J. Remote Sensing*, 15(3):619–631, 1994.

[3] P Gong, JR Miller, J Freemantle, and B Chen. Spectral decomposition of Landsat Thematic Mapper data for urban land-cover mapping. In *Proceedings of the 14th Canadian Symposium on Remote Sensing*, pages 458–461, 1991. Calgary, Alberta, Canada.

[4] HM Horwitz, RF Nalepka, PD Hyde, and JP 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.

[5] RP Pech, AW Davies, RR Lamcroft, and RD Graetz. Calibration of LANDSAT data for sparsely vegetated semi-arid rangelands. *Int. J. Remote Sensing*, 7(12):1729–1750, 1986.

[6] NA Quarmby, JRG Townshend, JJ Settle, KH White, M Milnes, TL Hindle, and N Silleos. Linear mixture modelling applied to AVHRR data for crop area estimation. *Int. J. Remote Sensing*, 13(3):415–425, 1992.

[7] JJ Settle and NA Drake. Linear mixing and the estimation of ground cover proportions. *Int. J. Remote Sensing*, 14(6):1159–1177, 1993.

[8] G Strang. *Linear Algebra and Its Applications.* Academic Press, New York, 2nd edition, 1980.