Article

Confidence Intervals and Regions for Proportions Under Various Three-Endmember Linear Mixture Models

Mark Berman [1]

[1] CSIRO Data61, Marsfield NSW 2122, Australia; mark.berman@data61.csiro.au

Abstract: Many papers in recent years have been devoted to estimating the per pixel proportions of three broad classes of materials (e.g. photosynthetic vegetation, non-photosynthetic vegetation and bare soil) using data from multispectral sensors. Many of these papers use estimation methods based on the linear mixture model. Very few of these papers assess the accuracy of their estimators. I show how to produce confidence intervals (CIs) and joint confidence regions (JCRs) for the proportions associated with various linear mixture models. There are two main models, both of which assume that the coefficients in the model are non-negative. The first model assumes that the coefficients sum to 1. The second does not, but uses rescaling of the estimated coefficients to produce estimated proportions. Three variants of these two models are also analysed. JCRs are shown to be particularly informative, because they are typically better at localising the information than CIs are. The methodology is illustrated using examples from Landsat Thematic Mapper data at 1169 locations across Australia, each of which has associated field observations. There is also discussion about the extent to which the methodology can be extended to hyperspectral data.

Keywords: proportion estimation; linear mixture model; confidence interval; confidence region; photosynthetic vegetation; non-photosynthetic vegetation; soil; Landsat Thematic Mapper

1. Introduction

There have been many papers in the remote sensing literature in recent years which have applied linear mixture models to data from multispectral sensors such as Landsat Thematic Mapper (TM) and MODIS to estimate the proportions (often called fractions or abundances) of a small number of *broad* classes. Most of these papers aim to estimate the per pixel proportions of three endmembers. Probably the most popular such model is for mixtures of photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV) and bare soil (BS) [1–12]. In [13], the three endmembers are PV, NPV and snow, while in [14], they are shade, vegetation and other landforms (including water, rock and sand).

In [15], shade is a fourth endmember, added to PV, NPV and BS. In [16], three of these endmembers are used, together with NPV-Ash instead of NPV, because they are interested in mapping burn severity after fires. Of course, the shade endmember is usually not of intrinsic interest, but is included to improve the fitting. So the interest is still in the *relative* proportions of the three *primary* endmembers.

Another fairly popular three-endmember model for urban scenes is the vegetation-impervious surface-soil (VIS) model [17–22]. All of these papers, except [17], analyse various Landsat TM data sets using a linear mixture model.

Apart from the AutoMCU method [1], which requires endmember "bundles" [23] and uses Monte Carlo sampling to estimate confidence intervals (CIs) for each of the three proportions, none of the methods used in the literature provides methods for assessing the accuracy of their estimators.

In this paper, I show how to produce CIs and joint confidence regions (JCRs) for the proportions associated with five three-endmember linear mixture models with different constraints. The requirement that the coefficients in the linear model are proportions complicates the model in two ways. First, the coefficients must sum to 1 and second they

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

must all be non-negative. These significantly constrain the solution space, so that even with moderately good fits, the CIs of individual proportions (especially of NPV and BS) can be quite large, and it can be difficult to interpret them in a meaningful way. The use of JCRs for the three proportions overcomes this difficulty.

I have focussed this paper on three-endmember models for two main reasons. First, the significant number of papers focussed on such models (some of which are listed above) suggests that such models are of great interest. Second, because the true and estimated proportions must sum to 1, they can be easily displayed inside a *two-dimensional* triangle, often in the form of a ternary diagram.

The results in this paper can be generalised to linear mixture models with more than three endmembers. However, the mathematics is more complicated (and will be published in a more theoretical journal). In addition, *visualisation* of the results is difficult except possibly for the four-endmember model, where true and estimated proportions can be displayed inside a tetrahedron using *three dimensional* visualisation software. For most readers, such displays are not as easy to interpret as are proportions inside triangles, and perhaps explains why three-endmember models are much more common in the remote sensing literature than models with four or more endmembers.

Some of the papers mentioned above have a *single* endmember spectrum representing each broad class [2,15,18–21]. This is often called spectral mixture analysis (SMA) or linear SMA (LMSA). Typically, the endmembers are obtained from the data set itself. Others obtain the endmembers by using suitable indices such as NDVI and the SWIR32 vegetation index to create a suitable two-dimensional space in which to unmix the data [3,4,9–12].

Some other papers recognise that within broad classes, there is "endmember variability" [24]. So they use libraries with multiple examples of pure "spectra" drawn from each class. *Single* spectra are then drawn from each of the classes for use in the linear mixture model in such a way that a best fit is achieved according to some criterion. Some of these papers use the popular multiple endmember spectral mixture analysis (MESMA) software [25] or variants of it [7,14,16,22]. MESMA builds two-, three- and sometimes four-endmember models in a *stepwise* manner (e.g. three-endmember models are built upon chosen two-endmember models). Usually "shade" (often zero reflectance in all bands [7]) is one of the endmembers included in the model.

AutoMCU [1,26] (called Monte Carlo SMA by [3]) also chooses one endmember from multiple candidates in each class (in their case, PV, NPV or BS). However, rather than use the stepwise approach of MESMA, they use Monte Carlo sampling as an approximation to investigating *all* PV/NPV/BS combinations. Typically, they randomly select e.g. 50 PV/NPV/BS combinations. For each combination, proportion estimates are obtained. For each spectrum, the mean and standard deviation of the 50 proportion estimates are used to obtain CIs for the PV, NPV and BS proportions.

While the theory presented in this paper can easily accommodate linear mixture models with a *single* candidate endmember for each class, it cannot deal directly with the case where one out of multiple candidate endmembers is chosen from each class. Calculating accurate CIs and JCRs after choosing a best model/subset according to some criterion is usually very difficult; see [27] and references therein. In Section 2.3.4, I offer an alternative approach, whereby the endmembers in each class are themselves modelled as mixtures of the *extreme* endmembers in that class. I will show how this can be done and discuss its advantages and disadvantages.

In order to construct CIs and JCRs associated with linear mixture models, one needs at least one degree of freedom (df) to estimate error variances. For six-band Landsat TM data, for instance, this means that the total number of endmembers that can be fitted in the model is 6 if the sum-to-one (proportion) constraint is enforced and 5 if the constraint is not enforced. These endmembers can consist of several *extreme* endmembers from some of the broad classes, or secondary endmembers (such as shade and/or water), as well as the primary endmembers of interest.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

The research presented here has been motivated by a data set of 1169 six-band Landsat TM spectra and associated field measurements collected at 913 sites around Australia. These have been analysed previously by [8]. The lead author of that paper asked me how to construct CIs for his proportion estimators. I will use exploratory analyses of the two associated data sets to suggest a suitable linear mixture model, consisting of five endmembers (1 PV, 2 NPV and 2 BS) for this data set. Two general mixture models will be analysed first (with and without the sum-to-one constraint respectively). These will then be specialised to deal with the cases of (i) primary and secondary endmembers, and (ii) multiple endmembers in some classes. Finally, I will show how to relax some error variance assumptions.

The paper is structured as follows. Section 2.1 discusses the above-mentioned data set and uses some exploratory data analysis to suggest plausible linear mixture models for it. Two main linear mixture models are then introduced in Section 2.2. Both of them assume that the coefficients in the model are *non-negative*. The first model assumes that the coefficients sum to 1. The second does not, but uses rescaling of the estimated coefficients to produce estimated proportions. Some assumptions about the error structure are also introduced. In Section 2.3, I show how to construct CIs and JCRs for the two main models, plus three variants of them. I do this via two principles, the latter of which I call the *Intersection Principle*. The JCRs for all the models involve the intersection of an ellipse and a triangle. This intersection is not easily represented by a small number of parameters, which can be used to summarise multispectral *image* data. In Section 2.4, I briefly discuss approximations which require only six parameters, and so which can be displayed as two color images. A summary and discussion are given in Section 3. This includes a discussion about the extent to which the methodology introduced in Section 2.3 can be extended to hyperspectral data.

2. Materials and Methods

2.1. A Data Set and Some Exploratory Data Analyses

2.1.1. The data set

The data set that I will analyse and that will motivate the more complex models has been analysed previously by [8]. It consists of 1169 Landsat TM spectra and associated field measurements collected at 913 sites around Australia; see [8, Figure 1]. Details of how the field measurements were obtained can be found in [8, Section 2.1]. Field-based estimates of PV, NPV and BS proportions over the 1169 sites were obtained. Surface reflectance for the six Landsat TM (non-thermal) bands was estimated by averaging a 3×3 pixel window closely corresponding to the area of each field measurement. The differences between the dates of the Landsat TM data and the corresponding field measurements had a mean of -0.6 days and a standard deviation of 8.2 days. However, these differences were as much as 60 days in some cases. Because of the the temporal discrepancies between the field and Landsat data sets, one needs to treat the field-based proportion estimates with care. I will call them the *nominal* proportions and will use the information that they provide as an informal guide, rather than treating them as the truth.

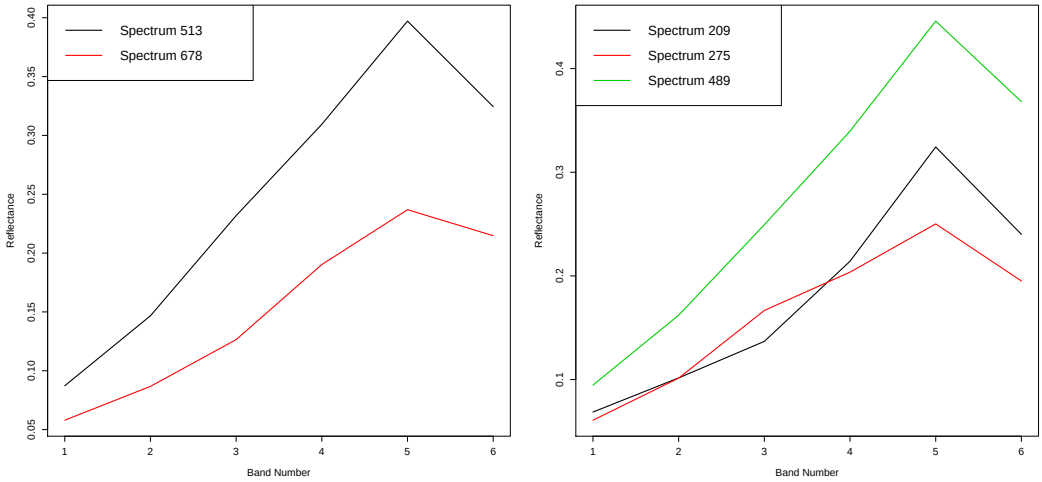
In [8] it is assumed that there are *unique* PV, NPV and BS endmembers. This is an unrealistic assumption considering that the samples have been collected across the large Australian continent. Unlike the above-mentioned studies which build endmember libraries directly from field or image spectra, [8] estimate the three endmembers using the nominal proportions and an inverse regression method [28] applied to the model (1). Because they have assumed unique PV, NPV and BS endmembers, they model the resulting variability by building a *non-linear* model with 24 predictors; see [8, Section II.D.3 and Figure 4] for details.

It is difficult to interpret the meaning of these 24 predictors. More importantly for our purposes, it does not easily lend itself to the construction of CIs and JCRs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

2.1.2. Some exploratory data analyses

The exploratory data analyses will take two simple forms. The first compares spectra in the data set with *identical* nominal proportions. There are in fact eleven pairs and one triplet of spectra with identical nominal proportions. Figure 1(a) shows one of these pairs, spectra 513 and 678, while Figure 1(b) shows the triplet, spectra 209, 275 and 489. The common nominal proportions, listed as p_{PV}, p_{NPV} and p_{BS}, are given in the captions of the two plots, correct to 3 decimal places (although they are in fact equal to 6 decimal places). What stands out in these plots (and many of the other plots of pairs of spectra with identical nominal proportions) is that many spectra with the same nominal proportions have similar shapes but different average values (i.e. brightnesses or albedos). The most obvious explanation for this is a shade component in the model, which is often included in the MESMA model. I will discuss ways of modelling shade in Section 2.2.



(a) Spectra 513 and 678:
$p_{PV} = 0.165, p_{NPV} = 0.329, p_{BS} = 0.506.$

(b) Spectra 209, 275 and 489:
$p_{PV} = 0.000, p_{NPV} = 0.550, p_{BS} = 0.450.$

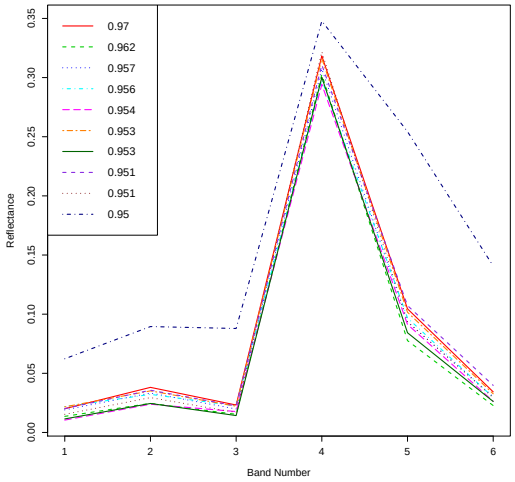
Figure 1. Five spectra from the data set.

The second form of exploratory data analysis involves an examination of the nominally purest spectra in each of the three classes. Figures 2 (a), (b) and (c) show the ten nominally purest PV, NPV and BS spectra respectively. The nominal proportions for the *dominant* groundcover type of each of the spectra are shown in the top left hand corner of each of the three plots. The purest PV spectra all have their maximum value in band 4, which easily distinguishes PV spectra from NPV and BS spectra, which both have their maximum value in band 5, partly explaining the greater difficulty in discriminating between them, especially in mixtures. The nine purest PV spectra are very similar, with $0.951 \leq p_{PV} \leq 0.970$. and $p_{BS} \leq 0.002$. The tenth purest PV spectrum has $p_{PV} = 0.950$ and $p_{BS} = 0.044$.

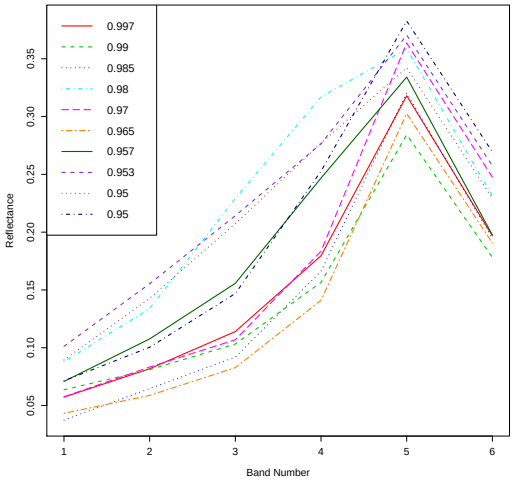
By contrast, the shapes of the ten purest NPV and BS spectra are much more variable, especially in bands 1, 2 and 3. For the ten purest NPV spectra, $0.950 \leq p_{NPV} \leq 0.997$, while for the ten purest BS spectra, $0.906 \leq p_{BS} \leq 0.960$.

Although none of the thirty spectra are perfectly pure, they are all more than 90% pure. The greater variability of the purest NPV and BS spectra suggests that more than one endmember is required to model these. For this reason, my final model will use 1 PV, 2 NPV and 2 BS endmembers.

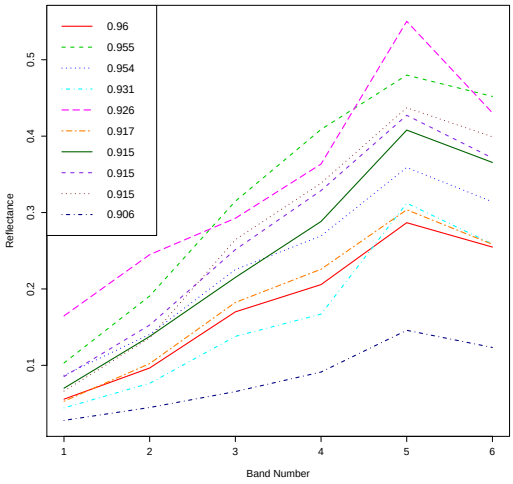
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1



(a) Ten nominally purest PV spectra



(b) Ten nominally purest NPV spectra



(c) Ten nominally purest BS spectra

Figure 2. Ten nominally purest PV, NPV and BS spectra and their PV proportions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

2.2. Two Linear Mixture Models and Error Assumptions

The linear mixture model is usually defined as follows. Let $X_i, i = 1, ..., N$ denote the d-dimensional column vector of observations for sample i (out of N). For Landsat TM data, $d = 6$. Often the samples are contiguous spectra in a multispectral (or hyperspectral) image. However, the samples analysed by [8, Figure 1] consist of 1169 spectra collected at 913 sites around Australia. Under the linear mixture model, if there are $M(< N)$ spectrally distinct materials in the data set, then

$$X_i = \sum_{k=1}^{M} p_{ik} E_k + \epsilon_i, i = 1, \ldots, N, \tag{1}$$

where (i) $E_k, k = 1, \ldots, M$ are the pure endmembers; (ii) p_{ik} are *proportions*, and (iii) ϵ_i are error terms. In Sections 2.3.1 and 2.3.2, where there is only one endmember per class, $M = 3$. However, in Sections 2.3.3 to 2.3.5, $M > 3$. They will then be reduced to three endmembers in various ways.

In most of the above-mentioned papers, because the weights in (1) are interpreted as proportions, we have for each i the constraints:

$$\sum_{k=1}^{M} p_{ik} = 1, \tag{2}$$

and

$$p_{ik} \geq 0, k = 1, \ldots, M. \tag{3}$$

I will call this the Proportion Linear (PL) model.

This model is inadequate to model the brightness variations seen in Figures 1 (a) and (b). This is often dealt with, in part, by adding a shade endmember, often modelled as zero reflectance in all bands, e.g. [7]. If we do this, (1) and (3) still hold, but (2) must be replaced by

$$\sum_{k=1}^{M} p_{ik} \leq 1. \tag{4}$$

This inequality will hold for all the spectra in a data set without having to enforce it if the non-shade endmembers are sufficiently bright. However, this may require some manual intervention. Alternatively, quadratic programming methods [29, Chapter 16] can be used to enforce the inequality. However, a simpler approach is to multiply the deterministic part of the right hand side of (1) by a *positive* scale factor, γ_i, i.e.

$$X_i = \gamma_i \sum_{k=1}^{M} p_{ik} E_k + \epsilon_i, i = 1, \ldots, N. \tag{5}$$

Equation (5) can be rewritten as

$$X_i = \sum_{k=1}^{M} \beta_{ik} E_k + \epsilon_i, i = 1, \ldots, N, \tag{6}$$

where

$$\beta_{ik} \geq 0, k = 1, \ldots, M, \tag{7}$$

$$\gamma_i = \sum_{k=1}^{M} \beta_{ik}, i = 1, \ldots, N, \tag{8}$$

and

$$p_{ik} = \beta_{ik} / \gamma_i, k = 1, \ldots, M; i = 1, \ldots, N. \tag{9}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

Note that the constraint (2) is no longer required. I will therefore call this model the Non-Negative Linear (NNL) model.

The errors in both models are typically a combination of instrumental noise, natural variation in spectra representing the same material, and small non-linearities in the mixing. Typically, the natural variation dominates the error [23, Introduction]. It is common to assume (implicitly or explicitly) that the errors have zero means, and are spatially and spectrally uncorrelated with a Gaussian distribution with *constant* variance σ^2. I will also assume Gaussian errors, but for most of the paper I will assume that

$$Var(\epsilon_{ij}) = \sigma_i^2, j = 1, \ldots, d; i = 1, \ldots, N, \tag{10}$$

where $\epsilon_i \equiv (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{id})$. This enables different samples to have different acceptable goodness-of-fit levels, which will prove useful in Section 2.3.

While I could assume a more general (sample-dependent) variance-covariance structure than (10), the formulae in the paper will be much more complicated. In Section 2.3.5, I will show how the more general variance-covariance structure can be converted to (10).

Under assumption (10), the maximum likelihood estimators of the weights in both models are obtained via least squares (LS) estimation, subject to the constraints (2) and (3) for the PL model, and (7) for the NNL model. I will call these the *constrained* estimators. It will also be useful to calculate the coefficients without the non-negativity constraints (3) or (7) imposed. It will call these the *unconstrained* estimators. I will show both the constrained and unconstrained estimators for selected examples in Section 2.3 using both the PL and NNL models with $M = 3$, and also the NNL model with $M = 5$.

2.3. Confidence Intervals and Joint Confidence Regions for Five Linear Mixture Models

In all five subsections of this section, when obtaining formulae for CIs and JCRs, I will give formulae using a general value of M. I will use specific values of M (e.g. 3 and 5) only where necessary, and in particular for the examples that I show. In Sections 2.3.1 and 2.3.2, I will give the formulae for the general PL and NNL models respectively. In Section 2.3.3, I will discuss the situation of primary and secondary endmembers, while in Section 2.3.4, I will discuss endmember variability [24]. The solutions in these two cases can be obtained via minor modifications of the PL and/or NNL models. In Section 2.3.5, I will discuss the relaxation of the error variance assumption (10).

In Sections 2.3.1 to 2.3.4, it will usually be unnecessary to include the subscript i (the sample number), so for simplicity, I will (mostly) omit it. I will reintroduce it in Section 2.3.5.

2.3.1. The Proportion Linear model

This is probably the most widely used model. It is also the easiest for illustrating two important general principles, which are also applicable to the other models that I will analyse.

It will be instructive to compare two estimators, the *unconstrained* LS estimators of the proportions (i.e. without imposing the non-negativity constraints (3), but imposing the sum-to-one constraint (2)) and the *constrained* LS estimators (i.e. with both constraints (2) and (3) imposed).

The unconstrained LS estimators have been derived by [30, eqn. (11)]. However, their formula assumes a general covariance matrix. This simplifies if it is assumed that the errors are uncorrelated; see (10). Using a simplified notation, their formula becomes

$$\hat{p}_u = \hat{p}_0 + \mu \mathbf{F} \mathbf{1}_M, \tag{11}$$

where $\hat{p}_u^T \equiv (\hat{p}_{1,u}, \hat{p}_{2,u}, \ldots, \hat{p}_{M,u})$, $\mathbf{1}_M$ is a vector consisting of M 1's, $\mathbf{E} \equiv (E_1, E_2, \ldots, E_M)$,

$$\mathbf{F} = (\mathbf{E}^T \mathbf{E})^{-1}, \tag{12}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

$$\mu = (1 - \mathbf{1}_M^T \hat{p}_0)/\delta, \tag{13}$$

$$\delta = \mathbf{1}_M^T \mathbf{F} \mathbf{1}_M \tag{14}$$

and

$$\hat{p}_0 = \mathbf{F}\mathbf{E}^T Y \tag{15}$$

is the standard LS estimator without *either* the constraints (2) and (3) imposed.

It will later be useful to know the covariance matrix of $\hat{p}_u$. In our less general situation, [30, eqn. (13)] simplifies to

$$Cov(\hat{p}_u) = \sigma^2 \mathbf{V}, \tag{16}$$

where

$$\mathbf{V} \equiv \mathbf{F} - \mathbf{F}\mathbf{1}_M \mathbf{1}_M^T \mathbf{F}/\delta. \tag{17}$$

It is straightforward to show that the residual sum of squares of the unconstrained fit, $RSS_{i,u}$, is given by:

$$RSS_u = Y^T \{\mathbf{I} - \mathbf{E}\mathbf{F}\mathbf{E}^T\} Y + \mu^2 \delta, \tag{18}$$

and that an unbiased estimator of σ^2 is given by [31, eqn. (4.29)]:

$$\hat{\sigma}^2 = RSS_u/(d - M + 1). \tag{19}$$

When $M = 3$ and $d = 6$, the denominator, the df, is $d - M + 1 = 4$.

Unlike the unconstrained solution, the constrained solution, $\hat{p}_c$ does not in general have an explicit algebraic solution. It can however be easily obtained using quadratic programming methods [29, Chapter 16]. Quadratic programming code is now widely available. I have used the R package quadprog[1]. Note that, if all the elements of $\hat{p}_u$ are non-negative, then both (2) and (3) are satisfied, in which case $\hat{p}_u = \hat{p}_c$.

I will shortly illustrate the first important principle using a sample spectrum from the data set. However, in order to do this, I need to remove the brightness variations that are apparent in Figures 1(a) and (b). I will do this by dividing each spectrum by its mean value, i.e.

$$Y \equiv X/\bar{X}, i = 1, \ldots, N, \tag{20}$$

where

$$\bar{X} = X^T \mathbf{1}_d/d, \tag{21}$$

where $\mathbf{1}_d$ is a vector of d 1's. I will call Y the *standardised* spectrum. It will also be convenient to assume that, for each endmember, the mean of its values is also 1, i.e.

$$E_k^T \mathbf{1}_d/d = 1, k = 1, \ldots, M. \tag{22}$$

Then, it is straightforward to show [32, Section 2.1] that, if there is *no error* in the NNL model (6), then the standardised spectra satisfy the PL model (1) with $p_k \equiv \beta_k / \sum_{l=1}^{M} \beta_l, k = 1, \ldots, M$. Therefore, *if the errors are not too large*, we can approximately remove the brightness variations *and* obtain a model which approximately satisfies (2) and (3) by standardising *both* the spectra and the endmembers.

In what follows, I will assume $M = 3$ endmembers. In order to satisfy (22), I will use the *standardised versions* of the nominally purest PV, NPV and BS spectra shown in

[1] https://cran.r-project.org/web/packages/quadprog/quadprog.pdf

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

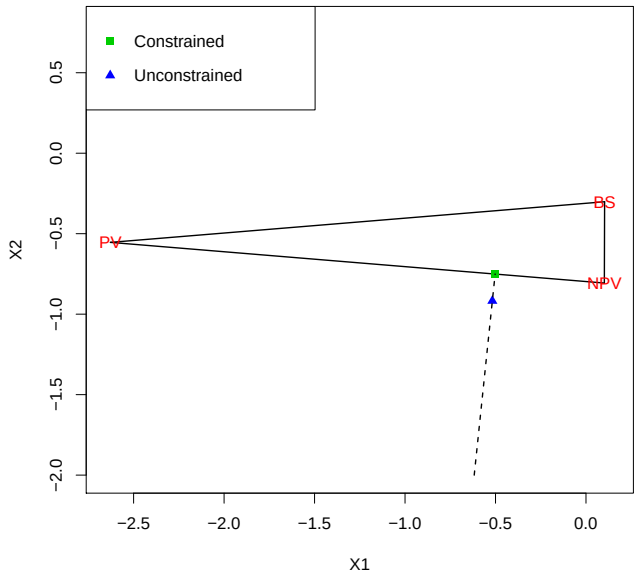


Figure 3. Triangle and plane determined by the standardised purest PV, NPV and BS spectra, and the constrained and unconstrained solutions for spectrum 1099 projected onto that plane.

Figures 2 (a), (b) and (c) respectively as the three endmembers. I will illustrate the first principle using *standardised* spectrum 1099 in the data set.

The three endmembers form the vertices of a triangle. Although this triangle lies in six-dimensional space, we can project it onto a two-dimensional plane (subspace). The projected triangle is shown in Figure 3. For reasons that will shortly become clear, I have ensured that the lengths of the two plotting axes are of equal length. In this figure I also show the projection of the data (Y) onto the plane determined by the endmembers (the blue triangle). This point corresponds to the unconstrained solution. Note that it lies outside the triangle. RSS_u is just the squared distance between Y and this point. I also show the nearest point to the blue triangle on the boundary of the triangle (the green square). It corresponds to the constrained solution. By Pythagoras' theorem, the RSS for the constrained fit, RSS_c, is just RSS_u plus the squared distance between the blue triangle and the green square.

I have also included a broken line between (and beyond) these two points. Because the lengths of the two plotting axes are equal, it can be seen that this broken line is perpendicular to the edge of the triangle nearest to the blue triangle. An important point to note is that *all unconstrained* solutions lying along the broken line have the *same* corresponding *constrained* solution. From a "confidence" perspective, if a point lies on the broken line but near the triangle, intuitively there must be a reasonable likelihood that the "true" point lies *inside* the triangle. However, if the point lies on the broken line but further away from the triangle, intuitively it is more likely that the true point actually lies on the *edge* of the triangle. This plot shows us that, although the constrained estimator is the best *point* estimate consistent with the constraints, it actually throws away information. *Hence statistical inference should be based on the unconstrained estimator, which does not throw away the relevant information.* This is the first major principle that this model enables us to demonstrate.

For the time being, I will ignore the non-negativity constraint (3) in considering confidence intervals and regions based on the unconstrained estimator. I will start with a confidence interval for a single proportion. In what follows, I will use subscripts k (or l) to represent any one (or two) of the M materials. Let v_{kl} denote the (k,l)th element of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

$\mathbf{V}$, given by (17). Then by (16), $\sigma^2 v_{kk}$ is the variance of $\hat{p}_{k,u}$. Hence by standard LS theory [31, eqn. (4.54)],

$$(\hat{p}_{k,u} - p_k)/(\hat{\sigma}\sqrt{v_{kk}}) \tag{23}$$

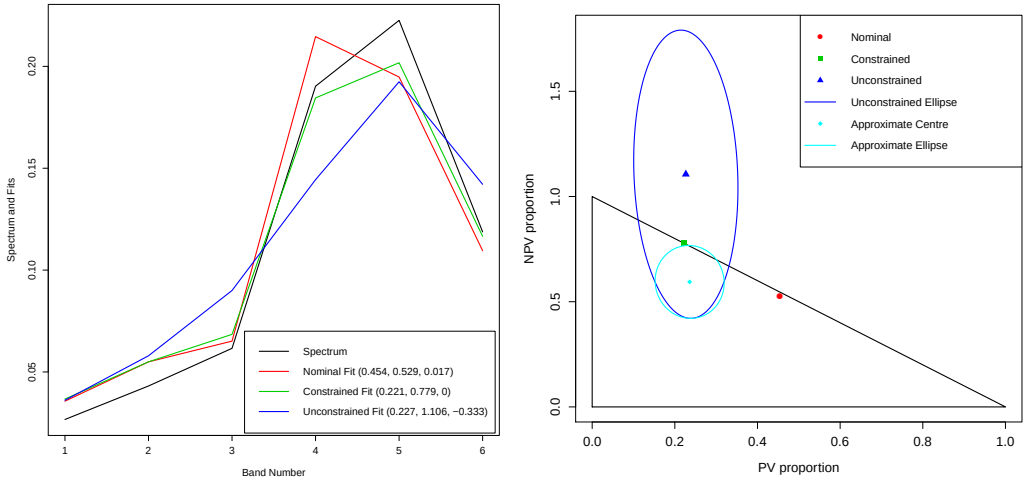
has a t distribution with $d - M + 1$ df, and hence (ignoring the non-negativity constraint (3)) a $100(1-\alpha)\%$ CI for p_k is given by

$$\hat{p}_{k,u} \pm t_{d-M+1,\alpha/2}/(\hat{\sigma}\sqrt{v_{kk}}), \tag{24}$$

where $t_{d-M+1,\alpha/2}$ is the *upper* $100(\alpha/2)$ percentage point of the t distribution with $d - M + 1$ df.

The interpretation of the CI is that it will include the true proportion, on average, $100(1-\alpha)\%$ of the time. However, the fact that the true proportion must lie in $[0,1]$ is additional information which does not invalidate this fact; it just helps us to reduce the size of the CI, *without altering the fact that the true proportion lies in the CI, on average,* $100(1-\alpha)\%$ *of the time*. So the *constrained* CI is just the *intersection* of the *unconstrained* CI with $[0,1]$. This is the second major principle, which I call the *Intersection Principle*, to be introduced in this subsection. This principle has been used previously in [33, Section 7.2] when deriving confidence intervals for linear combinations of non-negative parameters. However, to my knowledge, it has not been used previously for proportion estimation.

Figure 4 (a) shows the nominal, constrained and unconstrained fits for spectrum 1099 using the *standardised* versions of the nominally purest PV, NPV and BS spectra shown in Figures 2 (a), (b) and (c) respectively as the $M = 3$ endmembers. The corresponding PV, NPV and BS proportions are shown in the legend, while the corresponding 95% confidence intervals (based on (24) and the intersection principle) are given in the caption. There are a number of things to note. First, the constrained fit is fairly good (it will get better as the model becomes more sophisticated). Second, the nominal fit is much poorer because the nominal PV and NPV proportions are respectively significantly higher and lower than the corresponding constrained estimates. Third, the unconstrained fit is poorer still.



(a) Fits **(b)** JCR

Figure 4. Nominal, constrained and unconstrained fits and 95% JCR for PV and NPV estimates for spectrum 1099 under the three-endmember PL model. 95% CIs for PV, NPV and BS proportions: (0.13, 0.32), (0.60, 1.00), (0.00, 0.18). In (b), the constrained joint CR is just the intersection of the ellipse and the triangle. An elliptical approximation to this region is shown in cyan.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

Let us look more closely at the CIs given in the caption for Figure 4 (a), which are (0.13, 0.32), (0.60, 1.00) and (0.00, 0.18) for the PV, NPV and BS proportions respectively. While the lengths of the PV and BS CIs are not too large, the length of the NPV CI is much larger. It is difficult to see how the three CIs fit together, in particular within the sum-to-one constraint (2).

The way to deal with this is to consider a joint CR for any two of the proportions; in fact this is a JCR for all three proportions because of the sum-to-one constraint. Let $p_{kl}^T \equiv (p_k, p_l)$ denote the vector of any *two* proportions k and l, let $\hat{p}_{kl,u}$ denote the corresponding vector of the unconstrained estimators of these two proportions, and let $\mathbf{V}_{kl}$ denote the submatrix consisting of rows and columns k and l of $\mathbf{V}$ (given by (17)). Then the JCR for the two proportions based on their unconstrained estimators is an ellipse given by [31, eqn. 4.60]

$$(\hat{p}_{kl,u} - p_{kl})^T \mathbf{V}_{kl}^{-1} (\hat{p}_{kl,u} - p_{kl}) / (2\hat{\sigma}^2) \leq F_{2,d-M+1,\alpha}, \tag{25}$$

where $F_{\nu_1,\nu_2,\alpha}$ is the *upper* 100α percentage point of the F distribution with ν_1 and ν_2 df. It is straightforward to show that (25) is invariant to any linear transformation of the data, and so because of the sum-to-one constraint (2), it doesn't matter which two (out of three) proportion estimates are chosen.

Despite appearances to the contrary, (25) is an extension of (24). Because the latter equation is symmetric, we can rearrange some terms and square it to obtain

$$(\hat{p}_{k,u} - p_k)^2 / (\hat{\sigma}^2 v_{kk}) \leq F_{1,d-M+1,\alpha}, \tag{26}$$

noting that $t^2_{d-M+1,\alpha/2} = F_{1,d-M+1,\alpha}$, and so the extension is apparent.

Figure 4 (b) shows the nominal, constrained and unconstrained PV and NPV estimates for spectrum 1099, the 95% joint confidence ellipse for the true PV and NPV values based on (25) and the triangle determined by the constraints (2) and (3). I will call this the *feasible* triangle. By the intersection principle, the constrained JCR is just the intersection of the ellipse and the triangle. Although the ellipse is quite large, its intersection with the triangle is much smaller, and enables us to reconcile the CIs for the individual proportions in a coherent and interpretable way. (I also show an ellipse which approximates the JCR in cyan. I will briefly discuss this approximation in Section 2.4.)

2.3.2. The Non-Negative Linear model

In this subsection, I assume the model (6), and construct CIs and JCRs for the p_k's, defined by (9), based on suitable estimators of them. These will use the two principles that were introduced in the previous subsection, namely the principle that statistical inference should be based on the *unconstrained* estimator and the intersection principle.

First some standard LS theory is required. The unconstrained LS estimators of the coefficients in (6) are given by

$$\hat{\beta} = \mathbf{F}\mathbf{E}^T X. \tag{27}$$

where $\hat{\beta}^T \equiv (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_M)$. From (8) and (9), the unconstrained proportion estimates are then given by

$$\hat{p}_u = \hat{\beta} / \hat{\gamma}, \tag{28}$$

where

$$\hat{\gamma} = \hat{\beta}^T \mathbf{1}_M = \sum_{k=1}^{M} \hat{\beta}_k. \tag{29}$$

Again using standard LS theory, the covariance matrix of $\hat{\beta}$ is given by

$$Cov(\hat{\beta}) = \sigma^2 \mathbf{F}, \tag{30}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

and the unbiased estimator of σ^2 is now given by

$$\hat{\sigma}^2 = X^T\{\mathbf{I} - \mathbf{EFE}^T\}X/(d-M). \tag{31}$$

In what follows, it will be convenient to split the vector $\hat{p}_u$ into its M separate entries:

$$\hat{p}_{k,u} = \hat{\beta}_k/\hat{\gamma}, k = 1, \ldots, M. \tag{32}$$

Note that each of the M estimators in (32) is the ratio of two random variables. When the errors have a Gaussian distribution, the general solution for the CI of a ratio is given by [34]. In *Appendix A*, I use this theory to derive the $100(1-\alpha)\%$ CI for p_k under the NNL model, which is

$$(\hat{p}_{k,u} - g_1 C_k/V_\gamma \pm F^{\frac{1}{2}}_{1,d-M,\alpha}(\hat{\sigma}/\hat{\gamma})\sqrt{\Delta_k})/(1-g_1), \tag{33}$$

where

$$\Delta_k = V_k - 2\hat{p}_{k,u}C_k + \hat{p}^2_{k,u}V_\gamma - g_1(V_k - C_k^2/V_\gamma), \tag{34}$$

$\hat{p}_{k,u}$ is given by (32), g_1 is given by

$$g_1 = F_{1,d-M,\alpha}\hat{\sigma}^2 V_\gamma/\hat{\gamma}^2, \tag{35}$$

and

$$V_k = f_{kk}, C_k = \mathbf{f}_k^T\mathbf{1}_M, V_\gamma = \delta, \tag{36}$$

where δ is given by (14), $\mathbf{f}_k$ is the kth row (or column) of $\mathbf{F}$, given by (12), and f_{kl} is the (k,l)th element of $\mathbf{F}$. It will be useful later to note that V_k and V_γ are proportional to the variances of the numerator and denominator in (32) respectively (see also (30)), while C_k is proportional to their covariance.

There are a number of things to note about the CI (33). First, it is not centred on $\hat{p}_{k,u}$, but on $\hat{p}_{k,u} - g_1 C_k/V_\gamma$. This is because the ratio estimator $\hat{p}_{k,u}$ is a biased estimator of p_k. Second, for the CI to be a "valid" CI, Δ_k in (34) needs to be positive so that its square root in (33) is real. *It is straightforward to show that, if*

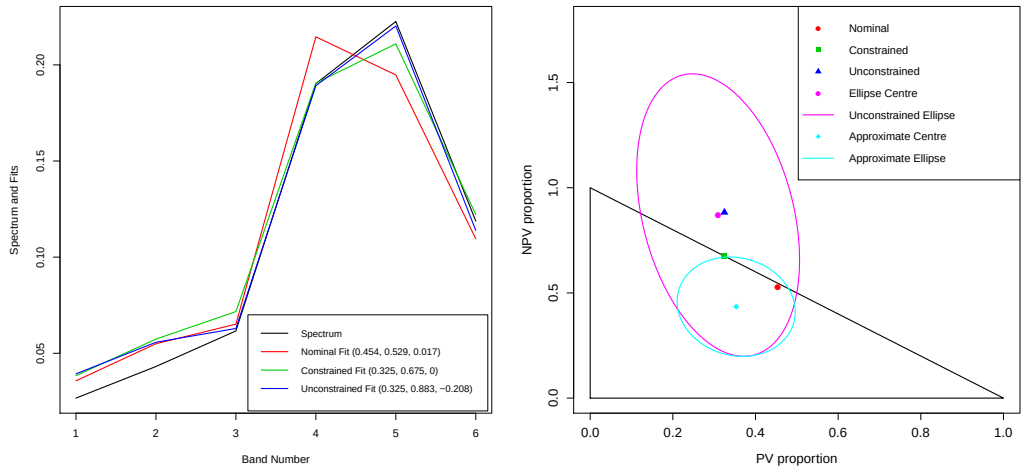
$$g_1 < 1, \tag{37}$$

then $\Delta_k > 0$. Details will not be given here. This is fortuitous, because if (37) holds, then the denominator in (33) is positive. It turns out that (37) is satisfied by all 1169 spectra in the data set.

The quantity g_1 is a useful *relative* goodness of fit measure. Excluding the first term on the right hand side of (35), the quantity $\hat{\sigma}^2 V_\gamma/\hat{\gamma}^2$ is the *estimated* variance of $\hat{\gamma}$ divided by $\hat{\gamma}^2$ (making it scale invariant). $\hat{\gamma}$ is the denominator in $\hat{p}_{k,u}$. So when the variance of the denominator is relatively large, the CI can become "invalid" (although that hasn't happened with any of the spectra in the data set being considered).

Analogous to Figure 4 (a) under the PL model, Figure 5 (a) shows the nominal, constrained and unconstrained fits for spectrum 1099 under the NNL model. As in the previous subsection, I will use the nominally purest PV, NPV and BS spectra shown in Figures 2 (a), (b) and (c) respectively as the three endmembers. However, unlike in the previous subsection, I will *not* standardise any of the endmember spectra. As previously, the constrained solution can be found using quadratic programming methods [29, Chapter 16]. For bands 1, 2 and 3, the constrained fit under the NNL model is a little worse than the constrained fit under the PL model. However, it is much better for bands 4, 5 and 6. On the other hand the unconstrained fit under the NNL model is clearly far superior to its counterpart under the PL model. This example clearly illustrates the advantage of fitting the NNL model, rather than standardising the spectra first and then fitting the PL model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1



(a) Fits **(b)** JCR

Figure 5. Nominal, constrained and unconstrained fits and 95% JCR for PV and NPV estimates for spectrum 1099 under the three-endmember NNL model. 95% CIs for PV, NPV and BS proportions: (0.17, 0.46), (0.39, 1.00), (0.00, 0.27). In (b), the constrained joint CR is just the intersection of the ellipse and the triangle. An elliptical approximation to this region is shown in cyan.

Although the unconstrained fit under the NNL model is visually much better than its counterpart under the PL model, the 95% CIs are somewhat longer; see the captions for Figures 4 and 5 for details. Although this is perhaps disappointing, it is probably more realistic.

I now turn to a JCR for any two proportions under the NNL model. As in Section 2.3.2, this is actually a JCR for all three proportions because of the sum-to-one constraint. I believe what follows to be original. In *Appendix B*, I will give an outline of the derivation of the JCR. However, I will give a more detailed derivation in a separate publication.

Let $\hat{p}_{kl,u}^T \equiv (\hat{p}_{k,u}, \hat{p}_{l,u})$ now denote a vector of two unconstrained estimators of the form (32) of $p_{kl}^T \equiv (p_k, p_l)$. The JCR based on these unconstrained estimators is based on an inequality involving a quadratic form. As pointed out previously, $\hat{p}_{k,u}$ is a biased estimator of p_k. A consequence of this is that the quadratic form is not centred on $\hat{p}_{kl,u}$. Let $\mathbf{F}_{kl}$ denote the submatrix consisting of rows and columns k and l of $\mathbf{F}$, and let $C_{kl}^T = (C_k, C_l)$, where C_k is given by (36). Let

$$q = p_{kl} - C_{kl}/V_\gamma, \hat{q} = \hat{p}_{kl,u} - C_{kl}/V_\gamma. \tag{38}$$

In its most succinct form, the $100(1-\alpha)\%$ JCR based on the unconstrained estimator $\hat{p}_{kl,u}$ is given by those values of p_{kl} satisfying

$$q^T \mathbf{A} q + b^T q + c < 0, \tag{39}$$

where

$$\mathbf{A} = \{(1-g_2) + \hat{q}^T \mathbf{B} \hat{q}\}\mathbf{B} - \mathbf{B}\hat{q}\hat{q}^T\mathbf{B}, \tag{40}$$

$$b^T = -2\hat{q}^T\mathbf{B}, \tag{41}$$

$$c = \hat{q}^T\mathbf{B}\hat{q} - g_2, \tag{42}$$

where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

$$\mathbf{B} = V_\gamma(\mathbf{F}_{kl} - C_{kl}C_{kl}^T/V_\gamma)^{-1}, \tag{43}$$

and

$$g2 = 2F_{2,d-M,\alpha}\hat{\sigma}^2 V_\gamma/\hat{\gamma}^2. \tag{44}$$

Compare (44) with (35). It can be shown that, if

$$g_2 < 1, \tag{45}$$

then (39) is the interior of an *ellipse*. A proof will be published elsewhere. Compare (45) with (37). We have

$$g2/g1 = 2F_{2,d-M,\alpha}/F_{1,d-M,\alpha}. \tag{46}$$

For the data considered in this paper $d = 6$. When $M = 3$ and $\alpha = 0.05$, $g2/g1 = 1.89$, so that the inequality (45) is more stringent that the inequality (37). Nevertheless, all 1169 spectra satisfy the inequality (45).

Figure 5 (b) shows the nominal, constrained and unconstrained PV and NPV estimates for spectrum 1099, the 95% joint confidence ellipse for the true PV and NPV values for the NNL model (based on (39)), the center of the ellipse and the feasible triangle. The X and Y ranges of the figure are the same as those for Figure 4 (b) (using the PL model), which allows for a direct comparison between the two. There are a number of things to note. First, while under the PL model the axes of the ellipse are approximately parallel to the X and Y axes, under the NNL model the ellipse is tilted. Second, the latter ellipse is somewhat shorter and fatter than the former. Third, $\hat{p}_{PV,u}$ has increased from 0.227 to 0.325. Together, these three facts mean that the nominal proportions are now inside the confidence ellipse (and the JCR), where they weren't previously. Also note the small difference between the center of the ellipse and the unconstrained estimator, due to the bias of the latter.

2.3.3. Primary and secondary endmembers

As previously, I assume that there are M endmembers, of which L are "primary" endmembers and $M - L$ are "secondary" endmembers. Without loss of generality, I will assume that the primary endmembers are the first L endmembers. An obvious example of this is where PV, NPV and BS are the primary endmembers, and (non-zero) shade and/or water are the secondary endmembers. In this case, we are interested in constructing CIs and JCRs for the *relative* proportions of the primary endmembers. The estimators of the relative proportions are just the estimators of the original proportions, divided by the sum of the original estimated proportions *of the primary endmembers only*. Hence they are ratios, as are the estimated (original) proportions under the NNL model (see (28)), and so only small adaptations of the NNL theory are required, whether one uses the PL or NNL model to begin with. Noting the comments after (36), all that is needed is to obtain new formulae for V_k, C_k and V_γ. The CI for the unconstrained estimators is then given by (33), while the JCR is given by (39).

Under the PL model, the relevant covariance matrix is given by (16). where $\mathbf{V}$ is given by (17). Let $\mathbf{V}_{L,L}$ denote the submatrix of $\mathbf{V}$ corresponding to its first L rows and L columns (corresponding to the L primary endmembers), and let $\mathbf{v}_{k,L}$ denote row k of $\mathbf{V}_{L,L}$. Then the relevant entries in (33) and (39) under the PL model are:

$$V_k = v_{kk}, C_k = \mathbf{v}_{k,L}^T\mathbf{1}_L, V_\gamma = \mathbf{1}_L^T\mathbf{V}_{L,L}\mathbf{1}_L, \tag{47}$$

where v_{kk} is the kth diagonal entry of $\mathbf{V}$ (as previously defined) and $\mathbf{1}_L$ is a vector of L1's.

Under the NNL model, the unconstrained *relative* estimated proportions are given by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

$$\hat{p}_{k,u} / \sum_{l=1}^{L} \hat{p}_{l,u} = \hat{\beta}_k / \sum_{l=1}^{L} \hat{\beta}_l, \tag{48}$$

by (32). The relevant covariance matrix is then given by (30), where $\mathbf{F}$ is given by (12). Let $\mathbf{F}_{L,L}$ denote the submatrix of $\mathbf{F}$ corresponding to its first L rows and L columns, and let $\mathbf{f}_{k,L}$ denote row k of $\mathbf{F}_{L,L}$. Then the relevant entries in (33) and (39) under the NNL model are:

$$V_k = f_{kk}, C_k = \mathbf{f}_{k,L}^T \mathbf{1}_L, V_\gamma = \mathbf{1}_L^T \mathbf{F}_{L,L} \mathbf{1}_L, \tag{49}$$

where f_{kk} is the kth diagonal entry of $\mathbf{F}$ (as previously defined).

I don't give an example of CIs and JCRs for *relative* proportions of *primary* endmembers here, because the data set analysed in this paper doesn't provide any secondary endmembers. However, I have included it because (i) it is a topic of some interest (e.g. [14–16], and (ii) the relevant theory is a relatively simple extension of the NNL model.

2.3.4. Endmember variability

This section is primarily motivated by the data set described in Section 2.1. In particular, note that in Figure 1(b), spectrum 275 is somewhat different in shape to the other two spectra, even though all three have the same nominal proportions. This is an indication that a model with *unique* PV, NPV and BS endmembers (such as (1) or (6)) is inadequate to model such variation.

A common approach to this endmember variability problem [24] is to use libraries with multiple examples of pure "spectra" drawn from each class. *Single* spectra are then drawn from each of the classes for use in the linear mixture model in such a way that a best fit is achieved according to some criterion (e.g. MESMA). Unfortunately, the approach presented in this paper is not easy to combine with this approach; see [27] and references therein.

I offer an alternative approach which can be useful in some circumstances. The idea is to model the endmembers in each class as linear mixtures of the *extreme* endmembers in that class. I will discuss how one might find these extreme endmembers shortly, but for the time being, let us assume that we have found them. Let L now denote the number of *broad* classes. Within broad class $j, j = 1, \ldots, L$, assume that there are n_j extreme endmembers. Then the total number of exteme endmembers is $M = \sum_{j=1}^{L} n_j$. In order to be able to estimate the error, we require $M \leq d$ under the PL model and $M < d$ under the NNL model, which is a significant limitation of the approach for small d.

Let $\mathcal{C}_j$ denote the indices k (between 1 and M) belonging to *broad* class $j, j = 1, \ldots, L$. Either the PL model or the NNL model is first fitted using all M extreme endmembers, and then the proportions of each *broad* class are modelled as:

$$p_j^* = \sum_{k \in \mathcal{C}_j} p_k, j = 1, \ldots, L. \tag{50}$$

There is an analogous formula for the unconstrained estimators of these parameters, $\hat{p}_j^*$, based on (11) for the PL model, and (28) for the NNL model. Equation (50) can be written in matrix notation. Let $p^T = (p_1, \ldots, p_M)$ denote the vector of notionally true proportions of the M extreme endmembers, and let $p^{*T} = (p_1^*, \ldots, p_L^*)$ denote the L broad class proportions. Let $\mathbf{H}$ denote an $L \times M$ matrix, with entry $h_{j,k}$ in row j and column k given by

$$\begin{aligned} h_{j,k} &= 1, k \in \mathcal{C}_j, \\ &= 0, \text{otherwise.} \end{aligned} \tag{51}$$

Then the matrix version of (50) is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

$$p^* = \mathbf{H}p, \tag{52}$$

with analogous formulae for the *estimated* unconstrained proportions, $\hat{p}_u^*$, under both the PL and NNL models.

For the PL model, it follows from (16) that

$$Cov(\hat{p}_u^*) \equiv \sigma^2 \mathbf{V}^* = \sigma^2 \mathbf{H}\mathbf{V}\mathbf{H}^T, \tag{53}$$

where $\mathbf{V}$ is given by (17). Then the formulae given for CIs and JCRs under the PL model in Section 2.3.1 apply with $\mathbf{V}$ everywhere replaced by $\mathbf{V}^*$. Note however that the formula for $\hat{\sigma}^2$ (19) remains unchanged.

For the NNL model, the relevant quantity is $\hat{\beta}^* = \mathbf{H}\hat{\beta}$, where $\hat{\beta}$ is given by (27). It follows from (30) that

$$Cov(\hat{\beta}^*) \equiv \sigma^2 \mathbf{F}^* = \sigma^2 \mathbf{H}\mathbf{F}\mathbf{H}^T, \tag{54}$$

where $\mathbf{F}$ is given by (12). Then the formulae given for CIs and JCRs under the NNL model in Section 2.3.2 apply with $\mathbf{F}$ everywhere replaced by $\mathbf{F}^*$. However, note that $\hat{\sigma}^2$ is still given by (31); $\mathbf{F}^*$ does *not* replace $\mathbf{F}$ in this equation.

Although the above approach has limitations when d is small, its advantage is that it can often model the variety of endmembers within a broad class in a *continuous* way. The usual approach relies on having enough endmembers in each broad class to represent *all* the variability among *mixtures* in the data set under consideration. This may not always be the case.

I now show how to apply the above theory to the data set discussed in Section 2.2 using the NNL model. Because $d = 6$, there can be at most $M = 5$ endmembers. The greater variability of the purest NPV and BS spectra in the data set (see Figures 2 (a), (b) and (c)) suggests that the model should use 1 PV, 2 NPV and 2 BS endmembers. It is not easy to find extreme endmembers in the latter two broad classes using automated methods. Fortunately, there are two significant subsets of the data set where the nominal proportion of one of the broad classes is < 0.01 (i.e. that class is almost absent): 181 spectra (15.5% of the total) have $p_{PV} < 0.01$, while 110 spectra (9.4%) have $p_{BS} < 0.01$. Plots of the first few PCs of the *standardised* spectra of these two subsets make it relatively easy to identify suitable candidates for the extreme endmembers for all three broad classes. Details of the procedure will not be given here. The five (unstandardised) endmembers found using this procedure are shown in Figure 6. Their nominal proportions are shown in the legend. While one each of the PV, NPV and BS endmembers has a nominal proportion which is either the highest or second highest nominal proportion in its broad class, the other NPV and BS endmembers have much lower nominal proportions, each a little over 0.90.

Figure 7 (a) shows the nominal, constrained and unconstrained fits for spectrum 1099 under the five-endmember NNL model. Compare this with the corresponding fits under the three-endmember NNL model in Figure 5 (a). While the constrained fit hasn't changed much (with the three proportions changing marginally from (0.325, 0.675, 0) to (0.320, 0.680, 0)), the unconstrained fit is now almost *visually* perfect (with the three proportions changing from (0.325, 0.883, -0.208) to (0.400, 0.609, -0.009)). This in turn means that $\hat{\sigma}$ is very small, which in turn means that the 95% CIs are also very small. In fact, it is so small, that the *unconstrained* CI for BS does not intersect [0, 1]. If $p_{BS} = 0$, one would expect this to happen about 2.5% of the time; another 2.5% of the time, the CI will intersect [0, 1] but exclude 0. There are two ways of dealing with this. First, one can reduce α (and hence expand the CI) until the CI intersects [0, 1]. Alternatively, we just set the CI to the nearest value in [0, 1], i.e. [0, 0], which is what I have written in the caption of Figure 7.

This problem is reflected in the 95% JCR, which is shown in Figure 7 (b). The confidence ellipse is slightly outside the triangle determined by the constraints. Nevertheless,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

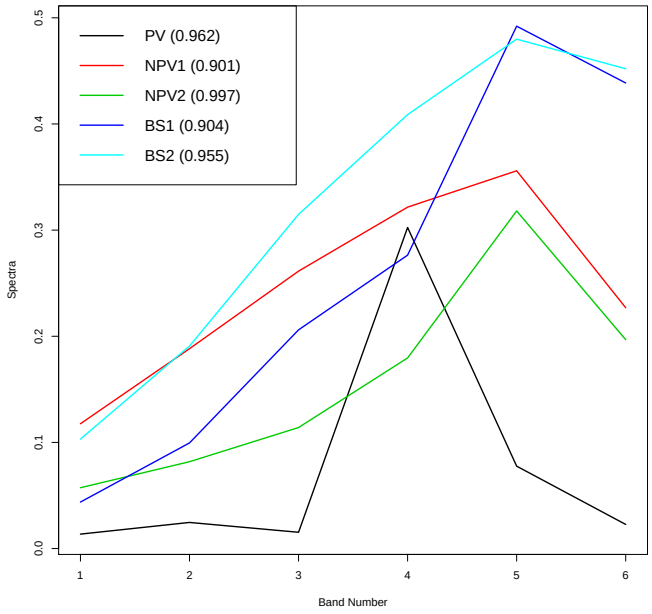
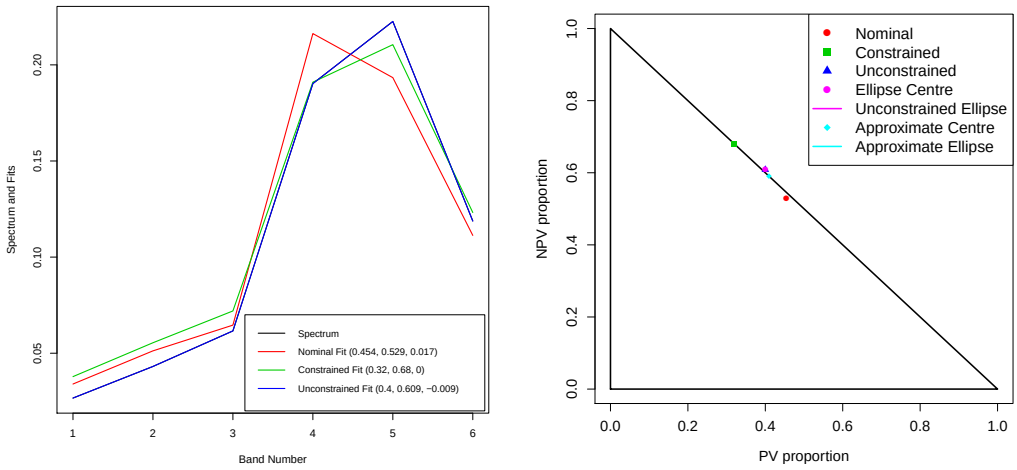


Figure 6. Endmembers used in the five endmember NNL model. Their nominal proportions are shown in the legend.

the plot gives us a good idea of where the true proportions are likely to be with a high level of confidence. More generally, the use of the five-endmember NNL model has considerably improved the fit over that produced by the three-endmember NNL model, at least for this spectrum.



(a) Fits **(b)** JCR

Figure 7. Nominal, constrained and unconstrained fits and 95% JCR for PV and NPV estimates for spectrum 1099 under the five-endmember NNL model. 95% CIs for PV, NPV and BS proportions: (0.400, 0.400), (0.608, 0.609), (0.00, 0.00). In (b), the constrained joint CR is just the intersection of the ellipse and the triangle. An elliptical approximation to this region is shown in cyan.

I have chosen spectrum 1099 as the exemplar for a number of reasons: (i) the unconstrained estimator lies outside the triangle determined by the constraints under all three models examined, (ii) the unconstrained fit under the five-endmember NNL

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

model is significantly better than the fit under the three-endmember NNL model, and (iii) under the five-endmember NNL model the 95% CI and JCR do not intersect [0, 1] and the feasible triangle respectively.

Unfortunately, there is a downside to the use of the five-endmember NNL model. Whereas all 1169 spectra satisfy the inequalities (37) and (45) (which are sufficient to ensure valid (unconstrained) CIs and JCRs) for the three-endmember NNL model, only 1101 and 926 of the spectra satisfy these inequalities respectively for the five-endmember NNL model. At first sight, this may be appear contradictory, because one would expect a better fit of the five-endmember model than for the three-endmember model. Indeed for 1093 of the 1169 spectra (93.5%), $\hat{\sigma}^2$, defined by (31), is smaller for the five-endmember model than it is for the three-endmember model. The reason that it is not 100% is partly because the endmembers in the three-endmember model are not a subset of the endmembers in the five-endmember model, but more importantly because the denominator in (31) (the df) is reduced from 3 to 1, so the five-endmember numerator needs to be considerably smaller than the three-endmember numerator to counteract this. In addition, note that in the definitions of $g1$ and $g2$ ((35) and (44) respectively), there are a number of factors, apart from $\hat{\sigma}^2$, that will change between the two models, in particular the factor $F_{l,d-M,\alpha}$, where $l = 1$ or 2 for $g1$ and $g2$ respectively. These factors are much higher when $d - M = 1$ than when $d - M = 3$. For instance, $F_{1,1,0.05}/F_{1,3,0.05} = 15.9$, while $F_{2,1,0.05}/F_{2,3,0.05} = 20.9$. So the other factors in $g1$ and $g2$ have a lot of work to do when the df is reduced from 3 to 1.

In the next section, I show how this problem can be ameliorated somewhat.

2.3.5. Relaxing the assumption (10)

In this section, it will be convenient to reintroduce the subscript i to represent spectrum i.

Up to this point, I have assumed that the errors in each band of any spectrum have the *same* variance and are uncorrelated; see (10). An examination of the ten purest PV, NPV and BS spectra (Figures 2 (a), (b) and (c) respectively) suggests that perhaps there is greater variability in bands 1, 2 and 3 than in bands 4, 5 and 6. So perhaps we should relax the assumption (10). Although I will do this shortly, for the sake of completeness, I will generalise (10) to

$$Cov(\epsilon_i) = \sigma_i^2 \Omega, i = 1, \ldots, N, \tag{55}$$

where Ω is assumed *known*, but σ_i is assumed *unknown*. It is straightforward to convert any of the three models considered so far with the assumption (55) into the analogous model with the assumption (10). This is done via an eigendecomposition of Ω:

$$\Omega = Q^T \Lambda Q, \tag{56}$$

where

$$Q^T Q = I. \tag{57}$$

Here Λ is the *diagonal* matrix of eigenvalues of Ω (which I will assume are all positive), and the columns of Q are its eigenvectors. Let

$$Z_i = \Lambda^{-1/2} Q X_i. \tag{58}$$

It follows easily from (56) and (57) that Z_i satisfies (10). So the theory of the previous three subsections will apply if we first transform X_i to Z_i via (58).

In principle, with a large enough library of *pure* spectra, it should be possible to estimate Ω and to then use the transformation (58); see for instance [22].

This is not the case with the data set discussed in Section 2.2. However, some progress is possible if one is prepared to assume that Ω is *diagonal*, i.e. the errors in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

different bands are uncorrelated. (Note that in this case $Q = I$ and $\Lambda = \Omega$.) I will call this the *variable* error variance model, and the model (10) the *constant* error variance model. Denote the diagonal entries of Ω by $\omega_1, \omega_2, \ldots, \omega_d$. For the five-endmember NNL model, how might one estimate these? A relatively simple, if crude, method is as follows. For spectrum i, consider the vector of residuals obtained from the constant error variance model $\rho_i = (\rho_{i1}, \rho_{i2}, \ldots, \rho_{id})$, given by

$$\rho_i = X_i - E\hat{\beta}_i, \tag{59}$$

where $\hat{\beta}_i$ is given by (27). If we square ρ_{ij}, each should on average be approximately proportional to ω_j. However, brighter spectra will tend to have larger residuals than darker spectra, so we should divide ρ_{ij}^2 by $\hat{\gamma}_i^2$, given by (29), and then average this statistic over all the fitted spectra in the data set, i.e.

$$\omega_j = \sum_{i=1}^{N} \rho_{ij}^2 / \hat{\gamma}_i^2, j = 1, \ldots, d. \tag{60}$$

I have done this for the five endmember model and all 1164 non-endmember spectra. The values of $\omega_j^{1/2} (\times 10^{-5})$ (the actual divisors of each entry of X_i) are: 271, 368, 147, 12.5, 4.1, 16.1. As expected, these are much larger for bands 1, 2 and 3 than they are for bands 4, 5 and 6.

When these values are used to produce Z_i, instead of X_i (equation (58)), the number of spectra satisfying (37) increases from 1101 (94.2%) to 1169 (100%), while the number of spectra satisfying (44) increases from 926 (79.2%) to 1148 (98.2%).

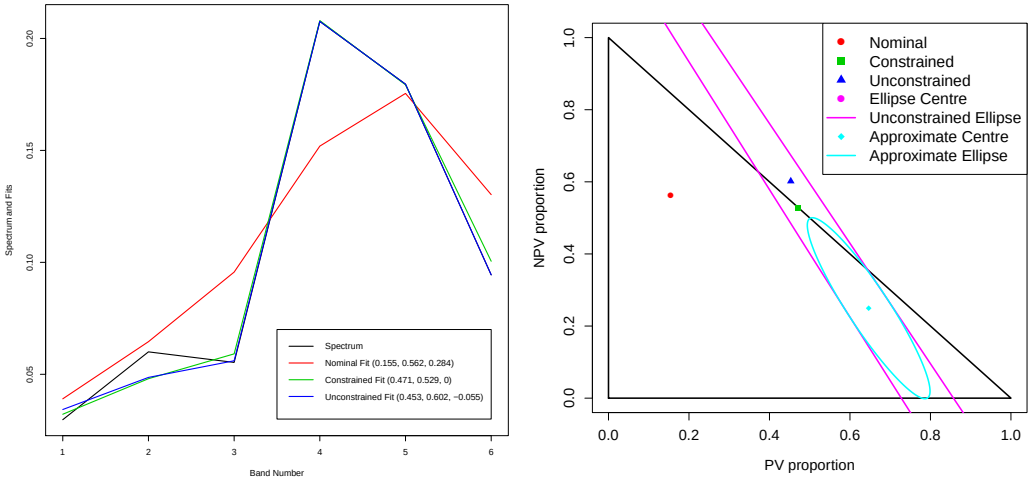
For spectrum 1099, the *unconstrained* estimators for the variable error variance model are the same as they are for the constant error variance model (to three decimal places): (0.400, 0.609, -0.009), while the *constrained* estimators change a little from (0.320, 0.680, 0.000) to (0.285, 0.715, 0.000). The CI and JCR (based on the unconstrained estimators) are also very similar to those shown in Figures 7 (a) and (b) respectively, so they are not shown here. Instead, I show the analogous figures for spectrum 77 in Figures 8 (a) and (b). This is the spectrum with the largest value of $g2$ less than 1 (0.995). Both the unconstrained and constrained estimators fit the spectrum quite well, except in band 2, the green (chlorophyll) band. This raises three issues. First, the green peak in Figure 8 (a) is much higher than the corresponding peaks of the ten purest PV spectra in the data set; see Figure 2 (a). So they are not sufficiently representative of "very green" PV spectra. The second issue is that, even were such samples present in the dataset, we would need both "green" and "very green" endmembers, but then df = 0, and so it would be impossible to generate CIs and JCRs. The third issue is that the nominal fit is much poorer, raising doubts about the accuracy of the nominal proportions. Unfortunately, because of the (relatively) poor fit in band 2, the individual 95% CIs for the three proportions are each [0, 1], which is totally uninformative. However, the JCR (shown in Figure 8 (b)) is much more informative. The confidence ellipse is long and narrow, so I have only shown the plot in the vicinity of the feasible triangle. It intersects with a relatively small part of the triangle, and so localises the JCR considerably. This example is a very good demonstration of the value of the JCR over the individual CIs.

2.4. Application to Multispectral Image Data

The intersection of an ellipse and a triangle cannot be represented algebraically. This is a drawback if one has multispectral *image* data. Such data are themselves usually summarised by a smaller number of images (e.g. NDVI images) for ease of interpretation. It would therefore be useful if the JCR (which, as can be seen in various examples above, is much more informative than three separate CIs) could be approximated by a small number of parameters which can be displayed as color images.

The basic idea presented in this section is to approximate the JCR itself by an ellipse. I will differentiate the "approximating" ellipse from the ellipse intersecting the feasible

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1



(a) Fits **(b)** JCR

Figure 8. Nominal, constrained and unconstrained fits and 95% JCR for PV and NPV estimates for spectrum 77 under the five-endmember NNL model. 95% CIs for PV, NPV and BS proportions: [0, 1], [0, 1], [0, 1]. In (b), the constrained joint CR is just the intersection of the ellipse and the triangle. An elliptical approximation to this region is shown in cyan.

triangle by calling the latter the "unconstrained" ellipse. Examples of the approximating ellipse are shown (in cyan) in Figures 4 (b), 5 (b), 7 (b) and 8 (b).

Although the idea is easy to explain, there are a number of different ways in which an ellipse can (or cannot) intersect a triangle, and it can be tedious to implement them all. The *vast majority* of examples fall into one of five categories. Three of these categories have been illustrated in this paper. Figures 4 (b) and 5 (b) exemplify the category where the unconstrained ellipse intersects the boundary of the triangle in two places and has one end inside and one end outside the triangle. Figure 7 (b) exemplifies the category where the unconstrained ellipse does not intersect the triangle at all. Figure 8 (b) exemplifies the category where the unconstrained ellipse intersects the boundary of the triangle in four places. The two other main categories are: (i) the unconstrained ellipse lies entirely inside the feasible triangle, and (ii) one edge of the unconstrained ellipse intersects the boundary of the triangle in two places (on two different sides), but both ends are outside the triangle. Typically, this happens near a corner of the feasible triangle.

An example of this category is shown in [35, Section 5.1]. I have also found rare pathological examples in the data set discussed in this paper, as well as two other data sets. One of these pathological examples is also shown in [35, Section 5.1].

As the examples shown in this paper illustrate, the approximating ellipse usually provides a reasonable approximation to the intersection of the unconstrained ellipse and the feasible triangle.

Once the major and minor axis lengths and the orientation of the approximating ellipse have been calculated at each pixel, they can be displayed as a colour image, as can three proportion estimates, either the constrained estimates or the centres of the approximating ellipse. There is extensive discussion about this issue in [35, Section 5.2]. Various issues are illustrated with the aid of two small image data sets.

3. Discussion and Conclusions

In this paper, I have demonstrated how to construct valid CIs and JCRs for proportions for several versions of the three-endmember linear mixture model. I have focussed on the three-endmember model, because there are many papers in the remote sensing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

literature which use it, especially for multispectral data, and also because JCRs for such models are easy to visualise, because the problem is essentially two dimensional.

The CIs and JCRs are based on two principles. The first is the principle that they should be based on the *unconstrained* estimator (i.e. ignoring the non-negativity constraint (3)) because the *constrained* estimator throws away important information about the variability of the estimator; see Figure 3. I call the second principle the Intersection Principle. Because of the constraints (2) and (3), CIs don't always provide very useful information. JCRs can often overcome this issue; see for instance Figures 4 (a) and (b), 5 (a) and (b), and 8 (a) and (b) (and the captions of these figures).

The first model that I analysed (in Section 2.3.1), the PL model, assumes that the coefficients in the model are proportions. However, unless the spectra in the data set discussed in Section 2.1 are first standardised, this model doesn't fit the spectra very well because of brightness variations in spectra with the same nominal PV/NPV/BS proportions; see Figures 1(a) and 1(b). In order to overcome this problem, I introduced and analysed (in Section 2.3.2) the NNL model (6), in which the coefficients are still constrained to be non-negative, but they are no longer constrained to sum to 1. The proportions are then estimated via *ratio* estimators. CIs for such estimators have previously been derived by [34]. I have extended this theory to provide JCRs under this model. A more detailed analysis will be published elsewhere. Unfortunately, when inequalities (37) and (45) aren't satisfied, the CIs and JCRs are respectively no longer valid. This happens when the unconstrained estimator is not sufficiently well fitted. For both the PL and NNL models, it is assumed (for simplicity) that the error variance is the same in all bands; see (10).

In the following three subsections, I analysed three variants of these two fundamental models. In Section 2.3.3, I considered the case where there are three *primary* endmembers and some *secondary* endmembers (i.e. of lesser interest), while in Section 2.3.4, I demonstrated how in some cases endmember variability can be modelled under either the PL or NNL framework. I exemplified this using an NNL model with 1 PV, 2 NPV and 2 BS endmembers. In Section 2.3.5, I showed how to relax the assumption (10).

Returning to the inequalities (37) and (45) for the NNL model, the issue is essentially a df ($= d - M$) problem. For the three-member NNL model (with df = 3), all 1169 spectra in the exemplar data set satisfy both inequalities when using a basic three-endmember model (in Section 2.3.2). However, when using the five-endmember model (with df = 1, in Section 2.3.4), not all the spectra satisfy the inequalities, although the situation is improved somewhat in Section 2.3.5 by using a more suitable error covariance structure.

Of course, this problem could be overcome by using a sensor with additional bands. An obvious sensor is MODIS (at least the seven bands with 500 m resolution or better). However, airborne hyperspectral sensors such as AVIRIS and HyMap (with 224 and 126 bands respectively) provide much greater redundancy in modelling and hence degrees of freedom! In such a case, the chances of the inequalities (37) and (45) being satisfied are greatly increased; see the discussion at the end of Section 2.3.4.

Hyperspectral sensors also (potentially) provide an opportunity to examine JCRs for more than three materials, even when some of these materials are considered to be variants within the same endmember class, using the theory presented in Sections 2.3.4 and 2.3.5. However, there are two issues to address here. First, there is the visualisation issue. When there are four (possibly broad) classes, because of the sum-to-one constraint (2), the JCR will need to be displayed in 3D, which is possible because of the ready availability these days of 3D visualisation software. Visualising JCRs for more than four materials is more difficult. However, there is a more fundamental issue: is the necessary theory available to obtain JCRs for more than three materials? For the PL model, it is straightforward to generalise (25) provided $M \leq d$. No such theory has been published for the NNL model for more than three materials. However, using fairly sophisticated matrix algebra, I have been able to generalise (39) provided $M < d$. I will publish the relevant theory elsewhere.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

Returning to the multispectral case, the JCR is just the intersection of the unconstrained ellipse and the feasible triangle (when they intersect). This intersection cannot be represented algebraically, which presents difficulties when summarising the information in multispectral *image* data. So in Section 2.4, I discussed how to approximate the intersection by another ellipse. The parameters of the approximating ellipse can be summarised by two colour images, one representing the center of the approximating ellipse, and the other the orientation and major and minor axis lengths of the ellipse.

Funding: This research received no external funding.

Acknowledgments: I would like to thank Juan Pablo Guerschman for bringing the confidence interval/region problem to my attention and for providing the main data set analysed in this paper. I would also like to thank Tony Traylen for his comments on an earlier version of this paper.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BS	Bare soil
CI	Confidence interval
JCR	Joint confidence region
LS	Least squares
MESMA	Multiple endmember spectral mixture analysis
NNL	Non-negative linear
NPV	Non-photosynthetic vegetation
PL	Proportion linear
PV	Photosynthetic vegetation
TM	Thematic Mapper
VIS	Vegetation-impervious surface-soil

Appendix A. Outline of derivation of (33)

Following [34], consider the quantity

$$r_k = \hat{\beta}_k - p_k\hat{\gamma}, \tag{A1}$$

where $\hat{\beta}_k$ is component k of $\hat{\beta}$, given by (27), and $\hat{\gamma}$ is given by (29). It follows from standard LS theory that $E(\hat{\beta}_k) = \beta_k$, and hence from (29) and (8) that $E(\hat{\gamma}) = \gamma$. It then follows from (9) and (A1) that

$$E(r_k) = 0. \tag{A2}$$

We next need to calculate the variance of r_k. it follows from (30) and (A1) that

$$Var(r_k) \equiv \sigma^2 w_{kk} = \sigma^2(V_k - 2p_kC_k + p_k^2V_\gamma), \tag{A3}$$

where V_k, C_k and V_γ are given by (36).

From standard statistical theory,

$$t_k \equiv (\hat{\beta}_k - p_k\hat{\gamma})/\{\hat{\sigma}(V_k - 2p_kC_k + p_k^2V_\gamma)^{\frac{1}{2}}\} \tag{A4}$$

has a t distribution with $d - M$ df, where $\hat{\sigma}^2$ is given by (31). Equivalently,

$$S_{1,k} \equiv (\hat{\beta}_k - p_k\hat{\gamma})^2/\{\hat{\sigma}^2(V_k - 2p_kC_k + p_k^2V_\gamma)\} \tag{A5}$$

has an F distribution with 1 and $d - M$ df. Hence, with probability $1 - \alpha$,

$$S_{1,k} \leq F_{1,d-M,\alpha}, \tag{A6}$$

where $F_{\nu_1,\nu_2,\alpha}$ is the *upper* 100α percentage point of the F distribution with ν_1 and ν_2 df.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 January 2023 doi:10.20944/preprints202301.0423.v1

This inequality provides the means of deriving a CI for the *unknown* parameter p_k. However, the problem is non-standard because p_k occurs in both the numerator and denominator in (A5). Fortunately, both the numerator and denominator are *quadratic* in p_k. So we can substitute the right hand side of (A5) into the left hand side of (A6), and multiply both sides by the denominator to obtain a quadratic inequality in p_k. This can be solved by standard means to obtain the bounds of the CI, which are given by (33).

Appendix B. Outline of derivation of (39)

The derivation generalises the approach taken in *Appendix A*. Let $\mathbf{R}_{kl} = (r_k, r_l)$, where r_k and r_l are given by (A1) and let $\sigma^2\mathbf{W}_{kl}$ denote the covariance matrix of $\mathbf{R}_{kl}$ (details of $\mathbf{W}_{kl}$ will be given shortly). Then generalising the statistic (A5)

$$S_{2,kl} \equiv \mathbf{R}_{kl}^T \mathbf{W}_{kl}^{-1} \mathbf{R}_{kl} / (2\hat{\sigma}^2) \tag{A7}$$

has an F distribition with 2 and $d - M$ df, where $\hat{\sigma}^2$ is given by (31). Hence, with probability $1 - \alpha$,

$$S_{2,kl} \leq F_{2,d-M,\alpha}. \tag{A8}$$

Let w_{kl} denote the (k,l)th element of $\mathbf{W}_{kl}$. w_{kk} (and hence w_{ll}) are given by (A3), while from (A1), w_{kl} is given by

$$w_{kl} = V_{kl} - p_k C_l - p_l C_k + p_k p_l V_\gamma, \tag{A9}$$

where V_γ and C_k (and hence C_l) are given by (36), and

$$V_{kl} = f_{kl}, \tag{A10}$$

the (k,l)th element of $\mathbf{F}$, given by (12). Using a standard formula for the inverse of a 2×2 matrix,

$$\mathbf{W}_{kl}^{-1} = \begin{pmatrix} w_{ll} & -w_{kl} \\ -w_{kl} & w_{kk} \end{pmatrix} / (w_{kk} w_{ll} - w_{kl}^2). \tag{A11}$$

After substituting (A11) into (A7), and then substituting (A7) into (A8), multiplying both sides of the inequality by the denominator on the left hand side, and using (A3) and (A9), one obtains a quadratic inequality. Its most succinct form is given by (39).

References

1. Asner, G.P.; Heidebrecht, K.B. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. *Inuternational Journal of Remote Sensing* **2002**, *23*, 3939–3958. doi:10.1080/01431160110115960.
2. Elmore, A.J.; Asner, G.P.; Hughes, R.F. Satellite monitoring of vegetation phenology and fire fuel conditions in Hawaiian drylands. *Earth Interactions* **2005**, *9*, 1–21. doi:10.1175/EI160.1.
3. Gill, T.K.; Phinn, S.R. Improvements to ASTER-derived fractional estimates of bare ground in a savanna rangeland. *IEEE Transactions on Geoscience and Remote Sensing* **2009**, *47*, 662–670. doi:10.1109/TGRS.2008.2004628.
4. Guerschman, J.P.; Hill, M.J.; Renzullo, L.J.; Barrett, D.J.; Marks, A.S.; Botha, E.J. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment* **2009**, *113*, 928–945. doi:10.1016/j.rse.2009.01.006.
5. Scarth, P.; Röder, A.; Schmidt, M.; Denham, R. Tracking grazing pressure and climate interaction-The role of Landsat fractional cover in time series analysis. Proceedings of the 15th Australasian Remote Sensing and Photogrammetry Conference, Alice Springs, Australia, 2010, pp. 13–17. doi:10.6084/m9.figshare.94250.
6. Guerschman, J.P.; Oyarzabal, M.; Malthus, T.; McVicar, T.; Byrne, G.; Randall, L.; Stewart, J. Evaluation of the MODIS-based vegetation fractional cover product. Technical report, CSIRO Land and Water, 2012.
7. Okin, G.S.; Clarke, K.D.; Lewis, M.M. Comparison of methods for estimation of absolute vegetation and soil fractional cover using MODIS normalized BRDF-adjusted reflectance data. *Remote Sensing of Environment* **2013**, *130*, 266–279. doi:10.1016/j.rse.2012.11.021.
8. Guerschman, J.P.; Scarth, P.F.; McVicar, T.R.; Renzullo, L.J.; Malthus, T.J.; Stewart, J.B.; Rickards, J.E.; Trevithick, R. Assessing the effects of site heterogeneity and soil properties when unmixing photosynthetic vegetation, non-photosynthetic vegetation and bare soil fractions from Landsat and MODIS data. *Remote Sensing of Environment* **2015**, *161*, 12–26. doi:10.1016/j.rse.2015.01.021.

9. Hill, M.J.; Zhou, Q.; Sun, Q.; Schaaf, C.B.; Southworth, J.; Mishra, N.B.; Gibbes, C.; Bunting, E.; Christiansen, T.B.; Crews, K.A. Dynamics of the relationship between NDVI and SWIR32 vegetation indices in southern Africa: implications for retrieval of fractional cover from MODIS data. *International Journal of Remote Sensing* **2016**, *37*, 1476–1503.
10. Hill, M.J.; Zhou, Q.; Sun, Q.; Schaaf, C.B.; Palace, M. Relationships between vegetation indices, fractional cover retrievals and the structure and composition of Brazilian Cerrado natural vegetation. *International Journal of Remote Sensing* **2017**, *38*, 874–905.
11. Wang, G.; Wang, J.; Zou, X.; Chai, G.; Wu, M.; Wang, Z. Estimating the fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil from MODIS data: Assessing the applicability of the NDVI-DFI model in the typical Xilingol grasslands. *International Journal of Applied Earth Observation and Geoinformation* **2019**, *76*, 154–166.
12. Zheng, G.; Bao, A.; Li, X.; Jiang, L.; Chang, C.; Chen, T.; Gao, Z. The potential of multispectral vegetation indices feature space for quantitatively estimating the photosynthetic, non-photosynthetic vegetation and bare soil fractions in Northern China. *Photogrammetric Engineering & Remote Sensing* **2019**, *85*, 65–76.
13. Okin, G.S. Relative spectral mixture analysis - A multitemporal index of total vegetation cover. *Remote Sensing of Environment* **2007**, *106*, 467–479. doi:10.1016/j.rse.2006.09.018.
14. Ballantine, J.A.C.; Okin, G.S.; Prentiss, D.E.; Roberts, D.A. Mapping North African landforms using continental scale unmixing of MODIS imagery. *Remote Sensing of Environment* **2005**, *97*, 470–483. doi:10.1016/j.rse.2005.04.023.
15. Schmidt, M.; Scarth, P. Spectral mixture analysis for ground-cover mapping. In *Innovations in remote sensing and photogrammetry*; Springer, 2009; pp. 349–359. doi:10.1007/978-3-540-93962-7_27.
16. Quintano, C.; Fernández-Manso, A.; Roberts, D.A. Multiple Endmember Spectral Mixture Analysis (MESMA) to map burn severity levels from Landsat images in Mediterranean countries. *Remote Sensing of Environment* **2013**, *136*, 76–88.
17. Ridd, M.K. Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. *International journal of remote sensing* **1995**, *16*, 2165–2185.
18. Ward, D.; Phinn, S.R.; Murray, A.T. Monitoring growth in rapidly urbanizing areas using remotely sensed data. *The Professional Geographer* **2000**, *52*, 371–386.
19. Phinn, S.; Stanford, M.; Scarth, P.; Murray, A.; Shyy, P. Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques. *International Journal of Remote Sensing* **2002**, *23*, 4131–4153.
20. Wu, C.; Murray, A.T. Estimating impervious surface distribution by spectral mixture analysis. *Remote sensing of Environment* **2003**, *84*, 493–505.
21. Weng, Q.; Lu, D. Landscape as a continuum: an examination of the urban landscape structures and dynamics of Indianapolis City, 1991–2000, by using satellite images. *International Journal of Remote Sensing* **2009**, *30*, 2547–2577.
22. Xu, F.; Cao, X.; Chen, X.; Somers, B. Mapping impervious surface fractions using automated Fisher transformed unmixing. *Remote Sensing of Environment* **2019**, *232*, 111311.
23. Bateson, C.A.; Asner, G.P.; Wessman, C.A. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing* **2000**, *38*, 1083–1094. doi:10.1109/36.841987.
24. Somers, B.; Asner, G.; Tits, L.; Coppin, P. Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment* **2011**, *115*, 1603–1616.
25. Roberts, D.A.; Gardner, M.; Church, R.; Ustin, S.; Scheer, G.; Green, R.O. Mapping chaparral in the Santa Monica Mountains using multiple endmember specral mixture models. *Remote Sensing of Environment* **1998**, *65*, 267–279. doi:10.1016/S0034-4257(98)00037-6.
26. Asner, G.P.; Lobell, D.B. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sensing of Environment* **2000**, *74*, 99–112. doi:10.1016/S0034-4257(00)00126-7.
27. Leeb, H.; Pötscher, B.M.; Ewald, K. On various confidence intervals post-model-selection. *Statistical Science* **2015**, *30*, 216–227.
28. Settle, J.; Campbell, N. On the errors of two estimators of sub-pixel fractional cover when mixing is linear. *IEEE Transactions on Geoscience and Remote Sensing* **1998**, *36*, 163–170. doi:10.1109/36.655326.
29. Nocedal, J.; Wright, S. *Numerical Optimization*; Springer: New York, 2006. doi:10.1007/978-0-387-40065-5.
30. Settle, J.J.; Drake, N.A. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing* **1993**, *14*, 1159–1177. doi:10.1080/01431169308904402.
31. Rawlings, J.O.; Pantula, S.G.; Dickey, D.A. *Applied Regression Analysis: a Research Tool*; Springer Science & Business Media, 2001.
32. Jiang, J.H.; Liang, Y.Z.; Ozaki, Y. On simplex-based method for self-modeling curve resolution of two-way data. *Chemometrics and Intelligent Laboratory Systems* **2003**, *65*, 51–65. doi:10.1016/S0169-7439(02)00103-X.
33. Rust, B.W.; O'Leary, D.P. Confidence intervals for discrete approximations to ill-posed problems. *Journal of Computational and Graphical Statistics* **1994**, *3*, 67–96. doi:10.2307/1390796.
34. Fieller, E.C. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B* **1954**, *16*, 175–185.
35. Berman, M. Confidence Intervals and Regions for Proportions Under Various Three-Endmember Linear Mixture Models. Technical Report EP2023-0047, CSIRO DATA61, 2023.