COMPARISON BETWEEN STATISTICAL AND OPTIMIZATION METHODS IN ACCESSING UNMIXING OF SPECTRALLY SIMILAR MATERIALS

Pravesh Debba
CSIR Built Environment, Logistics and Quantitative Methods, P.O. Box 395, Pretoria, 0001, South Africa

James Maina
CSIR Built Environment, Transport Infrastructure Engineering, P.O. Box 395, Pretoria, 0001, South Africa

Elias Willemse
CSIR Built Environment, Logistics and Quantitative Methods, P.O. Box 395, Pretoria, 0001, South Africa

ABSTRACT

This paper reports on the results from ordinary least squares and ridge regression as statistical methods, and is compared to numerical optimization methods such as the stochastic method for global optimization, simulated annealing, particle swarm optimization and limited memory Broyden-Fletcher-Goldfard-Sharon bound optimization method. We used each of the above mentioned methods in estimating the abundances of spectrally similar iron-bearing oxide/hydroxide/sulfate minerals in complex synthetic mixtures simulated from hyperspectral data. In evaluating the various methods, spectral mixtures were generated with varying linear proportions of individual spectra from the United States Geological Survey (USGS) spectral library. We conclude that ridge regression, simulated annealing and particle swarm optimization outperforms ordinary least squares method and the stochastic method for global optimization algorithms in estimating the partial abundance of each endmember. This result was independent of the error from either a uniform or gaussian distribution. For large remote sensing scenes, typically with millions of pixels and with many endmembers, we recommend using ridge regression.

1. BACKGROUND AND OBJECTIVE

Remote sensors often record scenes in which the spectral signatures of various materials on the ground contribute to the spectrum measured from a single pixel. This can occur due to two reasons, (i) the spatial resolution of the image is low and adjacent objects can jointly occupy a single pixel and the resulting spectrum will be a composite of the individual objects, and (ii) distinct materials on the ground are combined into complex mixtures, for example, mixtures of minerals in the ground, which can occur regardless of the spatial resolution of the sensor (Keshava, 2003). Given such mixed pixels, the objective of unmixing (or sometimes known as "abundance estimation" or "fractional estimation") is to identify the individual constituent materials present in the mixture, as well as the proportions in which they appear. Spectral unmixing is the procedure by which the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, or more commonly known in the field of remote sensing as endmembers, and a set of corresponding fractions, or abundances, that indicate the proportion of each endmember present in the pixel.

Spectral unmixing of hyperspectral remote sensing images is useful in determining abundances of different minerals. Most spectral unmixing techniques are variants of algorithms involving matrix inversion (Van der Meer & De Jong, 2000; Peddle & Smith, 2005; Miao et al., 2006; Settle, 2006). A major problem in spectral unmixing is the non-orthogonality of endmembers. The ability to estimate abundances in complex mixtures through spectral unmixing techniques is further complicated when considering very similar spectral signatures (Debba et al., 2006). It is known that iron-bearing oxide/hydroxide/sulfate minerals have similar spectral signatures and it is therefore difficult to estimate these abundances.

Over the last two decades, several different unmixing models have been implemented, including least squares methods (Quarmby et al., 1992; Settle & Drake, 1993; Adams et al., 1995), neural networks (Atkinson et al., 1997; Liu et al., 2004), fuzzy classifiers (Foody, 1996), regression and decision trees (DeFries et al., 1999), support vector regression (Walton, 2008), gaussian mixture discriminant analysis (Ju et al., 2003) and maximum likelihood classifiers (Foody et al., 1992; Schowengerdt,

Debba, P. (email pdebba@csir.co.za) is the corresponding and presenting author. Thanks to CSIR for funding.

1996). More recently, focus is on feature selection or feature extraction and using derivatives prior to spectral unmixing (Debba et al., 2006; Somers et al., 2009).

This paper reports on the results from several statistical and optimization methods in estimating the abundances of spectrally similar iron-bearing oxide/hydroxide/sulfate minerals in complex synthetic mixtures using hyperspectral data.

In using the various methods, spectral mixtures were generated with varying linear proportions of individual spectra of a set of iron-bearing oxide/hydroxide/sulfate minerals. The set of endmembers is commonly associated with sulphide-bearing mine wastes. Prior to unmixing, the mixed spectrum was first subjected to error from a uniform and gaussian distribution with signal-to-noise ratio of 500:1, which is typical for sensors like HyMap. Discussions on linear and non-linear mixtures and a literature survey on unmixing can be found in Keshava & Mustard (2002) and Keshava (2003).

2. METHODS OF SPECTRAL UNMIXING

Spectral unmixing is a deconvolution process for estimating the contribution of individual endmembers. If we have K spectral bands, and we denote the ith endmember spectrum as $\mathbf{s}_i$ $(K \times 1)$ and the abundance of the ith endmember as a_i, the observed spectrum $\mathbf{x}$ $(K \times 1)$ for any pixel in the scene can be expressed as

$$\begin{aligned} \mathbf{x} &= a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + \cdots + a_M\mathbf{s}_M + \mathbf{w} \\ &= \sum_{i=1}^{M} a_i\mathbf{s}_i + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w} \end{aligned} \tag{1}$$

where M is the number of endmembers, $\mathbf{S}$ $(K \times M)$ is the matrix of endmembers, and $\mathbf{w}$ $(K \times 1)$ is an error term accounting for additive noise (including sensor noise, endmember variability, and other model inadequacies). This model for pixel synthesis is the linear mixing model (LMM). In the past, the abundances have been commonly calculated using (i) an unconstrained linear spectral unmixing (LSU) algorithm, where the abundances are unconstrained and can assume any numeric value, (ii) a constrained LSU, where the sum of abundances equals one, and (iii) a fully constrained LSU, where the sum of abundances equals one and the abundance values cannot be less than zero. To be physically realizable, in this paper we consider the latter, namely, the abundance coefficients should be nonnegative and should sum to one.

Hence a solution to estimating $\mathbf{a}$ $(M \times 1)$ in Equation 1 is to minimize some function of the error term, for example, minimizing the sum of the squared errors i.e. $\mathbf{w}^T\mathbf{w}$. A solution using ordinary least squares (OLS) will be $\widehat{a} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{x}$. Under model assumptions, the OLS is the best linear unbiased estimator. Lagrange multipliers can be used to accommodate multiple constraints. This has been explained in detail by Settle & Drake (1993). However, it is well known that in case of near multicollinearity, where the matrix $\mathbf{S}^T\mathbf{S}$ is nearly singular (ill-conditioned), the OLS may perform poorly. Ridge Regression (RR), which was introduced by Hoerl & Kennard (1970) is one method that has been developed to deal with this problem. In RR, a real number $\delta \geq 0$ is added to the elements on the diagonal of the matrix to be inverted, yielding a modified estimator for $\mathbf{a}$ as $\widehat{a} = (\mathbf{S}^T\mathbf{S} + \delta\mathbf{I})^{-1}\mathbf{S}^T\mathbf{x}$ for $\delta \geq 0$. Again, in a similar way to the above, Lagrange multipliers can be used to accommodate multiple constraints.

Global optimization methods can be classified as deterministic, stochastic and hybrid strategies. For each optimization method, we minimized the sum of the squared errors, namely, we minimized

$$f = \mathbf{w}^T\mathbf{w} \tag{2}$$

so that we can also compare the results with OLS and RR solutions.

Stochastic Method for Global Optimization (SMGO), introduced by Boender et al. (1982), is a hybrid optimization strategy that combines global and local optimization methods. SMGO successfully applies local search to different starting points, resulting in a pool of local minima that may, potentially, include the global optimum. What distinguishes SMGO from a pure multisearch strategy is that it involves a combination of sampling, clustering and local search to increase the likelihood that the pool of local minima will include the global optimum.

For the spectral unmixing problem SMGO works as follows. The objective value of each possible solution, referred to as a point in the SMGO framework, is calculated through f. Let X^* be the set containing all the local minima points that were found so far, and let X_1 be the set of points that after the application of local search resulted in local minimum that have already been found in a previous iteration. Initially both sets are empty. SMGO then goes through the following five steps.

Step 1: Generate N points with uniform distribution U, and add the points to the cumulative sample C, which is initially empty.

Step 2: Transform the sample to T by taking the τ percentage of points with the lowest objective values, calculated with f, and improving each point through the Broydena-Fletchera-Goldfarba-Shanno (BFGS) local search algorithm (Csendes et al., 2008).

Step 3: Apply the single linkage clustering algorithm (Everitt, 1974) to T using first the elements of X^*, followed by elements of X_1. If all the elements of T can be assigned to a cluster then go to Step 5.

Step 4: Starting with the best solution not yet clustered, x_1, apply the BFGS local search algorithm and let x^* be the result. If $x^* \in X^*$, add x^* to X^* and choose x^* as the next seed point. Otherwise add x_1 to X_1 and choose x_1 as the seed point. Note that x_1 may not be a local optimum. After choosing a new seed point apply the clustering procedure. Repeat Step 4 until all points have been assigned to a cluster. If a new local minimum has been found, i.e., a point has been added to X^*, then go to Step 1. Otherwise go to Step 5.

Step 5: Return the best solution found, i.e. the point in X^* with the best objective value, and stop.

For a detailed description of SMGO we refer the reader to Boender et al. (1982) and Csendes et al. (2008).

Simulated Annealling (SA) is a general optimization method that has been widely applied to find the global optimum of an objective function called the fitness function $f(\omega)$ (defined in Equation 2 for this research) (Kirkpatrick et al., 1983; Bohachevsky et al., 1986; Aarts & Korst, 1989). The fitness function depends on the configuration of the estimates a_i, corresponding to ω that is to be minimized. As such, SA is a computer intensive search technique to find the optimum value of a function of the absolute difference between an image (mixed) spectra and a linearly combined reference spectrum, by continually updating this function at successive steps (Aarts & Korst, 1989). The problem of non-orthogonality in matrix inversion is thus avoided and reduced to solving a finite state space combinatorial problem. Unmixing of image spectra by means of optimization was previously addressed by applying simulated annealing (Penn, 2002; Debba et al., 2006) and by using a genetic optimization algorithm (Linforda & Platzman, 2004).

Starting with a random configuration of a_i, $f(\omega^0)$ is calculated. Let ω^i and ω^{i+1} represent two solutions with fitness $f(\omega^i)$ and $f(\omega^{i+1})$, respectively. Configuration ω^{i+1} is derived from ω^i by randomly replacing one point a_j^{old} of ω^i by a new point in a_j^{new} in $\left[0, a_j^{\text{old}} + a_k^{\text{old}}\right]$, where a_k^{old} is another randomly chosen point, but $a_k^{\text{new}} = a_j^{\text{old}} - a_j^{\text{new}} + a_k^{\text{old}}$ so that $\sum a_i = 1$. A probabilistic acceptance criterion decides whether ω^{i+1} is accepted or not. This probability $P_c(\omega^i \rightarrow \omega^{i+1})$ of ω^{i+1} being accepted is defined as

$$P_c(\omega^i \rightarrow \omega^{i+1}) = \begin{cases} 1, & \text{if } f(\omega^{i+1}) \leq f(\omega^i) \\ \exp\left(\dfrac{f(\omega^i) - f(\omega^{i+1})}{\mathbf{c}}\right), & \text{if } f(\omega^{i+1}) > f(\omega^i) \end{cases} \tag{3}$$

where $\mathbf{c}$ denotes a parameter. This parameter is reduced by a factor of 0.95 after several transitions are made, thereby decreasing the probability of accepting inferior moves. Reduction stops when the process stabilizes. A transition takes place if ω^{i+1} is accepted. Next, a solution ω^{i+2} is derived from ω^{i+1}, and the probability $P_c(\omega^{i+1} \rightarrow \omega^{i+2})$ is calculated with a similar acceptance criterion as Equation 3.

For computational feasibility the SA algorithm was implemented with the following parameters. The initial temperature was set to $\mathbf{c} = 10$, and decreased by a factor of 0.95, i.e., $\mathbf{c} \leftarrow 0.95\mathbf{c}$, after every 50 transitions, and the algorithm terminates after 1000 transitions have passed.

Particle Swarm Optimization (PSO) is a stochastic population-based metaheuristic inspired from swarm intelligence. The metaheuristic mimics the social behavior of natural swarms, such as bird flocking and fish schooling. PSO has been applied to a range of optimization problems (Kennedy & Ebenhart, 2001), including spectral unmixing (Omran et al., 2006).

The PSO algorithm developed for the spectral unmixing problem is a basic gBest PSO (Global Best Particle Swarm Optimization) derived from the framework proposed by Talbi (2009). The algorithm starts by randomly generating N particles. For each particle the initial abundance for each member, a_i^{temp}, was randomly taken from $(0, 1)$. Since $\sum_{i=1}^{M} a_i = 1$, the abundance was subsequently normalized by letting $a_i = a_i^{\text{temp}} / \sum_{k=1}^{M} a_k^{\text{temp}}$ for each endmember. Each particle's member abundance was further assigned a velocity v_i which was randomly taken from $[-V_{\max}, V_{\max}]$. Lastly, each particle is assigned a local best solution, p, which represents the best solution found by a particle. The global best solution, that is the best solution found by the entire swarm during the search, is given by g.

In each iteration, t, each particle first updates its velocity to

$$v_i(t) = w \times v_i(t-1) + \tau_1 \times (p_i - a_i(t-1)) + \tau_2 \times (g_i - a_i(t-1)),$$

where τ_1 and τ_2 are two random variables in the range $[0, 1]$, and w is the inertia weight that controls the impact of the previous velocity on the current one. To prevent the PSO system from exploding the velocity v_i is reset to $V_{\max}$ if it exceeds this value. Similarly, v_i is reset to $-V_{\max}$ if it falls below this value.

Next, each particle updates its position to

$$a_i^{\text{temp}} = \max(0, a_i(t) + v_i(t)).$$

The particle is then normalized by letting $a_i(t) = a_i^{\text{temp}}/sum_{k=1}^{M} a_k^{\text{temp}}$. For particles with $\sum_{i=1}^{M} a^{\text{temp}} = 0$ prior to normalization, the particles are first randomly re-initialized and then normalized. Each particle then updates its best local solution p to $a(t)$ if $f(a(t)) < f(p)$, and the best global solution of the swarm g is updated to $a(t)$ if $f(a(t)) < f(g)$ where f represents the objective value function. Each particle then again updates its velocity and position, and the local and global best solutions are updated. This updating process is iteratively applied until $T_{\max}$ iterations have passed, at which point g is returned as the best solution found during the search.

The gBest PSO was implemented with the following parameter values. The swarm size, N, was set to 30 particles; $V_{\max}$ was set to 0.5; and the algorithm was executed for $T_{\max} = 1000$ iterations. The inertia weight, w, was set to 0.9 and reduced per iteration by Δw, calculated as $\Delta w = (0.9 - 0.4)/T_{\max}$; thus, $w = 0.4$ when the algorithm terminated.

Limited Memory Broyden-Fletcher-Goldfard-Sharon Bound (LM-BFGS-B) algorithm is a popular quasi-Newton optimization technique used for solving large nonlinear optimization problems, subject to simple bounds on the variables (Zhu et al., 1997). The technique was developed by Lu et al. (1994) and uses a limited memory variation of the BFGS update to approximate the inverse Hessian matrix, making it ideal for solving large-scale optimization problems or for solving problems whose Hessian matrix is not practical to calculate.

Similar to the implementation of Zhu et al. (1997), LM-BFGS-B was implemented as follows to solve the spectral unmixing problem. At each iteration a limited memory BFGS approximation to the Hessian is updated, which is then used to define a quadratic model of f. A search direction is then computed by, firstly, identifying a set of active variables, which will be held at their bounds, through a gradient projection method. The quadratic model is approximately minimized in terms of the free variables, and a search direction is then defined as the vector leading from the current state to the approximate minimized quadratic model. Lastly, lines search is performed along the search direction.

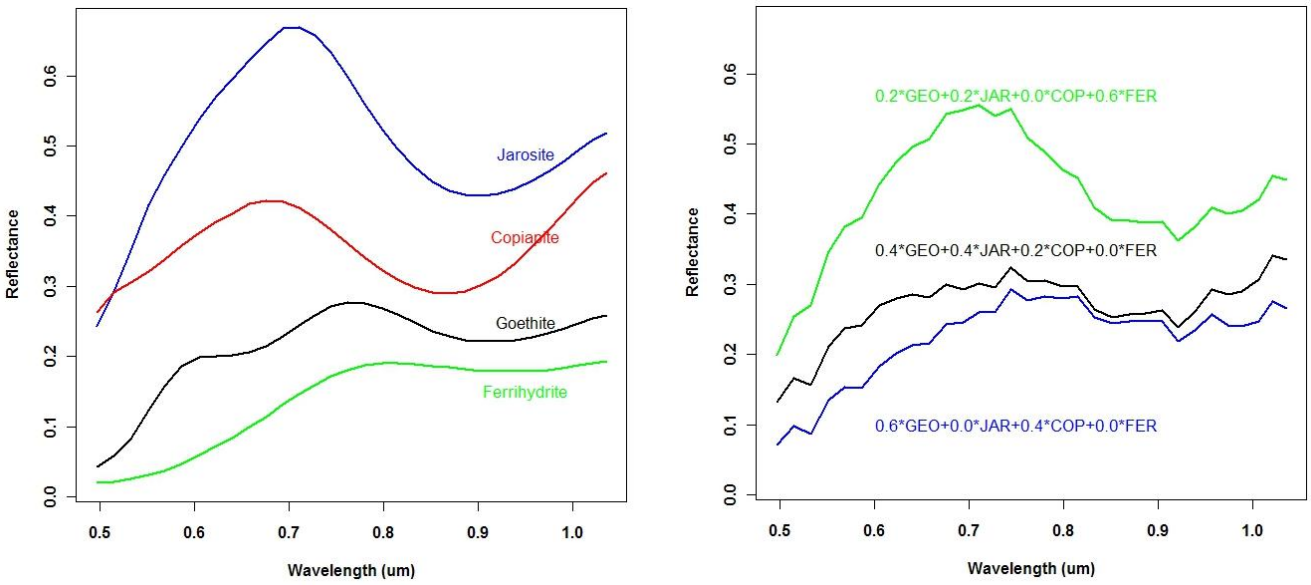
The above mentioned statistical (OLS and RR) and optimization methods (SMGO, SA, PSO and LM-BFGS-B) were compared for estimating the abundances of the endmembers.

3. ENDMEMBER SPECTRA AND SYNTHETIC MIXTURES

Synthetic spectral mixtures were created to test the various methods for spectral unmixing. Four secondary iron-bearing oxide/hydroxide/sulfate minerals that could form pyrite-rich mine wastes were selected to compose a set of endmembers, namely: ferrihydrite; copiapite; jarosite and goethite. Although each secondary iron-bearing oxide/hydroxide/sulfate mineral within a weathering sulfide-bearing mine waste shows distinctive spectral features in the 0.4–2.5 μm regions of the electromagnetic spectrum, this study was limited to the spectral range 0.5–1.1 μm, because this is where most of the iron-bearing oxide/hydroxide/sulfate minerals of interest have many and strong spectral features.

The individual spectrum of each of the four endmembers was selected from the USGS spectral library (Clark et al., 1993) and then linearly mixed with each other according to some proportions of each endmember. Since it has been observed previously that high spectral resolution results in high unmixing accuracy, the mixed spectrum was then degraded (commonly known as resampling) to an approximate 15 nm spectral resolution, to match typical sensors. Error from the uniform and gaussian distribution was then added to the mixed spectra. The resampling was performed (a) to simulate data with lower spectral resolution hyperspectral sensors (e.g., HyMap, DAIS, etc.) as compared to the spectral resolution of the original endmembers in the library, (b) to reduce dimensionality of the data, and (c) because it is a practical technique found effective for prediction of different soil properties (Ben-Dor & Banin, 1994).

Experiments were made on 100 different simple mixtures of selected endmembers, which could plausibly occur in real situations, namely, mixed spectrum $= a_1 \times \text{goethite} + a_2 \times \text{jarosite} + a_3 \times \text{copiapite} + a_4 \times \text{ferrihydrite}$, where $0 \leq a_i \leq 1,\ i = 1, \ldots, 4$ and $\sum_{i=1}^{4} a_i = 1$. For each mixture, error from the $U(-0.002, 0.002)$ or $N(0, 0.00018)$ was added, with corresponding SNR of 500:1, respectively, in the visible to near infrared regions. Graphs of the endmember spectra and the mixed spectra can be seen in Figure 1.



(a) Endmember spectra

(b) Mix spectra

Fig. 1. Endmember spectra and three mixed spectra with error from a uniform distribution.

4. ABUNDANCE ESTIMATION FROM SPECTRA

We simulated a total of 100 mixed spectra using the four endmembers each with both error from the uniform and gaussian distribution. For each of the methods used to unmix the spectra, namely OLS, RR, SMGO, SA, PSO and LM-BFGS-B, we ranked the results according to the absolute difference of the estimated abundance from the actual abundance. The ranks were then summed for the various methods to reflect the combined results of the various methods. Table 1 contains the results of using OLS, RR, SMGO, SA, PSO and LM-BFGS-B in estimating the abundances for each of the endmembers for the highest, the middle and the lowest rank with error from either a uniform or gaussian distribution. Table 1 also contains the summary results using the 100 mixed spectra.

In Table 1, for mixtures 1, 2 and 3, we defined the error for each of the methods as the total absolute deviation of the estimated abundance from the actual abundance, namely,

$$\text{Err} = \sum_{i=1}^{4} |\hat{a_i} - a_i| \ . \tag{4}$$

In Table 1, for the 100 simulations, the average deviations for each $a_1, \ldots, a_4$ is defined as

$$\text{AVG}_{\text{D}_i} = \frac{1}{100} \sum_{j=1}^{100} |\hat{a_{ij}} - a_{ij}|, \qquad i = 1, \ldots, 4 \tag{5}$$

the minimum deviation for each $a_1, \ldots, a_4$ is defined as

$$\text{MIN}_{\text{D}_i} = \min |\hat{a_{ij}} - a_{ij}|, \qquad i = 1, \ldots, 4 \tag{6}$$

and the maximum deviation for each $a_1, \ldots, a_4$ is defined as

$$\text{MAX}_{\text{D}_i} = \max |\hat{a_{ij}} - a_{ij}|, \qquad i = 1, \ldots, 4 \tag{7}$$

Hence, $\overline{\text{AVG}}_{\text{D}}$, $\overline{\text{MIN}}_{\text{D}}$ and $\overline{\text{MAX}}_{\text{D}}$ are the averages of those defined in Equations 5, 6 and 7 respectively.

For mixture 1, using errors from the uniform distribution, OLS and SMGO performed best while RR1 and RR2 performed worst. For mixture 2, using errors from the uniform distribution, RR2 performed best while all other methods performed almost

Table 1. Constraint statistical and optimization methods for estimating abundances with $a_i \geq 0$ and $\sum_{i=1}^{M} a_i = 1$.

	Mixed1					Mixed2					Mixed3				
Problem	a_1	a_2	a_3	a_4	Err	a_1	a_2	a_3	a_4	Err	a_1	a_2	a_3	a_4	Err
Actual abundance	0.6000	0.0000	0.4000	0.0000		0.2000	0.2000	0.0000	0.6000		0.0000	0.8000	0.0000	0.2000	
$\mathbf{w} \sim U(-0.002, 0.002)$															
OLS	0.5991	0.0000	0.4009	0.0000	*0.0018*	0.2571	0.1827	0.0009	0.5594	*0.1160*	0.0571	0.7827	0.0009	0.1594	*0.1160*
RR1[1]	0.6550	0.0000	0.4008	0.0000	*0.0557*	0.2564	0.1828	0.0009	0.5599	*0.1146*	0.0574	0.7825	0.0010	0.1590	*0.1170*
RR2[2]	0.6366	0.0000	0.3958	0.0000	*0.0408*	0.2046	0.1886	0.0096	0.5972	*0.0285*	0.1156	0.7674	0.0053	0.1118	*0.2416*
SMGO	0.5991	0.0000	0.4009	0.0000	*0.0018*	0.2571	0.1827	0.0009	0.5594	*0.1160*	0.0571	0.7827	0.0009	0.1594	*0.1160*
SA	0.6076	0.0000	0.3904	0.0020	*0.0192*	0.2582	0.1824	0.0009	0.5585	*0.1183*	0.0566	0.7829	0.0006	0.1599	*0.1143*
PSO	0.6107	0.0000	0.3893	0.0000	*0.0214*	0.2515	0.1844	0.0000	0.5640	*0.1030*	0.0676	0.7809	0.0000	0.1515	*0.1352*
LM-BFGS-B	0.6109	0.0000	0.3892	0.0000	*0.0217*	0.2589	0.1828	0.0000	0.5583	*0.1179*	0.0571	0.7826	0.0009	0.1593	*0.1161*
$\mathbf{w} \sim N(0, 0.00018)$															
OLS	0.6712	0.0000	0.4045	0.0000	*0.0757*	0.2712	0.1765	0.0045	0.5479	*0.1513*	0.0712	0.7765	0.0045	0.1479	*0.1513*
RR1[1]	0.6690	0.0000	0.4044	0.0000	*0.0734*	0.2704	0.1766	0.0045	0.5485	*0.1499*	0.0715	0.7763	0.0046	0.1476	*0.1522*
RR2[2]	0.6485	0.0000	0.3987	0.0000	*0.0498*	0.2153	0.1822	0.0146	0.5879	*0.0599*	0.1318	0.7604	0.0093	0.0984	*0.2823*
SMGO	0.6712	0.0000	0.4045	0.0000	*0.0757*	0.2712	0.1765	0.0045	0.5479	*0.1513*	0.0712	0.7765	0.0045	0.1479	*0.1513*
SA	0.6036	0.0000	0.3902	0.0061	*0.0195*	0.2702	0.1765	0.0047	0.5486	*0.0514*	0.0736	0.7758	0.0046	0.1459	*0.0541*
PSO	0.6126	0.0000	0.3874	0.0000	*0.0253*	0.2668	0.1763	0.0075	0.5494	*0.1486*	0.0674	0.7799	0.0000	0.1527	*0.1347*
LM-BFGS-B	0.6069	0.0000	0.3890	0.0041	*0.0220*	0.2710	0.1762	0.0049	0.5478	*0.1519*	0.0709	0.7765	0.0045	0.1481	*0.1507*
	AVG_D					MIN_D					MAX_D				
Summary (100 simulations)	a_1	a_2	a_3	a_4	$\overline{\text{AVG}}_\text{D}$	a_1	a_2	a_3	a_4	$\overline{\text{MIN}}_\text{D}$	a_1	a_2	a_3	a_4	$\overline{\text{MAX}}_\text{D}$
$\mathbf{w} \sim U(-0.002, 0.002)$															
OLS	0.0565	0.0144	0.0009	0.0268	*0.0247*	0.0009	0.0000	0.0009	0.0000	*0.0005*	0.0571	0.0173	0.0009	0.0580	*0.0333*
RR1[1]	0.0555	0.0142	0.0009	0.0262	*0.0242*	0.0537	0.0000	0.0008	0.0000	*0.0136*	0.0574	0.0175	0.0011	0.0567	*0.0332*
RR2[2]	0.0312	0.0116	0.0093	0.0133	*0.0164*	0.0002	0.0000	0.0017	0.0000	*0.0005*	0.1156	0.0384	0.0185	0.0882	*0.0652*
SMGO	0.0565	0.0144	0.0009	0.0268	*0.0247*	0.0009	0.0000	0.0009	0.0000	*0.0005*	0.0571	0.0173	0.0009	0.0580	*0.0333*
SA	0.0355	0.0112	0.0019	0.0243	*0.0182*	0.0030	0.0000	0.0000	0.0000	*0.0008*	0.0618	0.0187	0.0107	0.0444	*0.0339*
PSO	0.0354	0.0111	0.0025	0.0237	*0.0182*	0.0035	0.0000	0.0000	0.0000	*0.0009*	0.0723	0.0213	0.0140	0.0542	*0.0404*
LM-BFGS-B	0.0490	0.0142	0.0017	0.0238	*0.0222*	0.0016	0.0000	0.0000	0.0000	*0.0004*	0.0974	0.0264	0.0126	0.0725	*0.0522*
$\mathbf{w} \sim N(0, 0.00018)$															
OLS	0.0712	0.0195	0.0045	0.0345	*0.0324*	0.0712	0.0000	0.0045	0.0000	*0.0189*	0.0712	0.0235	0.0045	0.0757	*0.0437*
RR1[1]	0.0695	0.0193	0.0045	0.0338	*0.0318*	0.0677	0.0000	0.0043	0.0000	*0.0180*	0.0715	0.0237	0.0046	0.0743	*0.0435*
RR2[2]	0.0419	0.0163	0.0120	0.0186	*0.0222*	0.0069	0.0000	0.0001	0.0000	*0.0017*	0.1318	0.0435	0.0226	0.1016	*0.0749*
SMGO	0.0712	0.0195	0.0045	0.0345	*0.0324*	0.0712	0.0000	0.0045	0.0000	*0.0189*	0.0712	0.0235	0.0045	0.0757	*0.0437*
SA	0.0439	0.0154	0.0042	0.0315	*0.0237*	0.0027	0.0000	0.0001	0.0000	*0.0007*	0.0749	0.0245	0.0117	0.0553	*0.0416*
PSO	0.0403	0.0140	0.0031	0.0281	*0.0214*	0.0001	0.0000	0.0000	0.0000	*0.0000*	0.0890	0.0267	0.0129	0.0651	*0.0484*
LM-BFGS-B	0.0605	0.0193	0.0047	0.0305	*0.0287*	0.0028	0.0000	0.0000	0.0000	*0.0007*	0.1151	0.0335	0.0124	0.0839	*0.0612*

[1] Obtained by setting $\delta = 0.005$ and [2] Optimum is first estimated using Hoerl & Kennard (1970) approach

equally poorly. For mixture 3, using errors from the uniform distribution, LM-BFGS-B and SMO performed worst while all other methods performed almost equally well. For mixture 1, using errors from the gaussian distribution, SA performed best while OLS and SMGO performed worst. For mixture 2, using errors from the gaussian distribution, RR2 and SA performed best while all other methods performed almost equally poorly. For mixture 3, using errors from the gaussian distribution, SA performed best while RR2 performed worst. Based on all 100 simulations and using errors from the uniform distribution, RR2 performed best and this was closely followed by SA and PSO. OLS, SMGO and RR1 performed worst in this case. Based on all 100 simulations and using errors from the gaussian distribution, PSO performed best and this was closely followed by RR2 and SA. OLS and SMGO performed worst in this case. Hence it has been generally observed that RR2, SA and PSO outperforms OLS and SMGO.

5. CONCLUSIONS

We conclude that RR, SA and PSO outperforms OLS and SMGO in estimating the partial abundance of each endmember and this was independent of the error from either a uniform or gaussian distribution. The disadvantage of PSO and SA is that the implementations take longer to converge compared to RR, OLS, SMGO and LM-BFGS-B. On the test instances the average execution times of PSO and SA were 18 and 11 seconds, respectively, whereas the average execution times of SMGO and LM-BFGS-B were 0.025 and 0.00016 seconds, respectively, and for RR and OLS convergence was under 10^{-8} seconds. Hence for large remote sensing scenes, typically with millions of pixels and with many endmembers, we recommend using RR2.

REFERENCES

Aarts, E. & Korst, J. (1989). *Simulated Annealing and Boltzmann Machines*. New York: John Wiley.

Adams, J., Sabol, D., & Kapos, V. (1995). Classification of multispectral images based on fraction endmembers, application to land cover change in the Brazilian Amazon. *Remote Sensing of Environment*, 52, 137–154.

Atkinson, P., Cutler, M., & Lewis, H. (1997). Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing*, 18(4), 917–935.

Ben-Dor, E. & Banin, A. (1994). Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils. *Remote Sensing of Environment*, 48(3), 261–274.

Boender, C. G. E., Rinnooy Kan, A. H. G., Timmer, G. T., & Stougie, L. (1982). A stochastic method for global optimization. *Mathematical Programming*, 22(1), 125–140. 10.1007/BF01581033.

Bohachevsky, I. O., Johnson, M. E., & Stein, M. L. (1986). Generalized simulated annealing for function optimization. *Technometrics*, 28(3), 209–217.

Clark, R. N., Swayze, G. A., Gallagher, A. J., King, T. V. V., & Calvin, W. M. (1993). The U. S. Geological survey, digital spectral library: Version 1: 0.2 to 3.0 microns. U.S. Geological Survey Open File Report 93-592.

Csendes, T., Pl, L., Sendn, J., & Banga, J. (2008). The GLOBAL optimization method revisited. *Optimization Letters*, 2, 445–454. 10.1007/s11590-007-0072-3.

Debba, P., Carranza, E. J. M., van der Meer, F. D., & Stein, A. (2006). Abundance estimation of spectrally similar materials by using derivatives in simulated annealing. *IEEE Geoscience and Remote Sensing*, 44(12), 3649–3658.

DeFries, R., Townshend, J., & Hansen, M. (1999). Continuous fields of vegetation characteristics at the global scale at 1-km resolution. *Journal of Geophysical Research*, 104, 16911–16923.

Everitt, B. (1974). *Cluster analysis*. Heinmann, London.

Foody, G. (1996). Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17(7), 1317–1340.

Foody, G., Campbell, N., Trodd, N., & Wood, T. (1992). Derivation and applications of probabilistic measures of class membership from maximum-likelihood classification. *Photogrammery Engineering and Remote Sensing*, 58(9), 13351341.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

Ju, J., Kolaczyk, E. D., & Gopal, S. (2003). Gaussian mixture discriminant anlaysis and sub-pixel land cover characterization in remote sensing. *Remote Sensing of Environment*, 84, 550–560.

Kennedy, J. & Ebenhart, R. (2001). *Swarm Intelligence*. Morgan Kaufmann, San Francisco.

Keshava, N. (2003). A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1), 55–78.

Keshava, N. & Mustard, J. F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1), 44–57.

Kirkpatrick, S., Gelatt, J. C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.

Linforda, N. & Platzman, E. (2004). Estimating the approximate firing temperature of burnt archaeological sediments through an unmixing algorithm applied to hysteresis data. *Physics of the Earth and Planetary Interiors*, 147, 197–207.

Liu, W., Seto, K., Wu, E., Gopal, S., & Woodcock, C. (2004). ART-MMAP: A neural network approach to subpixel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(9), 1976–1983.

Lu, P., Nocedal, J., Zhu, C., Byrd, R. H., & Byrd, R. H. (1994). A limited-memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5), 1190–1208.

Miao, X., Gong, P., Swope, S., Pu, R., Carruthers, R., Anderson, G. L. Heaton, J. S., & Tracy, C. R. (2006). Estimation of yellow starthistle abundance through CASI-2 hyperspactral imagery using linear spectral mixture models. *Remote Sensing of Environment*, 101, 329–341.

Omran, M. G., Engelbrecht, A. P., & Salman, A. (2006). A PSO-based end-member selection method for spectral unmixing of multispectral satellite images. *International Journal of Computational Intelligence*, 2(2), 124–134.

Peddle, D. R. & Smith, M. (2005). Spectral mixture analysis of agricultural crops: endmember validation and biophysical estimation in potato plots. *International Journal of Remote Sensing*, 26, 4959–4979.

Penn, B. S. (2002). Using simulated annealing to obtain optimal linear end-member mixtures of hyperspectral data. *Computers & Geoscience*, 28, 809–817.

Quarmby, N. A., Townshend, J. R. G., Settle, J. J., & White, K. H. (1992). Linear mixture modeling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing*, 13, 415–425.

Schowengerdt, R. (1996). On the estimation of spatial-spectral mixing with classifier likelihood functions. *Pattern Recognition Letter*, 17(13), 1379–1387.

Settle, J. (2006). On the effect of variable endmember spectra in the linear mixture model. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 389–396.

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

Somers, B., Delalieux, S., Verstraeten, W. W., Verbesselt, J., Lhermitte, S., & Coppin, P. (2009). Magnitude- and shape-related feature integration in hyperspectral mixture analysis to monitor weeds in citrus orchards. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11), 3630–3642.

Talbi, E. (2009). *Metaheuristics: From Desing to Implementation*. Willey, New Jersey.

Van der Meer, F. D. & De Jong, S. M. (2000). Improving the results of spectral unmixing of Landsat Thematic Mapper imagery by enhancing the orthogonality of end-members. *International Journal of Remote Sensing*, 21(15), 2781–2797.

Walton, J. T. (2008). Subpixel urban land cover estimation: Comparing cubist, random forests, and support vector regression. *Photogrammetric Engineering & Remote Sensing*, 74(10), 1213–1222.

Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4), 550–560.