Mixed Pixel Classification in Remote Sensing - Literature Survey

Technical Report VSSP–TR–1/2004

O Duran and M Petrou
School of Electronics and Physical Sciences,
University of Surrey,
Guildford, GU2 7XH, United Kingdom

February 23, 2004

Abstract

Surface patches imaged as individual pixels may contain more than one cover-type. However, even if different covers are present in a pixel, a single signature is recorded per pixel. Pixels containing more than one cover type are called *mixed pixels*, and those containing a single cover are called *pure pixels*. Mixed pixel classification is an important issue in industrial inspection processes and especially in remote sensing applications. The unmixing of such pixels involves finding the proportions of the pixel covered by each component or cover class. It is noted that different cover types may have distinctive spectral signatures. Taking advantage of the spectral information, spectral unmixing is performed. Over the last three decades, several models and methods to spectrally unmix pixels have been proposed. The best-known is the linear mixture modelling approach. That method and others are reviewed here.

1 Introduction

Remote sensing is defined as the extraction of information from the Earth's surface using images acquired from an overhead perspective, and electromagnetic radiation [15]. Depending on the type of illumination used, the sun or an artificial source, passive or active sensing is applied respectively.

When the radiation reaches the surface of the Earth, part of it is reflected and part of it is absorbed and reemitted (thermal energy). Figure 1 shows the reflectance spectra of two materials. This figure shows that the fraction of the irradiance that is reflected is dependent on the wavelength and differs more or less for each material [32]. Information about a scene can be extracted by measuring its amount of electromagnetic radiation reflected or emitted and comparing it with the spectral reflectance curves of known materials such as the ones shown in Figure 1. In order to measure the reflected and emitted radiation, usually imaging scanners aboard an aircraft or satellite are used. The resulting images are corrected and transformed to a digital image consisting of picture elements or pixels. The spatial resolution is defined by the instantaneous field of view (IFOV) of the sensor, that is defined as the geometrical projection of a single detector element at the scene under observation (Figure 2). This is approximately equivalent to the ground area represented by a pixel. Apart from space resolution, spectral and temporal resolutions are also defined. The temporal resolution is related to the interval between two successive recordings of the same scene. In case the scanner is carried by a satellite, the temporal resolution is determined by the satellite's orbit. The spectral resolution denotes the width of the wavelength interval at which the electromagnetic radiation is recorded [22].

Lately there has been great interest in civil as well as in military remote sensing applications for more detailed information about the materials contained within a pixel. Thus there have been significant improvements in the spectral breadth and resolution of remote sensing sensors. However, surfaces viewed by individual pixels may still contain more than one cover-type. Mixed pixels can occur when the resolution of the sensor is low, for instance for remote platforms performing wide-area surveillance. Moreover, even with sensors of relatively high resolution, mixed pixels are expected to be in the scene. For instance, mixed pixels can be a consequence of homogeneous combination of the materials, independently of the sensor resolution [19]. Figure 3 shows different situations that will produce mixed pixels:

- Sub-pixel objects: small objects, such as a house or a tree, are included in a pixel.

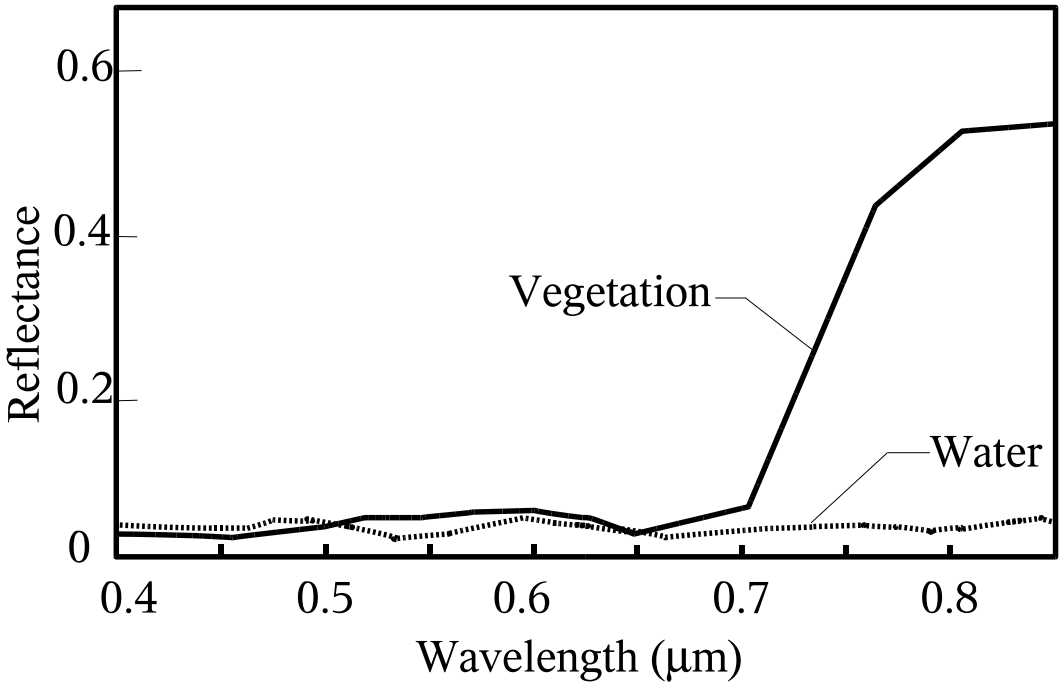

Figure 1: Spectral reflectance curves for different materials (adapted from [20]).

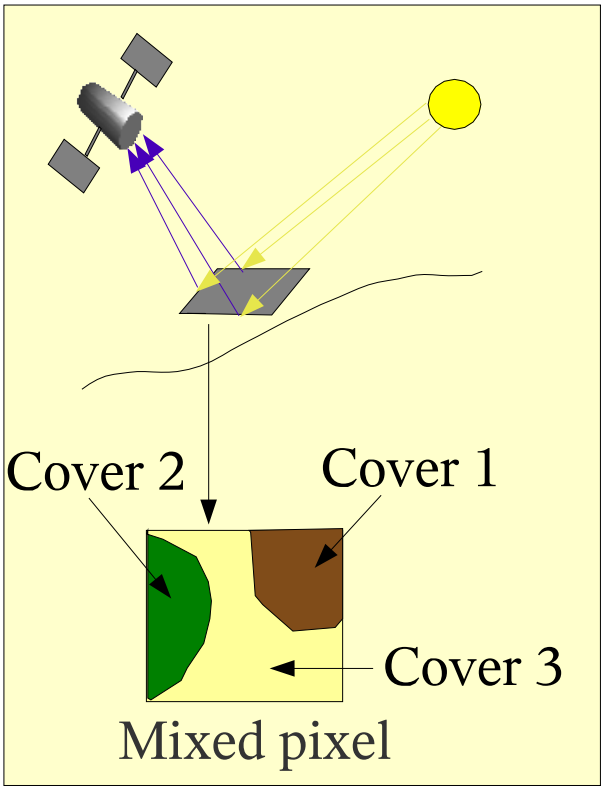

Figure 2: Instantaneous field of view of the sensor and resulting mixed pixel.

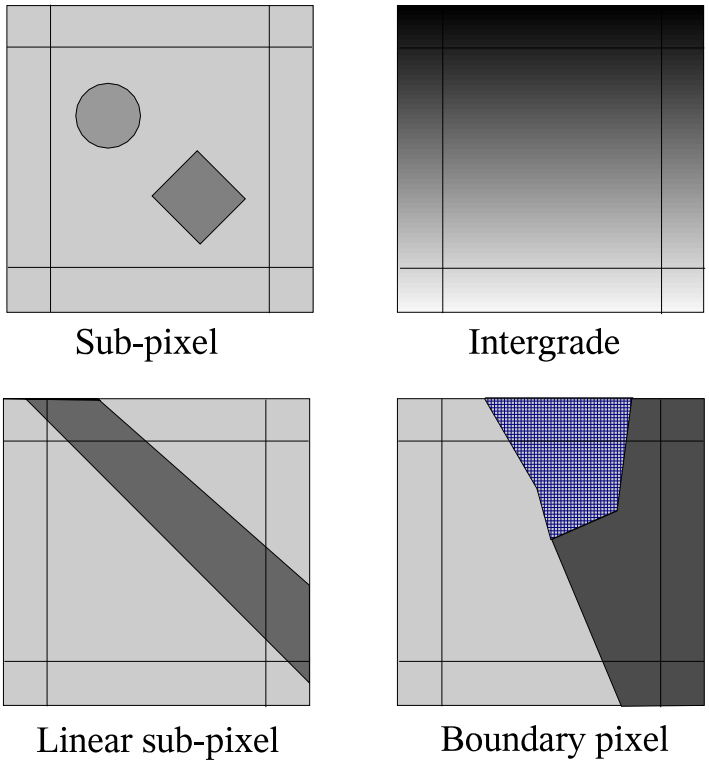

Figure 3: Mixed pixel types (adapted from [11])

- Intergrade: a transition zone between two or more different objects, such as transitions between vegetation types.
- Linear sub-pixel: a thin long structure, such as a road, partly covers several pixels.
- Boundary pixel: the boundary between two or more objects, for instance the boundary of different vegetation areas, passes through a pixel.

Spectral unmixing consists of decomposing the measured spectrum of a mixed pixel into a set of pure class spectra, or *endmembers*, and computing their corresponding fractions, or *abundances*. Pure cover classes correspond to usual components in the scene, such as water, soil, or metal. Unmixing algorithms were first investigated for studying the properties of chemical mixtures using reflectance spectroscopy in the early-mid nineties. This became an important industrial inspection tool. Reflectance spectroscopy and remote sensing merged in order to provide analysis of geophysical phenomena. In order to extract information from the remotely sensed data, first, physical models of terrestrial phenomena and the subtraction of atmospheric effects permitted passive multispectral and hyper-spectral radiance observations to be converted to reflectance values; second, models to represent reflectance spectra in mathematical terms and methods to extract the information had to be developed [19].

Two different directions have been taken in the spectral unmixing problem: the classification technique and the modelling of the relationship between the classes and the fractions of components in the pixel and the spectral response [36]. Several types of model have been proposed, such as the linear, probabilistic, geometric, physical and fuzzy as well as the use of Artificial Neural Networks (ANN). In the linear model, the reflectance of a pixel for a given band is computed as the linear combination of the pure-class reflectances weighted by their respective surface proportions [35] [1] [17] [18]. Probabilistic methods are based on maximum likelihood techniques [17] [24]. In the geometric models, the geometry of the objects in the scene, such as trees, are taken into account to evaluate the effects of shadows in pixels [18]. The physical model is a variant of the linear model that takes into account the correlations between pixels [14]. Fuzzy and ANN models employ, as their names show, fuzzy set theory and neural networks respectively [12][15].

Research of the different mixture models and the methods for unmixing the pixels has been carried out in parallel, the linear model being the most popular. However, it has often been criticised, mainly for its inadequacy to model backscatter and because it can only be used to unmix at most as many components as there are bands. Nonlinear mixing models incorporating backscatter effects are also researched, however the resulting models are more complex when compared with the linear ones.

Besides the models, many unmixing methods have been proposed that deal with the above mentioned models. Different classifications of methods exist. Keshaba and Mustard [19] classify the methods for unmixing pixels using the linear model in least squares methods (LMM) and Minimum Variance Methods (SVD). Gebbinck [15] differentiates between methods based on the linear model, and a selection of statistical-based methods. Besides those and other reviews of mixed pixel classification methods, there is a lack of comparative studies in the field of mixed pixel classification. Researchers use different data, in many cases synthetic data, to prove the capabilities and limitations of the methods they propose. This latter issue makes difficult the choice of one method for a particular application. However, a few comparative studies exist. Horwith *et al* [17] introduced and compared analytically some inversion methods for the linear model. Ichoku and Karnieli [18] conducted a comparative study of the following five models: linear, probabilistic, geometric, stochastic geometric, and fuzzy models for land cover estimation. They considered aspects such as model formulation, parameters, spectral dimensionality requirements, unmixing procedures, accuracy and applicability. Foody and Cox compared the linear and fuzzy models [12]. Besides those studies, a

detailed descriptive and comparative study has clearly become a need for the research, as well as for the end-user communities. Such a study would provide a better understanding of the available techniques where the strong and weak aspects would be emphasised, and would ease the choice of a model or method for a particular application.

2 Spectral unmixing

In this section and the subsequent subsections the stages for the unmixing process and the methods described in the literature will be reviewed. It is noted that the determination of the pure class members or endmembers and the estimation of the abundances are often considered as two separate problems. Thus, an examination of the literature reveals that many abundance estimators assume that the endmembers are already known. On the other hand, other researchers concentrate their efforts on the determination of the endmembers present in the scene and their spectra. Here, the two steps will be considered in different sections. Many methods have been developed, however there is a lack of a comparative study. Some researchers have reviewed some of the methods, but there is no benchmark. Most of the methods are tried either with synthetic data or with application-based data, making a fair comparison between them difficult. A summary of the unmixing methods is given in Table 1 with some comparative description. This table is further discussed in Section 3 where the models are rated and compared.

The two different stages of the unmixing, namely, the pure class selection and the unmixing, considered in the literature as two different problems are reviewed next.

2.1 Pure-class selection

The success of spectral unmixing depends on the selection of the endmember components. The objective of the algorithms described here is to determine the pure class members and their respective spectra. Selection of the pure-class spectra can be achieved in two ways:

- from a spectral library;
- from the pixels in the image.

The first way consists of using laboratory-derived reflectance values. This makes use of collections of reflectance spectra which have been put together in spectral libraries. The second approach relies

upon ground information. The fractional covers may be known for part of the images or they may be inferred, for instance by analysing a higher resolution image of part of the scene [35]. It is noted that spectral libraries represent specific cases and do not account for intraclass variability (Figure 8). Samples used as pure classes may have been obtained under different conditions or from different geographical areas than the one under inspection. Moreover, spectra of some substances change with physical parameters, such as chlorophyll content, water content, roughness, etc. Thus, for a quantitative study of spectral variability, a number of examples for each class should be included in a library [32].

Besides that classification, the selection can be static or dynamic. While static algorithms keep the same end-member subset for the whole scene, dynamic methods identify the optimum end-member subset for each pixel. Dynamic algorithms take into account the fact that the number of endmembers required to unmix an entire scene may be considerably greater than the number present in any individual pixel. Thus, each pixel in a scene may utilise a different subset of endmembers and those are calculated dynamically. One of the proposed methods performs unconstrained least squares inversion, discards those endmembers yielding negative values, and then performs the inversion without the discarded endmembers, repeating the entire process until all endmembers still present have nonnegative abundance values. To enforce the additivity condition, the sum of abundances is then normalised to one [33]. In other approaches, first the best pair of endmembers is found and additional endmembers are added to fit the model [25]. In the work proposed by Maselli *et al* [25], first the available end-members are normalised as follows:

$$v_i = \frac{a_i}{\| a_i \|} \tag{1}$$

where a_i is the endmember spectral signature and v_i is the corresponding normalised signature. Each candidate pixel is then projected onto all available normalised endmembers, and the largest projection coefficient, c_{max}, identifies the first selected end-member with spectral signature v_{max}. The spectral signal is then orthogonally subtracted from the spectral signature of the pixel, r, to yield the residual spectral signature, r_r, which is analysed further in the same way:

$$r_r = r - c_{max} v_{max} \tag{2}$$

The process may continue until a subset is identified or until the residual does not reduce its value significantly. The residual may become very small or negative even after the first iteration. In [23],

equation (2) is arbitrarily modified as follows, to avoid zero or negative values of r_r:

$$r_r = r - 0.5 c_{max} v_{max} \tag{3}$$

In that work, a method for selecting a subset of endmebers from a larger set of possible endmembers for sets of pixels was also proposed. The method exploits pixel neighbourhood information and the fact that generally, a mixed pixel is likely to have similar endmembers as its neighbouring mixed or pure pixels. For a set of pixels, votes are given to the endmembers depending on some neighbourhood rules. The endmembers are then sorted by votes and those with the most votes make up the subset.

Some problems arise in endmember selection algorithms. It is noted that to ensure the uniqueness of a solution to the linear mixture model, the set of endmembers must be linearly independent. Another constraint of the linear model limits the number of endmembers in a scene to be equal to the number of spectral bands plus one. Strategies to overcome this issue have been proposed [6] [10]. Moreover, materials of different physical composition may have similar spectral signatures or linear combinations of other component signatures (mixtures). Materials under different conditions may also exhibit different spectral responses [32]. This issue makes the selection of endmembers difficult if the computation relies on the spectral information only.

2.2 Statistical Linear Mixing Model

The linear mixing model was first proposed by Horwitz [17]. It is based on the physical assumption that there is no multiple scattering between the different cover types; each photon interacts with a single cover (Figure 4). Taking into account this condition, the energy received by the sensor can be considered as the sum of the energies received from each cover type.

Horwitz [17] assumed pixels containing small objects belonging to n different classes. The signature of component j, where $j = 1, \ldots, n$, is represented by an m-dimensional Gaussian distribution with mean m_j and covariance matrix N_j, with $i = 1, \ldots, m$ indexing the spectral bands. The proportions of the pure class j are defined as f_j and $f = (f_1, \ldots, f_n)^T$. Then the signature of this combination of classes is represented by mean $\mu(f)$ and covariance matrix $N(f)$. In order to find expressions for $\mu(f)$ and $N(f)$, let us consider a pixel containing k_j elements only of class j, represented by random variables with mean μ_j^* and variance-covariance matrix N_j^* [17]. Then, that pixel will contain k_j times stronger reflectance with variance k_j times the variance of one individual object

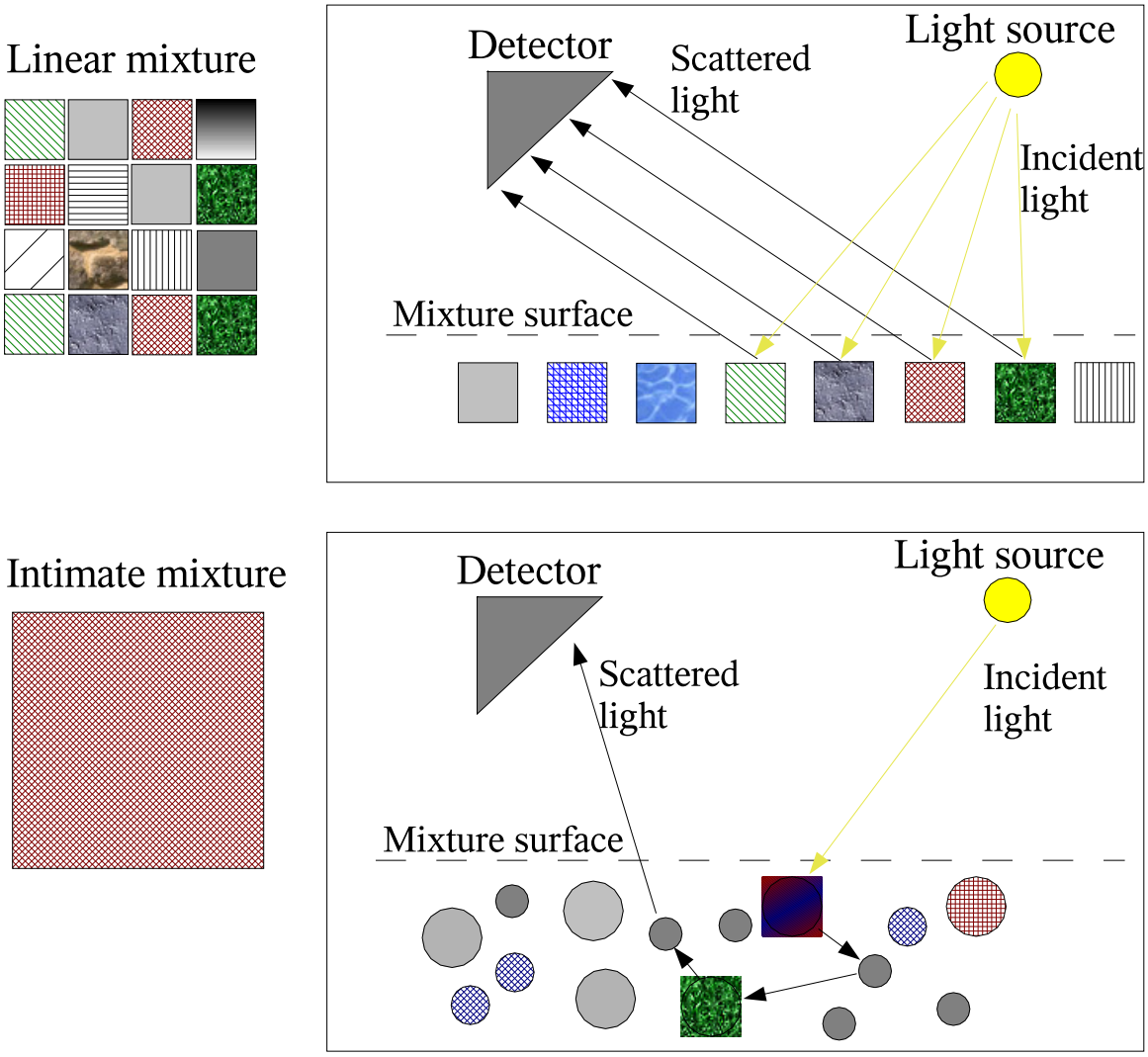

Figure 4: Linear mixture assumption. Photons interact will a single cover type (top), as opposed to an intimate mixture where the photons interact with more than one component (bottom). (Adapted from [29])

of that class (see appendix A),

$$\mu_j = k_j \mu_j^* \tag{4}$$

and

$$N_j = k_j N_j^* \tag{5}$$

The proportions of the classes f are then considered following the same reasoning as above and consequently if the number of elements of class j is $f_j k_j$ instead of k_j, the composite signature is,

$$\mu(f) = \sum_{j=1}^{m} f_j k_j \mu_j^* = \sum_{j=1}^{m} f_j \mu_j \tag{6}$$

and

$$N(f) = \sum_{j=1}^{m} f_j k_j N_j^* = \sum_{j=1}^{m} f_j N_j \tag{7}$$

Equations (6) and (7) constitute the model that has been followed by many researchers and repeated in many papers [17] [35].

2.3 Decomposition based on the linear model

According to the model presented in the previous section, each cover in the scene will contribute to the received signal with a signal characteristic of the cover type and proportional to the area covered. The reflectance value of a mixed pixel in band i can then be given by the following formula:

$$r_i = \sum_{j=1}^{n} a_{ij} f_j + e_i \tag{8}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, where a_{ij} denotes the reflectance of the jth component of the pixel in the ith spectral band and e_i is the error term in the ith spectral band.

It is noted that (8) is a system of m linear equations that can be expressed in matrix form, as follows:

$$r = Af + e \tag{9}$$

where r is $(r_1, r_2, \ldots, r_m)^T$, A is an $m \times n$ matrix containing the reflectances of the n components in the m spectral bands, f is the vector of proportions $(f_1, f_2, \ldots, f_n)^T$, and e the vector of errors, equal to $(e_1, e_2, \ldots, e_m)^T$, representing the statistical fluctuations around the linearly predicted value Af. The vector of errors e satisfies $E(e) = 0$ and $E(ee^T) = N$, where E denotes the expectation operator. A pixel with a mixture f, will then fluctuate around the mean value Af, characterised by a covariance matrix N, that is assumed to be independent of f [35].

Equations (8) and (9) can be solved in several ways to determine the proportions f. Since the proportions have to sum to unity, a linear constraint, $f_1 + f_2 + f_3 \ldots + f_n = 1$ can be included in the system of equations, as follows:

$$\begin{array}{c} a_{11}f_1 + a_{12}f_2 + \cdots + a_{1n}f_n = r_1 \\ a_{21}f_1 + a_{22}f_2 + \cdots + a_{2n}f_n = r_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}f_1 + a_{m2}f_2 + \cdots + a_{mn}f_n = r_m \\ f_1 + f_2 + \cdots + f_n = 1 \end{array}$$

Generally the number of unknowns or number of components, n, has to be less than or equal to the number of equations, here given by the number of bands m, or $m+1$ if the linear constraint is considered. If n is less than m (or $m+1$), the system of equations is overdetermined and it can be solved by the least square error method. Several least square error techniques have been proposed [36] [35]. Summarising,

if $n = m + 1$, there are as many linear equations as unknowns and if one solution exists, it will be unique;

if $n \leq m + 1$, there are more equations than unknowns and the system may be solved in the least square error sense;

if $n \geq m + 1$, there are fewer equations than unknowns and there is an infinite number of solutions, from which there is not a clear way to define the best among them.

A second constraint, apart from the sum-to-unity, that should be satisfied, is the positivity constraint, where no component of a mixed pixel can make a negative contribution:

$$f_j \geq 0 \quad \text{for} \quad j = 1, \ldots, n \tag{10}$$

Satisfaction of the latter constraint is often difficult and may require some specialised techniques (see Section 2.3.3.). The mixing equations alongside the two constraints describe a model that must be solved for each mixed pixel that is to be decomposed, i.e. given r and A, f is to be determined, subject to the constraints [35] [15].

Next, the methods for model inversion or decomposition found in the literature for the linear model will be described.

2.3.1 Linear spectral unmixing for single pixels

For the mixed pixel problem, the following assumptions are made:

- A pixel contains several objects;
- an endmember is the reflectance of a pixel that contains objects of a single class j;
- the spectrum of an endmember is drawn from an m-dimensional normal distribution, where m is the number of bands;
- there is no correlation between the bands of the sensor.

The spectral reflectance of a mixed pixel with a mixture f, since it is described by the linear model, is also considered a random variable, normally distributed, with mean value Af, and a covariance matrix N.

Maximum likelihood estimation is a supervised classification technique. This technique assumes that the probability density function of the spectral distribution of the mixed pixel can be represented by a multivariate normal distribution, with mean vector Af and a variance-covariance matrix N, where A contains the endmember spectra and f the mixing proportions.
The classification of a pixel of unknown composition consists of finding the likelihood with which a pixel has a given set of proportions. The decision is based on maximising the joint probability density function, $p(r, f)$, of the reflectance r and proportions f [20]. According to Bayes's theorem, the probability of r and f jointly occurring is equal to the probability that r occurs after knowing that f has occurred or to the probability of f occurring after knowing that r has occurred,

$$p(r, f) = p(r|f)p(f) = p(f|r)p(r) \tag{11}$$

The conditional probability density function $p(r|f)$ in the case of a multivariate normal distributions is given by:

$$p(r|f) = \frac{1}{(2\pi)^{\frac{n}{2}}|N|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(r - Af)^T N^{-1}(r - Af)\right) \tag{12}$$

The objective is to find the proportions that maximise $p(r, f)$,

$$p(r, f) = \frac{1}{(2\pi)^{\frac{n}{2}}|N|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(r - Af)^T N^{-1}(r - Af)\right) p(f) \tag{13}$$

where $p(f)$ is the prior probability density of the mixing proportions f. Often, instead of maximising the above equation, its logarithm is maximised, or the negative logarithm is minimised:

$$\tilde{Q} \equiv \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln|N| + \frac{1}{2}(r-Af)^T N^{-1}(r-Af) - \ln(p(f)) \tag{14}$$

If we do not have any prior knowledge about the relative abundance of each class in the scene, we may assume that all classes are equally likely and set $p(f)$ equal to a constant. Ignoring the constants in (14) and multiplying by 2, we define,

$$Q \equiv \ln|N| + (r-Af)^T N^{-1}(r-Af) = \ln|N| + e^T N^{-1} e \tag{15}$$

where we made use of equation (9). The term $(r-Af)^T N^{-1}(r-Af)$ is the Mahalanobis distance between r and Af. If it is assumed that N is independent of f and equal to the common noise matrix N, then the term $\ln|N|$ becomes also a constant and can be omitted from equation (15). The resulting estimator would be simply the Mahalanobis distance that expresses the distance of the observed r from Af. Figure 5 shows schematically this approach for two bands.

The brute force approach is the simplest method to minimise Q [17]. It consists of a search through the entire solution space. For a specific band, every f that satisfies the sum-to-one and positivity constraint, the corresponding error vector e and variance-covariance matrix N are calculated, as well as the criterion used for maximum likelihood classification (see equation (15)) [17]. The f that minimises this cost function is taken as the solution. This approach is high in computational complexity since the maximum likelihood criterion has to be evaluated in the order of c^{n-1} times, where c is the step size. Furthermore, it leads to an approximation of the optimal f because the solution space is searched taking discrete steps. By varying the sampling rate, the accuracy of the approximation can be traded against computational costs [15].

If the term $\ln|N|$ is ignored in equation (15), the least square error solution is obtained. Let us consider the quantity we have to minimise in this case:

$$\begin{aligned}
\hat{Q} &\equiv (r-Af)^T N^{-1}(r-Af) \\
&= (r^T - f^T A^T)(N^{-1}r - N^{-1}Af) \\
&= r^T N^{-1} r - r^T N^{-1} Af - f^T A^T N^{-1} r + f^T A^T N^{-1} Af
\end{aligned} \tag{16}$$

We must take the first derivative of $\hat{Q}$ with respect to f and set it to zero. To differentiate $\hat{Q}$ with respect to a vector, we apply the following formulae [30]:

$$\frac{\partial f^T a}{\partial f} = a \tag{17}$$

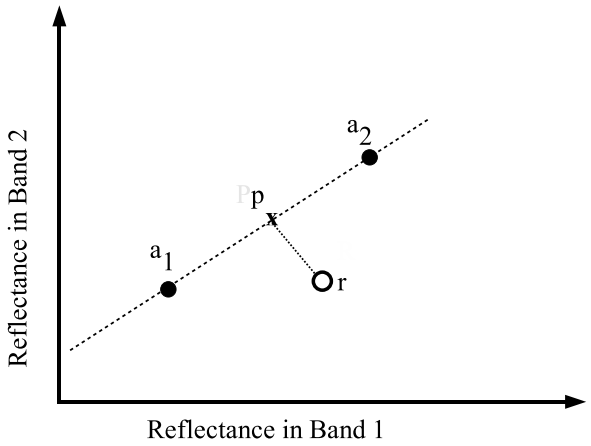

Figure 5: Maximum likelihood solution of the statistical linear mixture model. In this two-dimensional feature space, linear combinations of endmembers a_1 and a_2 satisfying the sum-to-one constraint occupy the dashed line; those that satisfy the positivity constraint as well lie on $a_1 a_2$ segment. To find the optimal solution, point p is moved along $a_1 a_2$ until the Mahalanobis distance from point r to p (plus a log-term) is minimal (from [15])

$$\frac{\partial b^T f}{\partial f} = b \tag{18}$$

$$\frac{\partial f^T U f}{\partial f} = (U + U^T) f \tag{19}$$

where f, a and b are vectors and U is a matrix. We apply these formulae in order to differentiate $\hat{Q}$ with respect to f, using the following correspondences:

$$A^T N^{-1} r \leftrightarrow a \tag{20}$$

$$r^T N^{-1} A \leftrightarrow b^T \tag{21}$$

$$A^T N^{-1} A \leftrightarrow U \tag{22}$$

The first derivative of $\hat{Q}$ with respect to f then is,

$$\begin{aligned} \frac{\partial \hat{Q}}{\partial f} &= -(r^T N^{-1} A)^T - A^T N^{-1} r + (A^T N^{-1} A + (A^T N^{-1} A)^T) f \\ &= -A^T N^{-1} r - A^T N^{-1} r + (A^T N^{-1} A + A^T N^{-1} A) f \\ &= -2 A^T N^{-1} r + 2 A^T N^{-1} A f \end{aligned} \tag{23}$$

where we set $(N^{-1})^T = N^{-1}$ because N is a covariance matrix and therefore symmetric. Setting $\frac{\partial \hat{Q}}{\partial f} = 0$ results in the following solution for f:

$$\hat{f} = (A^T N^{-1} A)^{-1} A^T N^{-1} r \tag{24}$$

If N is replaced by the identity matrix, as suggested in [35],

$$\hat{f} = (A^T A)^{-1} A^T r \tag{25}$$

Equation (25) is the classical least squares estimator, since it is the standard expression for least squares approximations for the system of equations $r = Af$ [35].

The last term in equation (14) can be exploited in order to introduce constraints to the solution. We may for instance introduce the sum to 1 constraint as follows [35]: We may say that the prior probability $p(f)$ is given by

$$p(f) = \frac{1}{C} e^{-\alpha(1-\sum_j f_j)^2} \tag{26}$$

where α and C are some constants. This probability density function favours solutions of f which satisfy the sum to 1 constraint. Omitting constant terms, the quantity we then have to minimise is,

$$\hat{\hat{Q}} \equiv \frac{1}{2}(r - Af)^T N^{-1}(r - Af) + \alpha \left(1 - \sum_j f_j\right)^2 \tag{27}$$

which becomes

$$\begin{aligned} \hat{\hat{Q}} &= r^T N^{-1} r - r^T N^{-1} A f - f^T A^T N^{-1} r + f^T A^T N^{-1} A f + \alpha \left(1 - \sum_j f_j\right)^2 \\ &= \hat{Q} + \alpha \left(1 - \sum_j f_j\right)^2 \end{aligned} \tag{28}$$

Differentiating $\hat{\hat{Q}}$ with respect to each of the components of vector f will produce $\frac{\partial \hat{Q}}{\partial f}$ plus a term which will be identical for all components and equal to $-2\alpha(1 - \sum_j f_j)$. We define a vector k with all its elements equal to 1, and write,

$$\begin{aligned} \frac{\partial \hat{\hat{Q}}}{\partial f} &= \frac{\partial \hat{Q}}{\partial f} - 2\alpha \left(1 - \sum_j f_j\right) k \\ &= -2A^T N^{-1} r + 2A^T N^{-1} A f - 2\alpha \left(1 - \sum_j f_j\right) k \end{aligned} \tag{29}$$

Setting this to 0 and solving for f we obtain the solution with the sum to 1 constraint incorporated:

$$\hat{\hat{f}} = (A^T N^{-1} A)^{-1} A^T N^{-1} r + \alpha \left(1 - \sum_j f_j\right) (A^T N^{-1} A)^{-1} k \tag{30}$$

Note that this is not yet the full solution as the unknowns f_j appear on the right-hand-side of equation (30). Note also that $\sum_j f_j$ can be written as $k^T f$. Assuming then that the sum to 1

constraint will simply add a perturbation to the solution obtained without the constraint (given by $\hat{f}$ of equation(24)), we may substitute f on the right hand side of (30) by $\hat{f}$:

$$\hat{\hat{f}} = (A^T N^{-1} A)^{-1} A^T N^{-1} r + \alpha(1 - k^T (A^T N^{-1} A)^{-1} A^T N^{-1} r)(A^T N^{-1} A)^{-1} k \tag{31}$$

Settle and Drake in [35] showed that α should be set to $(k^T (A^T N^{-1} A)^{-1} k)^{-1}$. It is noted that for a given matrix U, $k^T U k$ is the sum of all the elements of the matrix. Then, α is set equal to the inverse of the sum of the elements of matrix $(A^T N^{-1} A)^{-1}$. Matrix $A^T N^{-1} A$ gives the square brightnesses of all pure classes in all bands inversely weighted by the spread in each band. The inverse of this, grossly gives us the uncertainty in each input value of matrix A. Summing all elements then of matrix $(A^T N^{-1} A)^{-1}$ gives us the total uncertainty of the values we use to estimate f. It is obvious that the sum to 1 constraint can not be enforced more strictly than the uncertainty we have, so α, being the inverse of the total uncertainty gives more weight to the constraint as the uncertainty is reduced and less weight as the uncertainty grows.

Satisfaction of the positivity constraint cannot be achieved with a similar transformation, but requires some post-processing of the initial estimate or an alternative approach to solve the model's equations [15].

Settle and Darke proposed also the regularised estimator [35]. This uses the $\ln p(f)$ term in equation (14) to introduce a regularisation effect in the obtained solution. The method is a modification of the classical estimator that aims to take into account the noise contained in each pixel. The individual estimates are directed towards some favoured mixture g, provided that the noise level in the overall estimate remains the same. Let us assume that the prior probability in (14) is

$$p(f) = \frac{1}{C} e^{-\lambda \sum_j (f_j - g_j)^2} \tag{32}$$

where C and λ are some constants. We may replace $\sum_j (f_j - g_j)^2$ by $(f-g)^T (f-g)$ and so write for the quantity to be minimised,

$$\begin{aligned}
\hat{\hat{Q}} &\equiv \hat{Q} + \lambda (f-g)^T (f-g) \\
&= (r - Af)^T N^{-1} (r - Af) + \lambda (f^T - g^T)(f - g) \\
&= r^T N^{-1} r - r^T N^{-1} A f - f^T A^T N^{-1} r + f^T A^T N^{-1} A f + \lambda (f^T I f - f^T g - g^T f + g^T g)
\end{aligned} \tag{33}$$

where I is the identity matrix. By differentiation of the above expression using formulae (17), (18) and (19),we obtain:

$$\begin{aligned}\frac{\partial \hat{\hat{Q}}}{\partial f} &= -2A^T N^{-1} r + (2A^T N^{-1} A) f + \lambda(2If - 2g) \\ &= 2(\lambda I + A^T N^{-1} A) f - 2\lambda g - 2A^T N^{-1} r \end{aligned} \tag{34}$$

Setting this to zero, we obtain the solution:

$$\hat{\hat{f}} = (\lambda I + A^T N^{-1} A)^{-1} (\lambda g + A^T N^{-1} r) \tag{35}$$

In [35], it was proven that if g is set equal to the proportions of the averaged signal of the pixels in a region, the optimum value of λ can be determined.

Gebbinck and Schouten [14] presented an alternative solution to the statistical linear model, based on a physical model. In the proposed model, the noise is attributed to variations in physical factors such as humidity, soil type and elevation instead of statistical noise as it is done in the standard approach. Since the scale of the variations in these physical factors usually is larger than the size of a single pixel, correlations between neighbouring pixels are likely to be found and therefore values of neighbouring pixels are not statistically independent. According to this, equations of the mixing model take the form:

$$r = (A + E) f \tag{36}$$

where, as before, r is the m-dimensional pixel reflectance, f is the proportions vector and A is the matrix whose columns contain the endmember spectra. For a given pixel, the error $m \times n$ matrix E contains the endmember spectra deviations due to physical factors [15]. Figure 6 shows a schematic representation of the mixing process according to the physical model, showing a mixture of two classes. This representation can be compared with the representation in figure 5 to see the difference with the statistical model. Equation (36) can be rewritten explicitly as follows:

$$\begin{aligned} r_1 &= (a_{11} + e_{11}) f_1 + (a_{12} + e_{12}) f_2 + \cdots + (a_{1n} + e_{1n}) f_n \\ r_2 &= (a_{21} + e_{21}) f_1 + (a_{22} + e_{22}) f_2 + \cdots + (a_{2n} + e_{2n}) f_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_m &= (a_{m1} + e_{m1}) f_1 + (a_{m2} + e_{)} f_2 + \cdots + (a_{mn} + e_{mn}) f_n \end{aligned}$$

The criterion used in [15] to determine the optimal solution, is that the deviations from the endmember spectra should be minimised. It is assumed that cover class j has a multivariate normal

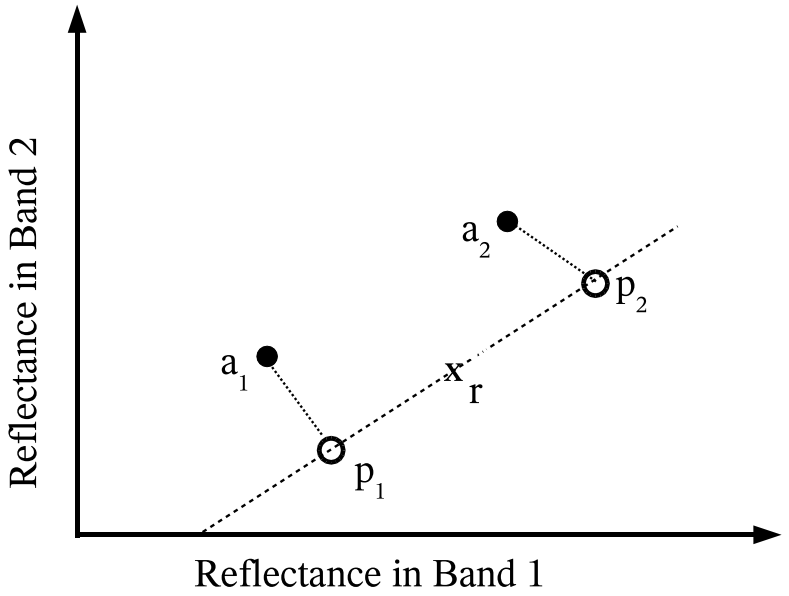

Figure 6: Mixing process according to the physical linear mixture model. In this two-dimensional feature space, a mixture of two classes with global endmembers a_1 and a_2 is shown. Pixel r is a linear combination of the component spectra p_1 and p_2 that deviate from a_1 and a_2 respectively, due to local variations in physical factors (from [15]).

distribution with mean a_j and covariance matrix N_j. The values therefore $e_j \equiv (e_{1j}, e_{2j}, \ldots, e_{mj})^T$ which have to be added to its mean reflectance values $a_j = (a_{1j}, a_{2j}, \ldots, a_{mj})^T$ to produce values with which pure class j contributes to the reflectance of the mixed pixel, are drawn from the probability density function,

$$p(e_j) = \frac{1}{(2\pi)^{\frac{m}{2}} |N_j|^{\frac{1}{2}}} e^{-e_j^T N_j^{-1} e_j} \tag{37}$$

Assuming that the pure classes are independent, the joint probability density function of all e_{ij} values to arise is

$$p(E) = p(e_1)p(e_2)\ldots p(e_n) = \frac{1}{(2\pi)^{\frac{mn}{2}} \prod_{j=1}^{n} |N_j|^{\frac{1}{2}}} e^{-\sum_{j=1}^{n} e_j^T N_j^{-1} e_j} \tag{38}$$

We wish to choose values f so that the combination of correction values e_{ij} we need is as probable as possible, i.e. we wish to choose f so that $P(E)$ is maximised. At the same time we wish to choose solutions which have highest probability to satisfy equation (36) and highest probability to satisfy the sum to one constraint. This implies that we wish to choose solutions which maximise the probability density function

$$p(f) = \frac{1}{C} e^{\left(-2\alpha^T (r-(A+E)f) - 2\beta(k^T f - 1)\right)} \tag{39}$$

where α is a vector of constants and C and β are some constants. Combining these requirements, means that we are seeking solutions which maximise the joint probability density function,

$$p(E, f) = p(E)p(f) \tag{40}$$

This leads to the minimisation of

$$\begin{aligned} \tilde{\tilde{Q}} &\equiv \sum_{j=1}^{n} e_j^T N_j^{-1} e_j + 2\alpha^T (r - (A+E)f) + 2\beta(k^T f - 1) \\ &= trace(E^T N^{-1} E) + 2\alpha^T (r - (A+E)f) + 2\beta(k^T f - 1) \end{aligned} \tag{41}$$

Expression (41) is then differentiated with respect to f and E and the derivatives are set to zero. The resulting equations are combined to eliminate one of the constants, which results in the following expression,

$$(A - rk^T)^T N^{-1} (A - rk^T) f = f(\alpha^T N \alpha)(f^T f) \tag{42}$$

This is a $Bf = \lambda f$ system, with $B \equiv (A - rk^T)^T N^{-1} (A - rk^T)$ and $\lambda \equiv (\alpha^T N \alpha)(f^T f)$. The solution for f is one of the eigenvectors of matrix B. Gebbinck [15] suggests to select the eigenvector which corresponds to the smallest eigenvalue as the optimum solution, since it can be shown that that solution minimises $\tilde{\tilde{Q}}$.

2.3.2 Unmixing sets of mixed pixels

Classification of mixed pixels can be difficult due to the small number of spectral bands that are available. Sometimes the number of spectral bands can be less than the number of pure classes to be identified [16].

Bosdogianni *et al* [6] address the problem of mixed pixel classification when whole regions of mixed pixels have to be classified. They treated the observed reflectance value of a mixed pixel in a particular band as a random variable, and the collection of the mixed pixels in the region as the instantiations of this random variable. They also treated the reflectance values of the pure classes as random variables drawn from some distributions. This interpretation of the mixing equations implies that the moments of the distribution of the mixed pixels can be expressed in terms of the moments of the distributions of the pure classes, which are assumed independent. Then two sets of equations that relate the means and second order moments of the pure and mixed sets or random variables may be written:

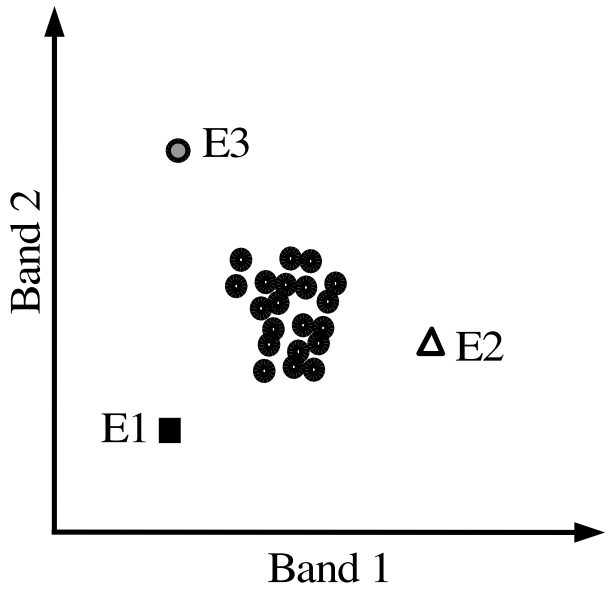

Figure 7: In the classical approach the pure classes are assumed to have fixed known spectra and each pixel represents a different linear combination of them.

$$\hat{r}_i = \sum_{j=1}^{n} \hat{a}_{ij} f_j \tag{43}$$

where $\hat{r}_i$ is the mean spectral density of the set of mixed pixels in the ith spectral band, $\hat{a}_{ij}$ is the mean spectral intensity of the set of pure pixels of the jth class in the ith spectral band, f_j is the proportion of the jth class, assumed the same for the whole set of mixed pixels, and n is the number of pure classes. The sum to 1 constraint may also be used to supplement equation (43). A second equation to express the relationship of the second order moments is:

$$N_{ik} = \sum_{j=1}^{n} f_j^2 N_{jik} \tag{44}$$

where N_{ik} is the ik element of the covariance matrix of the set of mixed pixels and N_{jik} is the ik element of the covariance matrix of the set of the jth pure class pixels (see appendix A). In this model, all pixels in the set are considered to have identical proportions. The error is only due to intraclass variability (Figures 7 and 8) [32] [28], while in the classical model the error is due to the variability of the individual proportions that contribute to a single pixel.

Bosdogianni *et al* combined equations (43) and (44) in a single set of equations and proposed two approaches for solving it, by least square error and using robust methods [6] [5].

Because the error with which the components entering equations (43) and (44) are computed is different, the total error, which had to be minimised for the least square error solution of this system

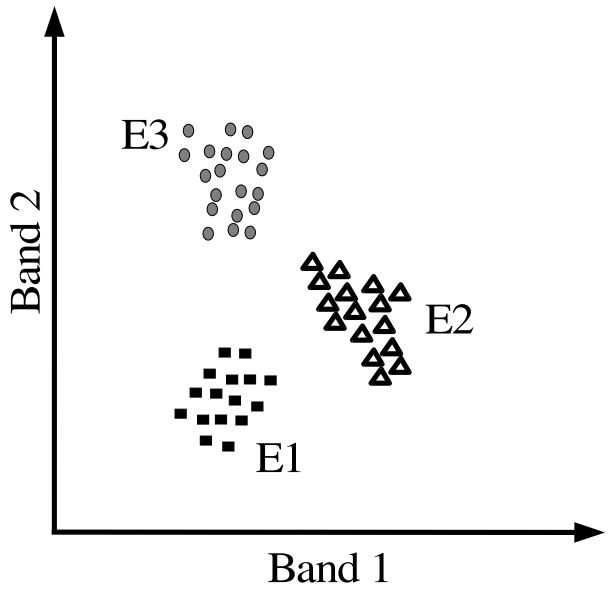

Figure 8: Intraclass variability of three pure classes (E1, E2 and E3). Due to interclass variability, identical mixtures may lead to a variety of observed mixed spectra by creating linear combinations of different pure class pixels.

of equations, was computed by weighting the error of each equation inversely proportionally by the standard error with which mean and variances can be estimated. The standard error for estimating the mean is given by $\frac{N_{ii}}{\sqrt{p}}$ where N_{ii} is the standard deviation for band i, as calculated from the data and p is the number of samples used to represent the given distribution. The standard error for the co-variance is given by $\frac{\sqrt{2}N_{ij}}{\sqrt{p}}$. So the total error function that had to be minimised was defined as

$$\sum_{i=1}^{n}\left(\sqrt{\frac{p}{N_{ii}}}\left(\hat{r}_i-\sum_{j=1}^{m}f_j a_{ij}\right)\right)^2+\sum_{i=1}^{n}\sum_{k=1}^{n}\left(\frac{\sqrt{p}}{\sqrt{2}N_{ik}}\left(N_{ik}-\sum_{j=1}^{m}f_j^2 N_{j_{ik}}\right)\right)^2 \tag{45}$$

The minimum of expression (45) can be found by setting its partial derivatives with respect to each f_i to zero and solving the resulting equations. The sum-to-one constraint may also be included in the set of equations. The main advantage of this model is that it still keeps the same number of unknowns, but the number equations is increased, which means that the region can be unmixed into more classes than the $m+1$ limit of the classic model [10]. It is useful for the cases when the exact mixing proportions of individual pixels are of no interest, but instead one is interested in identifying the mixing proportions which describe on average the whole set of mixed pixels. Such is the case, for example of a burned forest being regenerated and one is interesting in estimating the average cover of new trees over the whole region.

The second approach proposed by Bosdogianni *et al* is based on a Hough transform decomposition

[5]. The Hough Transform (HT) is an image analysis tool, where the data are mapped into a suitable parameter space which is then searched for the most likely parameter values. The HT has been widely used for finding straight lines hidden in images. The method proposed in [4] and [5], uses the Hough transform by parametrising the linear mixing equation in the parametric $f_1, f_2, \ldots, f_n$ space. An $(n-1)$D accumulator array is defined in the parametric $f_1, f_2, \ldots, f_n$ space. For each set of $n+1$ values $(a_{i1}, a_{i2}, \ldots, a_{in}, r_i)$, a different hyperplane is defined in the $f_1, f_2, \ldots, f_n$ space. The cells in the accumulator array that are intersected by the hyperplane are incremented by one. This is repeated until all possible sets of values $(a_{i1}, a_{i2}, \ldots, a_{in}, r_i)$ have been considered. The highest peak in the accumulator defines the best values for the mixing parameters $f_1, f_2, \ldots, f_n$.

The Hough transform has the advantage of being robust against outliers. Moreover, with the Hough transform, coherent outliers can also be identified as secondary peaks in the parameter space. However, some disadvantages also arise: The parameter space size is proportional to the power of the number of parameters. Thus, the method is limited to a few dimensions. It is also noted that the method depends on the detection of the peak, that may require filtering and thresholding to eliminate insignificant peaks. Another factor that may influence the accuracy of the detection is the quantisation of the parameter space. A wrong bin size may influence the likelihood of finding a peak [34]. This problem may be overcomed using the hypothesis testing transform [4]. In order to speed-up the process, a hierarchical two-stage algorithm may be used [8]. In a first stage a coarse grid is used to determine the approximated proportions. A new accumulator space smaller but with higher resolution is created for the second stage. The true solution may then be found by using hill-climbing with sub-bin accuracy. The Randomised Hough Transform (RHT) [7] was proposed to deal with the computational cost of using the Hough transform for the case of many pure classes, i.e. a large accumulator array. The RHT differs from the standard HT in the vote accumulation process, instead of analysing the parameter space at the end of the process, it is done dynamically. The data sets are randomly sampled to choose $n-1$ sets of values $(a_{i1}, a_{i2}, \ldots, a_{in}, r_i)$ at a time. The random sampling continues until an evident candidate for a global peak is detected. A fixed threshold may be chosen to detect the peak [7]. It was also proposed in [7] to use a variable threshold where the threshold level is adjusted according to the input data. Alternative robust methods have also been proposed. For example, Bosdogianni *et al* in [5] proposed the use of trimmed means as a robust statistic and Rosin in [34] proposed the least median square solution (LMeds). The Trimmed Means

(TM) algorithm proposed in [5] is as follows,

1) Select the set of mixed pixels.

2) Determine the mixing proportions from each pixel in the set.

3) Identify the mixing proportions representing the highest number of pixels.

4) Compute the trimmed mean of the mixing proportions around the maximum values found under (3), within arbitrarily chosen intervals, discarding the outliers. In [5], the trimmed mean was computed using a window in the accumulator (parameter) space with size equal to the proportions uncertainties.

The Lmeds is an estimator that instead of minimising the usual squared residuals, it minimises the median of the squared residuals. The algorithm proposed in [34] is as follows:

1) Randomly sample pixel tuples.

2) Determine the mixing proportions from each tuple.

3) Select the mixing proportions that minimise the median error over the complete dataset.

4) Identify outliers.

5) Re-determine the mixing proportions from the inliers.

Rosin [34] used linear mixing with the endmember fractions estimated by simple LS to provide local estimates that were fed into the LMedS method as described above.

2.4 Other Unmixing Methods

Apart from the linear mixture modelling, many other techniques have been proposed. Some of those approaches are based on nonlinear reflectance models. Those methods will not be reviewed here. Most of the methods are based on physical models of the process that takes place during image formation.

2.4.1 Vegetation Index approach

This approach exploits the spectral information that can be associated with the physical characteristics of materials. The method is based on the reflectance properties of natural land areas and the compression of the original spectral information into a smaller number of features. Addition, subtraction and division of the pixel radiance of two bands define the vegetation indices. The vegetation index relates to the near-infrared/red spectrum. A simple vegetation index is computed as follows,

$$VI = NIR - R \tag{46}$$

where R represents the radiance or reflectance reaching the sensor in the visible red part of the spectrum, and NIR in the near infrared part of the spectrum. Since green vegetation exhibits high reflectance in the NIR band and low in the red band, while soil has similar reflectance in the two bands, areas with vegetation will have high vegetation index while soil will produce indices close to zero.

One of the most commonly used measures are the *Normalised difference vegetation index* ($NDVI$),

$$NVDI = \frac{Near\ \ Infrared\ \ band\ \ -\ \ Red\ \ band}{Near\ \ Infrared\ \ band\ \ +\ \ Red\ \ band} \tag{47}$$

and the *Perpendicular vegetation index* (PVI):

$$PVI = \sqrt{(soil\ \ in\ \ NIR -\ \ point\ \ in\ \ NIR)^2 + (soil\ \ in\ \ the\ \ R - point\ \ in\ \ the\ \ R)^2} \tag{48}$$

It was noticed that soil pixels orient themselves along a *soil line* in the 2D space defined by the two spectral bands (NIR and Red). Vegetation pixels of a particular cover type or with the same percentage of vegetation, lie on lines parallel to the soil line. Such lines of constant vegetation divide the plot area into regions that can be associated with percentages of vegetation contributing to a mixed pixel. A measure of the distance of a point from the soil line is used as an index of the vegetation. The PVI defines the geometrical distance of a point from the line of soil; as the vegetation increases, the PVI increases [38] [39].

The vegetation index method works well in many vegetation types. However, problems arise when soil brightness changes from one area to another. The cover line is then used with more accuracy to detect changes over time (temporal vegetation index) rather than determining the absolute cover across a range of soils. The temporal approach will be further discussed in Section 3.2. In terms of classification performance, the vegetation index-based approaches were found inferior to spectral unmixing [38] [39].

2.4.2 Fuzzy approaches

Another method proposed is based on fuzzy sets theory [3]. Here, the algorithm partitions the pixels to be classified into C-fuzzy sets or classes. A nondegenerate C-fuzzy set U of p pixels is defined as follows,

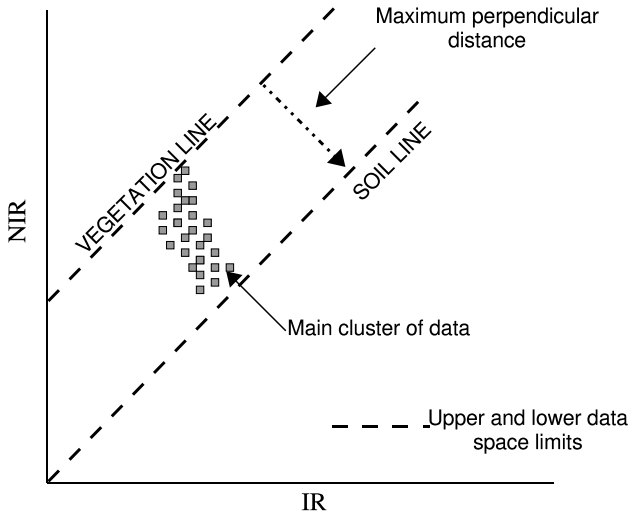

Figure 9: Vegetation index (adapted from [31])

$$U = \left\{ \mu_{ik} \mid \mu_{ik} \in [0,1]; \sum_{k=1}^{p} \mu_{ik} > 0, i = 1, \ldots, n; \sum_{i=1}^{n} \mu_{ik} = 1, k = 1, \ldots, p \right\} \tag{49}$$

where U is a fuzzy C-partition of n fuzzy groups and p observations and μ_{ik} are the membership grades representing the membership of an observation r_k to the ith class. The number of classes is to be known a priori and every class has to be represented in the set of pixels (The constraint $\sum_{k=1}^{p} \mu_{ik} > 0, \quad i = 1, \ldots, n;$ ensures that there is no cluster with no members).

Several techniques were proposed to find the optimal fuzzy C-partitions [3]. The most popular consists of minimising the generalised least-squared errors functional, $J(U, A)$,

$$J(U, A) = \sum_{k=1}^{p} \sum_{i=1}^{n} (\mu_{ik})^q d(r_k, a_i)^2 \tag{50}$$

where A is a $n \times m$ matrix, the rows, $a_i = a_{i1}, a_{i2}, \ldots, a_{im}$, of which represent the mean or centre of the ith class for each of the m bands; $d(r_k, a_i)^2$ is the Mahalanobis distance between the kth pixel and the class vector centroid a_i, and q is a weighting parameter that indicates the degree of fuzziness [3]. $J(U, A)$ is then a weighted measure of the squared distance, $d(r_k, a_i)^2$, that can be determined as follows,

$$d(r_k, a_i)^2 = (r_k - a_i)^T N^{-1} (r_k - a_i) \tag{51}$$

where N is is the covariance matrix of the data set.

The algorithm to minimise J is described in [3]. The optimal fuzzy clusterings are defined as pairs

$(\hat{U}, \hat{A})$ that minimise J. However, one should sound a note of caution here: The obtained minimum of J is not necessarily the global minimum, as there is nothing in the algorithm to help avoid local minima. To minimise J, we differentiate it with respect to a_i and set the derivative to zero:

$$\begin{aligned}\frac{\partial J}{\partial a_i} &= \sum_{k=1}^{p}(\mu_{ik})^q \frac{\partial}{\partial a_i}\left[(r_k^T - a_i^T)(N^{-1}r_k - N^{-1}a_i)\right] \\ &= \sum_{k=1}^{p}(\mu_{ik})^q \frac{\partial}{\partial a_i}\left[(r_k^T N^{-1} r_k - r_k^T N^{-1} a_i - a_i^T N^{-1} r_k + a_i^T N^{-1} a_i)\right] \end{aligned} \tag{52}$$

Using formulae (17) to (19), we obtain,

$$\begin{aligned}\frac{\partial J}{\partial a_i} &= \sum_{k=1}^{p}(\mu_{ik})^q \left[(r_k^T N^{-1})^T - N^{-1}r_k + 2N^{-1}a_i\right] \\ &= 2N^{-1}\sum_{k=1}^{p}(\mu_{ik})^q \left[-r_k + a_i\right] = 0 \Rightarrow \\ \Rightarrow \quad \hat{a}_i &= \frac{\sum_{k=1}^{p}(\mu_{ik})^q r_k}{\sum_{k=1}^{p}(\mu_{ik})^q}; \qquad 1 \le i \le n; \end{aligned} \tag{53}$$

It is customary in C-means fuzzy clustering to choose the membership function to a class i to be inversely proportional to the distance of observation k from the cluster that represents class i. Let us say that we choose μ_{ik} to be:

$$\mu_{ik} = \frac{A_k}{d^{\frac{2}{q-1}}(r_k, a_i)} \tag{54}$$

where A_k is some constant of proportionality. We may define it by using the sum to 1 constraint:

$$\sum_{i=1}^{n}\mu_{ik} = 1 \Rightarrow \frac{A_k}{d^{\frac{2}{q-1}}(r_k, a_i)} = 1 \Rightarrow A_k = \frac{1}{\sum_{i=1}^{n}\frac{1}{d^{\frac{2}{q-1}}(r_k, a_i)}} \tag{55}$$

Then

$$\begin{aligned}\mu_{ik} &= \frac{1}{d^{\frac{2}{q-1}}(r_k, a_i)\left[\sum_{j=1}^{n}\frac{1}{d^{\frac{2}{q-1}}(r_k, a_j)}\right]} \\ &= \frac{1}{\sum_{j=1}^{n}\left[\frac{d(r_k, a_i)}{d(r_k, a_j)}\right]^{\frac{2}{q-1}}} \\ &= \left[\sum_{j=1}^{n}\left(\frac{d(r_k, a_i)}{d(r_k, a_j)}\right)^{\frac{2}{q-1}}\right]^{-1}; \qquad 1 \le k \le p; \quad 1 \le i \le n; \end{aligned} \tag{56}$$

Let us use this value of u_{ik} into (50):

$$J = \sum_{i=1}^{n}\sum_{k=1}^{p}\mu_{ik}\mu_{ik}^{q-1}d^2(r_k, a_i)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\sum_{k=1}^{p}\mu_{ik}\left[\frac{A_k}{d^{\frac{2}{q-1}}(r_k,a_i)}\right]^{q-1} d^2(r_k,a_i) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{p}\mu_{ik}\frac{A_k^{q-1}}{d^2(r_k,a_i)}d^2(r_k,a_i) \\
&= \sum_{k=1}^{p}A_k^{q-1}\sum_{i=1}^{n}\mu_{ik} = \sum_{k=1}^{p}A_k^{q-1} \\
&= \sum_{k=1}^{p}\frac{1}{\left[\sum_{i=1}^{n}\left(\frac{1}{d^2(r_k,a_i)}\right)^{\frac{1}{q-1}}\right]^{q-1}} \qquad (57)
\end{aligned}
$$

We observe that minimising J is equivalent to maximising the sum of the inverses of the square distances of each observation from all classes:

$$J = \sum_{k=1}^{p}\left[\sum_{i=1}^{n}\left(d^2(r_k,a_i)\right)^{\frac{1}{1-q}}\right]^{1-q} \qquad (58)$$

J is optimised by looping back and forth between equations (53) and (56) until the iteration where the changes in successive entries of $\hat{U}$ become small.

It is easy to show that when $q \to 1$ a 'hard' classification may be obtained. This is because when $q \to 1$, $\mu_{ik} = 1$ if $d(r_k,a_i) < d(r_k,a_j)$ for all $j \neq i$ and otherwise $\mu_{ik} = 0$. This can be shown as follows,

$$
\begin{aligned}
\mu_{ik} &= \left[\sum_{j=1}^{n}\left(\frac{d(r_k,a_i)}{d(r_k,a_j)}\right)^{\frac{2}{q-1}}\right]^{-1} \\
&= \left[\left(\frac{d(r_k,a_i)}{d(r_k,a_i)}\right)^{\frac{2}{q-1}} + \sum_{j\neq i}\left(\frac{d(r_k,a_i)}{d(r_k,a_j)}\right)^{\frac{2}{q-1}}\right]^{-1} \\
&= \left[1 + \sum_{j\neq i}\left(\frac{d(r_k,a_i)}{d(r_k,a_j)}\right)^{\frac{2}{q-1}}\right]^{-1} \qquad (59)
\end{aligned}
$$

Equation (59) shows that μ_{ik} is equal to the inverse of the sum of 1 plus a term. When $q \to 1$, if $d(r_k,a_i) < d(r_k,a_j)$ for all $j \neq i$, the second term in (59) will tend to zero, since it is raised to a very high power, thus $\mu_k = 1$. Otherwise, if $d(r_k,a_i) > d(r_k,a_j)$ for any $j \neq i$, at least one of the components of the second term in (59) will tend to infinity, thus the inverse of the sum will tend to zero ($\mu_k \simeq 0$). There is no optimal value for q. However, in most studies a value in the range $1.5 < q < 3.0$ is used [3].

In [12], regression equations were derived from the relationship between a pixel's proportion of a pure class and the observed fuzzy membership function of the pixel to the particular class. These

equations were then used to derive the predicted proportions for test pixels from the values of their membership functions.

Petrou and Foschi [28] proposed a fuzzy linear unmixing approach to model the variability of each pure class using non parametrical representations of the constituent pure classes, with individual samples being treated as bona fide representatives of their classes. Spatial arrangement information is used to identify such samples. They noticed that due to intraclass variability of the mixing components, more than one linear combination can lead to the same mixture spectrum [32]. The fuzzy linear unmixing approach estimates the membership of a mixed pixel in various possible mixing proportions. The method relies on the creation of model mixtures and the estimation of their local density in the vicinity of the pixel under consideration. For each possible mixture, a confidence is calculation using the density of the n-class model mixture. For all possible mixtures, all possible combinations of pure class pixels are considered, and a mixture cluster of these particular proportions is created. These mixture classes are considered to be model mixtures for the particular locality in the feature space. Given a mixed pixel, a feature space disc of radius r (defined arbitrarily, for instance as half the maximum distance between points) is considered around it. Inside the disc, the number of model samples is counted form each of the possible combinations. The confidence is calculated by dividing the number of model samples by the total number of samples in the model mixture cluster. This can be used not only to identify the mixing proportions but also to give a confidence to them. In [28] the method is described and tested for two classes, but it could be extended to higher dimensions and more pure classes.

2.4.3 Artificial Neural Networks

Artificial Neural Networks (ANN) are computer models that imitate the biological nerve cell networks. An ANN can be compared with a black box having multiple inputs and multiple outputs which operates using a large number of mostly parallel connected simple arithmetic units incorporating a non-linear dependence between the inputs and outputs (Figure 10). A MLP consists of neurons organised in an input layer, one or more hidden layers, and an output layer. These neurons are arranged in such a way that each of them has a weighted connection coming from every neuron in the previous layer. Each neuron performs a summation of all its inputs and passes the value through a non-linear function before sending it to the output (Figure 10). Feed-forward networks based on the multi-layer perceptron (MLP) are mainly used because of their training capabilities.

Training an ANN is done by presenting a set of examples to the network that after a number of iterations should be able to generalise and learn. ANNs have been used in many classification problems, and in the recent years, their potential as a method to decompose mixed pixels has been demonstrated. Foody [13] found that the outputs of an ANN trained with pure pixels can be regarded as class membership grades, whose strength can be used to derive class proportions. Foody [13] and Schouten and Gebbinck [37] have proven that ANNs can also be trained successfully using mixed pixels as inputs, having class proportions directly as the outputs. A pixel spectral vector is transformed into a fractions vector. The network has then n inputs corresponding to the spectral bands and m outputs corresponding to the pure classes. The architecture of the ANN should be based on the characteristics of the data (Figure 10).

The accuracy of the network depends on the architecture and weight selection. The training algorithm used is based mainly on backpropagation, where the difference between the example and the actual output is computed and fed back through the network, changing the weights in proportion to the error of the neurons they connect. There are a number of parameters that need to be adjusted and the training can be a long procedure. However, once the network is trained, mixed pixels can be decomposed into their fractions with a speed comparable to that of the LSE in the linear mixture model, with higher accuracy [37] [13].

2.4.4 Decomposition based on Geometric models

In this model, the scene is considered as a plane ground containing trees, that can produce shadows in the image. Usually four ground covers are considered, namely, canopy, shadowed canopy, background and shadowed background. The reflectance of each pixel is calculated as a linear combination of the refectances of these components. The component areas depend on the scene geometry, i.e. shape and size of the trees, heights, densities, etc. Otterman [26] modelled forest trees and desert vegetation as thin vertical cylinders of random height and density. Other authors have tried to fit stochastic models [18]. Since landscapes are also characterised by variations in vegetation and soil background, some algorithms treat the reflectance terms as random variables. The new sets of equations provided to the linear model are useful when few bands are used. However, in some cases, those models can introduce further unknowns, and the inversion would then present some difficulties [18] [26].

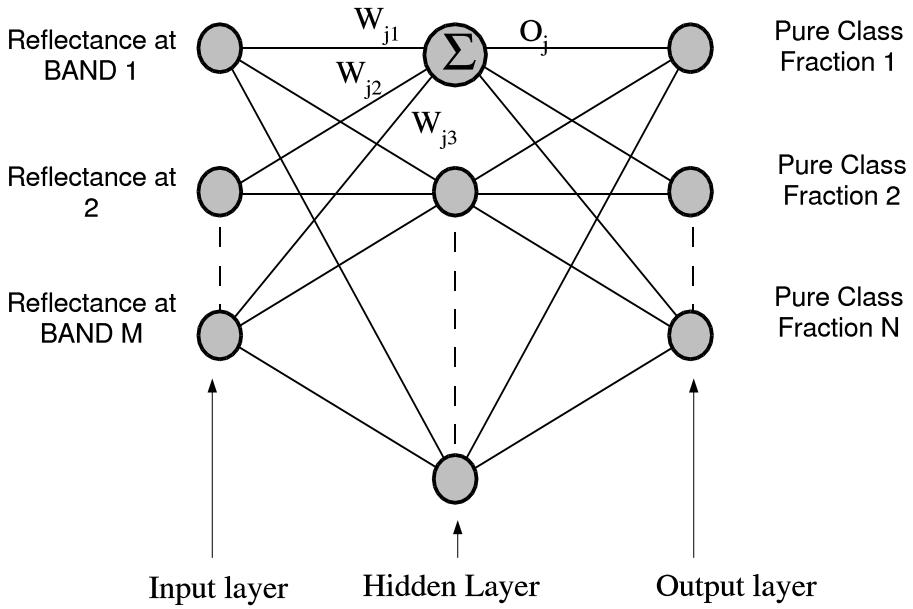

Figure 10: ANN in mixed pixel classification. Here an example with a single hidden layer is shown. In the input layer, information about the reflectance at bands $1, \ldots, .., m$ is fed. At every neuron a weighted summation of the inputs is calculated and passed to the next layer. In the output layer, information about the fractions is obtained.

3 Summary and comparative studies

Different methods to decompose mixed pixels have been reviewed. The most widely used techniques are based on the linear model, that relies on the following main assumptions:

- The spectrum of the mixed pixel is a linear combination of the spectra of the constituent pure classes, weighed by their proportions.
- There are no intimate mixtures: spectra proportions of the pure classes reflect the corresponding cover proportions on the ground.

This approach has been used to solve two different problems:

- Single pixels classification, and
- classification of sets of pixels

Most of the methods reviewed here follow the linear model. Different theoretical and experimental studies exist in the literature. Settle and Drake [35] compared the linear model classical estimator and its variants. They calculated the errors associated with the different estimators, namely,

- the unconstrained classical estimator (i),

- normalisation of (i), by setting to zero negative fractions, and rescaling the rest of the proportions to sum to one, (ii),
- partially constraint estimator, forcing the proportions to sum to one (given by (31)), (iii),
- and a fully constrained estimation, (iv).

For each signal, two measures were calculated,

- a measure of the reconstruction performance with the estimated proportions, (a),
- The RMS (root mean square error) of the proportions vector (b)

These quantities are calculated as follows,

$$e_r = \sqrt{< \frac{(r - Af)^T (r - Af)}{m} >} \tag{60}$$

The RMS is defined as follows,

$$e_f = \sqrt{< \frac{(f - f_t)^T (f - f_t)}{n} >} \tag{61}$$

where f_t are the true proportions. In that study, it was noted that the most the estimate was constrained, the poorer the fit was, but the better the accuracy of the estimate. They also showed that with the regularisation estimator, the RMS error of the proportions estimate could be reduced by half [35].

Ichoku and Karnieli [18] reviewed the linear model, the linear discriminant, geometric model and fuzzy models. They made a theoretical comparative study where they examined different aspects of the models. The number of spectral bands required was examined (see Table (1)). Inversion methods, accuracy of pixel unmixing, assumptions and simplifications, and applications were also comparatively described.

Robust methods applied to sets of pixels were also presented. The use of second order moments [6] has the advantage of allowing the classification of more pure classes than available bands. In principle, the method could be extended to the third, fourth, etc order moments. However, the number of samples available for each class is the main limitation [29]. The Hough transform-based method has the advantage of being robust against outliers. However, the Hough transform presents a main drawback: for large sets, it can be very time-consuming. The randomised Hough transform

Unmixing method	n classes discriminated	Advantages	Disadvantages
LSE - Linear model	$n \leq m+1$	Widely used and well-known	Limited number of pure classes discriminated
Fuzzy classification	n not limited by m	The variability of pixels can be identified	The accuracy can be compromised if $n >> m$
Geometrical mixture model	n not limited by m	The number of equations can be increased; extra information	Inversion is complicated and further parameters are required; Limited application, usually to only four classes
Artificial Neural networks	n not limited by m	Accurate results	Training can be time consuming
Second order moments	$n > m$	n is increased with respect to the linear model, and can be further increased using further moments	Number of pixels required can be prohibitive
Hough transform	$n \leq m+1$	Robust to outliers; multiple mixtures may be identified	Time consuming

Table 1: Comparative table of some unmixing methods. n represents the number of classes that can be discriminated and m the number of bands available [24] [15] [18]

[7], the Trimmed means [5] and the LMeds [34] methods were proposed to solve the time consuming problem of the standard Hough transform method [34].

Neural networks, fuzzy C-means and the linear mixture modelling approaches were compared by Atkinson *et al* [2], who concluded that neural networks gave the most accurate estimates. Schouten and Gebbinck [37], found neural network results comparable with the brute force solution of the linear mixture model and better than those achieved by the analytical approach.

Table 1 summarises the advantages and disadvantages of some of the methods described here. The maximum number of pure classes that can be unmixed with respect to the number of bands is also considered.

Appendix A - The variance of the sum of two random variables

Let us assume that $z = ax + by$, where a and b are some constants and x and y are random variables with mean $\bar{x}$ and $\bar{y}$ respectively and variances σ_x^2 and σ_y^2 respectively. We wish to calculate the variance of z, σ_z^2. We denote by $< >$ the expectation operation, i.e. taking the average value of a random variable over all its instantiations. Then,

$$\bar{z} =< z >= a < x > +b < y >= a\bar{x} + b\bar{y} \tag{62}$$

$$\begin{aligned}
\sigma_z^2 &= < (z - \bar{z})^2 >=< (ax + by - a\bar{x} - b\bar{y})^2 > \\
&= < a^2(x - \bar{x})^2 + b^2(y - \bar{y})^2 + 2ab(x - \bar{x})(y - \bar{y}) > \\
&= a^2 < (x - \bar{x})^2 > +b^2 < (y - \bar{y})^2 > +2ab < (x - \bar{x})(y - \bar{y}) > \\
&= a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab < (x - \bar{x})(y - \bar{y}) >
\end{aligned} \tag{63}$$

Suppose that the random variables are independent. Then $< (x - \bar{x})(y - \bar{y}) >= 0$ and

$$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 \Rightarrow \sigma_z \tag{64}$$

If x and y are correlated, we may write $y = \alpha x$ and $\bar{y} = \alpha\bar{x}$. Upon substitution in (63), we obtain:

$$\begin{aligned}
\sigma_z^2 &= a^2 < (x - \bar{x})^2 > +b^2 < (\alpha x - \alpha\bar{x})^2 > +2ab < (x - \bar{x})(\alpha x - \alpha\bar{x}) > \\
&= a^2\sigma_x^2 + b^2\alpha^2\sigma_x^2 + 2ab\alpha\sigma_x^2 \\
&= (a + b\alpha)^2\sigma_x^2
\end{aligned} \tag{65}$$

Now assume that a pixel consists of k_j replicas of different instantiations of the same object, say a tree. Assume that the brightness b of this object is a random variable with mean μ_j and variance σ_j^2. Then the total brightness of the pixel is

$$b_T = b_1 + b_2 + \ldots + b_{kj} \tag{66}$$

The mean value of b_T will be

$$1\times < b_1 > +1\times < b_2 > +\ldots + \ 1\times < b_{k_j} >= k_j\mu_j \tag{67}$$

and the variance of its brightness would be

$$1^2 \times \sigma_{b_1}^2 + 1^2 \times \sigma_{b_2}^2 + \ldots + \ 1^2 \times \sigma_{b_{k_j}}^2 = k_j\sigma_j^2 \tag{68}$$

References

[1] Adams, J.B., Smith, M.O. and Johnson, P.E. "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site", *Journal of geophysical research*, Vol. 91 (B8), pp. 8098–8112, 1986.

[2] Atkinson, P.M., Cutler, M.E.J. and Lewis, H. "Mapping sub-pixel proportional land cover with AVHRR imagery", *International Journal of Remote Sensing*, Vol. 18 (4), pp. 917–935, 1997.

[3] Bezdek, J., Ehrlich, R. and Full, W. "FCM: the fuzzy c-means clustering algorithm", *Computers and Geosciences*, Vol. 10, pp. 191–203, 1984.

[4] Bosdogianni, P., Petrou, M. and Kittler, J. "Robust mixed pixel classification using the hypothesis testing Hough transform", *International Geoscience and Remote Sensing Symposium, (IGARSS '96)*, Vol. 2, pp. 1376 –1378, 1996.

[5] Bosdogianni, P., Petrou, M. and Kittler, J. "Mixed pixel classification with robust statistics", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 35 (3), pp. 551–559, 1997.

[6] Bosdogianni, P., Petrou, M. and Kittler, J. "Mixture models with higher order moments", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 35 (2), pp. 341–353, 1997.

[7] Bosdogianni, P., Kälviäinen, H., Petrou, M. and Kittler, J. "Robust unmixing of large sets of mixed pixels", *Pattern Recognition Letters*, Vol. 18 (5), pp. 415–424, 1997.

[8] Bosdogianni, P., Petrou, M. and Kittler, J. "Classification of sets of mixed pixels with the hypothesis-testing Hough transform", *IEE Proceedings: Vision, Image and Signal Processing*, Vol. 145 (1), pp. 57–64, 1998.

[9] Brown, P.J. "Multivariate calibration", *Journal of the Royal Statistical Society*, Vol. 44 (3), pp. 287–321, 1982.

[10] Faraklioti, M. and Petrou, M. "Recovering more classes than available bands for sets of mixed pixels in satellite images", *Image and Vision Computing*, Vol. 18 (9), pp. 705–713, 2000.

[11] Fisher, P. "The pixel: a snare and a delusion", *International Journal of Remote Sensing*, Vol. 18 (3), pp. 679–685, 1997.

[12] Foody, G.M. and Cox, D.P. "Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions", *International Journal of Remote Sensing*, Vol. 15, pp. 619–631, 1993.

[13] Foody, G.M., Lucas, R.M., Curran, P.J. and Honzak, M. "Non-linear mixture modelling without end-members with an artificial neural network". *International Journal of Remote Sensing*, Vol. 18, pp. 937–953, 1997.

[14] Gebbinck, M. and Schouten, T. "Decomposition of mixed pixels", J. Desachy, *Image and Signal Processing for Remote Sensing II*, pp. 104–115, Paris, 1995.

[15] Gebbinck, M. Decomposition of mixed pixels in remote sensing images to improve the area estimation of agricultural fields, Thesis dissertation, Katholieke Universiteit Nijmegen, 1998.

[16] Heinz, D.C. and Chang, C.-I. "Unsupervised fully constrained squares linear spectral mixture analysis method for multispectral imagery", *IEEE 2000 International Geoscience and Remote Sensing Symposium, (IGARSS 2000)*, Vol. 4 , pp 1681–1683, 2000.

[17] Horwitz, H.M., Nalepka, R.F., Hyde, P.D. and Morgenstern, J.P. "Estimating the Proportions of Objects within a Single Resolutions Element of a Multispectral Scanner", International Symposium on Remote Sensing of Environment, pp. 1307–1320, 1971.

[18] Ichoku, C. and Karnieli, A. "A review of mixture modeling techniques for sub-pixel land cover estimation", *Remote Sensing Reviews*, Vol. 13, pp 161–186, 1996.

[19] Keshava, N. and Mustard, J.F. "Spectral unmixing", *IEEE Signal Processing Magazine*, Vol. 19 (1), pp. 44–57, 2002.

[20] Landgrebe, D. Signal Theory Methods in Multispectral Remote Sensing, John Wiley and Sons, 2003.

[21] Li, X. and Strahler, A.H. "Geometric-optical modeling of a conifer forest canopy", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 23, pp. 705–721, 1985.

[22] Lillesand, T.M and Ralph, W.K. Remote Sensing and Image Interpretation. John Wiley and Sons, New York, USA, 1994.

[23] Luo, J., King, L. and Younan, N. "An Unmixing Algorithm based on vicinnal information", *IEEE 2002 International Geoscience and Remote Sensing Symposium, (IGARSS 2002)*, Vol. 3, pp. 1453–1455, 2002.

[24] Marsh, S.E., Switxer, P. and Kowalik, W.S. "Resolving the percentage of component terrains within single resolution elements", *Photogrammetric Engineering and Remote Sensing*, Vol. 46 (8), pp. 1079–1086, 1980.

[25] Maselli, F. "Multiclass spectral decomposition of remotely sensed scenes by selective pixel unmixing", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 36, pp. 1809–1820, 1998.

[26] Otterman, J. "Albedo of a forest modeled as a plane with dense protrutions", *Journal of Climate Applicaltions Meteorology*, Vol. 22, pp. 297–307, 1984.

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

[28] Petrou, M. and Foschi, P.G. "Confidence in linear spectral unmixing of single pixels", *IEEE Transactions in Geoscience and Remote Sensing*, Vol. 37 (1), pp. 624–626, 1999.

[29] Petrou, M. "Mixed pixel classification: an overview", In *Information Processing for Remote Sensing*, C.H. Chen, World scientific Publishing, Singapore, pp. 69–83, 1999.

[30] Petrou, M. and Bosdogianni, P. Image Processing, the Fundamentals, Chichester, John Wiley, 1999.

[31] Pickup, G., Chewings, V.H. and Nelson, D.J. "Estimating changes in vegetation cover over time in arid rangelands using Landsat MSS data", *Remote Sensing of Environment*, Vol. 43, pp. 243–263, 1993.

[32] Price, J. C. "How Unique Are Spectral Signatures?", *Remote Sensing of Environment*, Vol. 49 (3), pp. 181–186, 1994.

[33] Ramsey, M.S. and Christensen, P.R. "Mineral abundance determination: Quantitative deconvolution of thermal emission spectra", *Journal of Geophysical Research*, Vol. 103, pp. 577–596, 1998.

[34] Rosin, P.L. "Robust pixel unmixing", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39 (9), pp. 1978–1983, 2001.

[35] Settle, J.J. and Drake, N.A. "Linear mixing and the estimation of ground cover proportions", *International Journal of Remote Sensing*, Vol. 14, pp. 1159–1177, 1993.

[36] Shimabukuro, Y.E. and Smith, J.A. "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 29 (1), pp. 16–20, 1991.

[37] Schouten, T.E. and Gebbinck, M.S. "A neural network approach to spectral mixture analysis", In *Neurocomputation in Remote Sensing Data Analysis*, I. Kanellopoulos, G.G. Wilkinson, F. Roli, J. Austin, Springer, pp. 79–85, 1997.

[38] Tucker, C.J. "Red and photographic infrared linear combinations for monitoring vegetation", *Remote Sensing of the Environment*, Vol. 8, pp. 127–150, 1979.

[39] Tucker, C. J., Newcomb, W. W., Los, S. O. and Prince, S. D. "Mean and inter-year variation of growing-season normalized difference vegetation index for the Sahel 1981–1989", *International Journal of Remote Sensing*, Vol. 12, pp. 1113–1115, 1991.