AGU PUBLICATIONS

Geochemistry, Geophysics, Geosystems

Supporting Information for

New methods for unmixing sediment grain size data

Greig A. Paterson[1] and David Heslop[2]

[1] Key Laboratory of the Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China. E-mail: greig.paterson@mail.iggcas.ac.cn

[2] Research School of Earth Sciences, Australian National University, Bldg. 142, Mills Rd., Canberra, ACT 2601, Australia.

Introduction

This Supporting Information outlines the details of the hierarchal alternating least squares non-negative matrix factorization (HALS-NMF) algorithm described in the main paper. The convexity errors from the comparison of different algorithms are given in Figure 1. The comparison of end members and abundances from the algorithm comparison are presented in Figure 2 and Figure 3, respectively. Examples of the type of plots that can be generated and save by the AnalySize software package are shown in Figure 4.

The HALS-NMF algorithm

For a data set of *N* specimens with *P* grain size bins resulting from mixtures of *J* end members, the unmixing problem can be expressed in matrix notation as:

$$\mathbf{X} = \mathbf{AS} \qquad (1)$$

where **X** is a *N*×*P* matrix of observed data (one specimen per row), **A** is the *N*×*J* abundance matrix of the constituent end members whose forms are given by the *J*×*P* matrix **S** (one end-member per row). Equation (1) is subject to abundance non-negativity and sum-to-one constraints: $\forall_i : \mathbf{A}_i \geq 0$ and $\sum_i \mathbf{A}_i = 1$ (or 100%). These constraints correspond to the physical requirement that abundances cannot be negative and represent parts that must sum to a whole. In the case of unmixing GSD data, it is also required that end member signatures, which are normalized particle counts, are non-negative and sum-to-one: $\forall_k : \mathbf{S}_k \geq 0$ and $\sum_k \mathbf{S}_k = 1$.

Non-negative matrix factorization (NMF) aims to find two non-negative matrices, $\hat{\mathbf{A}}$ and $\hat{\mathbf{S}}$, such that

$$\mathbf{X} \approx \hat{\mathbf{A}}\hat{\mathbf{S}} \tag{2}$$

The commonly adopted NMF method of Lee and Seung (1999) finds a solution to equation (2) that corresponds to a local minimum or saddle-point in the residual quadratic error (RQE):

$$RQE(\mathbf{A},\mathbf{S}) = \left\|\mathbf{X} - \mathbf{AS}\right\|_F^2 \tag{3}$$

where $\left\|\bullet\right\|_F$ is the Frobenius norm (i.e., the quadratic norm).

Hierarchal alternating least squares (HALS)-NMF (Cichocki et al., 2008) differs by defining the residuals using each row of **S** and each column of **A** (i.e., they are based on the residuals associated with each end member). The residuals are given by:

$$\mathbf{X}^{(k)} = \mathbf{X} - \sum_{i \neq k} \mathbf{A}_i \mathbf{S}_i = \mathbf{X} - \mathbf{AS} + \mathbf{A}_k \mathbf{S}_k \tag{4}$$

for k = 1, 2, …, J, where $\mathbf{S}_k$ represents a single end member (1×P) with corresponding fractional abundances $\mathbf{A}_k$ (N×1). Substituting equation (4) into (3) yields the HALS-NMF RQE function (i.e., the objective function to be minimized):

$$RQE(\mathbf{A}_k,\mathbf{S}_k) = \left\|\mathbf{X}^{(k)} - \mathbf{A}_k\mathbf{S}_k\right\|_F^2. \tag{5}$$

The partial derivatives of the HALS-NMF RQE function are:

$$\begin{aligned}\frac{\partial RQE(\mathbf{A}_k,\mathbf{S}_k)}{\partial \mathbf{S}_k} &= 2\left(\left\|\mathbf{A}_k\right\|^2 \mathbf{S}_k - \mathbf{A}_k^T\mathbf{X}^{(k)}\right)\\ \frac{\partial RQE(\mathbf{A}_k,\mathbf{S}_k)}{\partial \mathbf{A}_k} &= 2\left(\mathbf{A}_k\left\|\mathbf{S}_k\right\|^2 - \mathbf{X}^{(k)}\mathbf{S}_k^T\right).\end{aligned} \tag{6}$$

The update rules for basic HALS-NMF can be obtained by setting the above gradients to zero:

$$\mathbf{S}_k \leftarrow \left[\frac{\mathbf{A}_k^T\mathbf{X}^{(k)}}{\left\|\mathbf{A}_k\right\|^2}\right]_{[0,1]}, \quad \mathbf{A}_k \leftarrow \left[\frac{\mathbf{X}^{(k)}\mathbf{S}_k^T}{\left\|\mathbf{S}_k\right\|^2}\right]_{[0,1]}. \tag{7}$$

The $[\delta]_{[0,1]}$ function constrains each element of δ to lie within the range [0,1]:

$$[\delta]_{[0,1]} = \forall_{ij} \begin{cases} 0, \text{ if } \delta_{ij} < 0 \\ \delta_{ij}, \text{ if } 0 < \delta_{ij} < 1 \\ 1, \text{ if } \delta_{ij} > 1 \end{cases}. \tag{8}$$

Chen and Guillaume (2012) introduced a number of constraints to the basic HALS-NMF routine. They expressed the optimization function for the HALS-NMF routine in a more general form:

$$f\left(\mathbf{A}_k,\mathbf{S}_k\right)=RQE\left(\mathbf{A}_k,\mathbf{S}_k\right)+\sum_i \alpha_i D_i\left(\mathbf{A}_k\right)+\sum_j \beta_j D_j\left(\mathbf{S}_k\right), \tag{9}$$

where $D_i\left(\mathbf{A}_k\right)$ are constraint functions on the abundances, weighted by α_i, and $D_j\left(\mathbf{S}_k\right)$ are constraint functions on the end members, weighted by β_i. Of interest to grain size data unmixing are the abundance sum-to-one constraint and the so-called "minimum distance" constraint.

The abundance sum-to-one constraint ensures that the sum of the abundances of each row of **A** are equal to 1. This can be defined as:

$$D_1\left(\mathbf{A}_k\right)=\left\|\mathbf{A}_k+\sum_{i\neq k}\mathbf{A}_i-\mathbf{1}_{(N\times 1)}\right\|_F^2 \tag{10}$$

where $\mathbf{1}_{(N\times 1)}$ is a $N\times 1$ column vector of ones. The derivative with respect to $\mathbf{A}_\text{k}$ is

$$\frac{\partial D_1\left(\mathbf{A}_k\right)}{\partial \mathbf{A}_k}=2\left(\mathbf{A}_k+\sum_{i\neq k}\mathbf{A}_i-\mathbf{1}_{(N\times 1)}\right). \tag{11}$$

The minimum distance constraint encourages HALS-NMF to find a solution that minimizes the distance of each end member to the centroid of the simplex, which encourages the simplex to shrink and bound the observations closely. This can be expressed as

$$D_2\left(\mathbf{S}_k\right)=\left\|\left(\mathbf{I}_{(P\times P)}-\frac{1}{P}\mathbf{1}_{(P\times P)}\right)\left(\mathbf{S}_k-\frac{1}{J}\left(\mathbf{S}_k+\sum_{i\neq k}\mathbf{S}_i\right)\right)\right\|_F^2, \tag{12}$$

where $\mathbf{I}_{(P\times P)}$ is a $P\times P$ identity matrix. The derivative with respect to $\mathbf{S}_\text{k}$ is

$$\frac{\partial D_2\left(\mathbf{S}_k\right)}{\partial \mathbf{S}_k}=2\left(\mathbf{I}_{(P\times P)}-\frac{1}{P}\mathbf{1}_{(P\times P)}\right)\left(1-\frac{1}{J}\right)\left(\mathbf{S}_k-\frac{1}{J}\left(\mathbf{S}_k+\sum_{i\neq k}\mathbf{S}_i\right)\right). \tag{13}$$

By combining equations (6), (9), (11), and (13) the update rules for constrained HALS-NMF are given by

$$\begin{aligned}
\mathbf{S}_k &\leftarrow \left[\frac{\mathbf{A}_k^T\mathbf{X}^{(k)}+\beta_2\dfrac{1}{J}\left(1-\dfrac{1}{J}\right)\left(\mathbf{I}_{(P\times P)}-\dfrac{1}{P}\mathbf{1}_{(P\times P)}\right)\sum_{i\neq k}\mathbf{A}_i}{\left\|\mathbf{A}_k\right\|^2 I_{(P\times P)}+\beta_2\left(\mathbf{I}_{(P\times P)}-\dfrac{1}{P}\mathbf{1}_{(P\times P)}\right)\left(1-\dfrac{1}{J}\right)^2}\right]_{[0,1]}, \\
\mathbf{A}_k &\leftarrow \left[\frac{\mathbf{X}^{(k)}\mathbf{S}_k^T+\alpha_1\left(\mathbf{1}_{(N\times 1)}-\sum_{i\neq k}\mathbf{A}_i\right)}{\left\|\mathbf{S}_k\right\|^2+\alpha_1}\right]_{[0,1]}.
\end{aligned} \tag{14}$$

It should be noted that the formulation of the HALS-NMF does not allow for an explicit constraint that the end member signatures must sum-to-one. Although, given the other constraints on the problem and the fact that the data fit these constraints, the sum of each end members tends to be ≈

1. Nevertheless, after each iterative update of the non-negative matrices, the end member signatures are normalized to sum-to-one.

The constrained HALS-NMF algorithm for grain size unmixing is thus as follows.

```
1. Initialize S using the SISAL (Bioucas-Dias, 2009) and setting S < 0 to 0. Using X and the initial
   S, initialize A using fully constrained least squares (Heinz and Chang, 2001).
2. while stopping criteria are not met…
        for $k = 1, 2, \ldots, J$
            Update $\mathbf{S}_k$ with equation (14). Normalize $\mathbf{S}_k$ to sum-to-one.
            Update $\mathbf{A}_k$ with equation (14).
        end
   end
```

The algorithm is stopped if one of three conditions is met.

1. The maximum number of iterations (5000) is reached.
2. The maximum relative change of both **A** and **S** between successive iterations is small (≤ 1e-6%).
3. The relative change in misfit (RQE) between successive iterations is small (≤ 1e-6%).

Convexity errors from the algorithm comparison

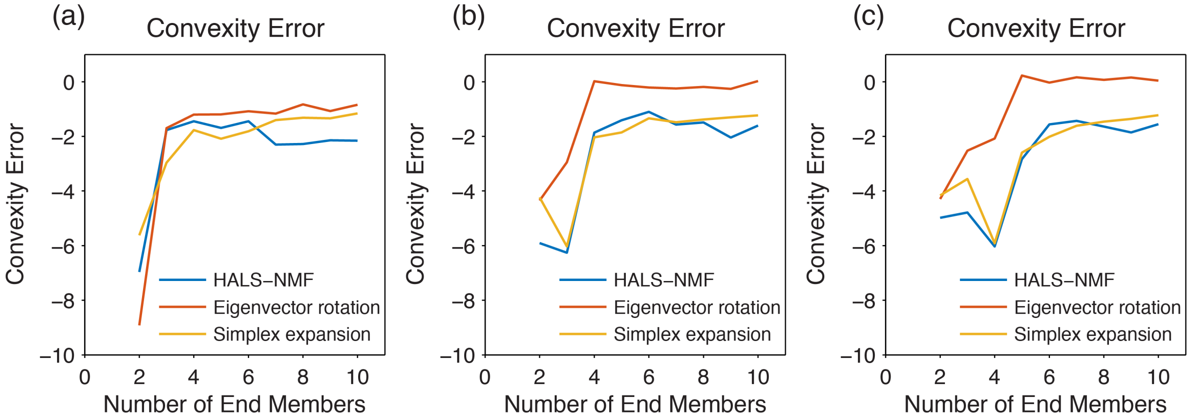

Figure 1. Convexity error for unmixing of data sets of 200 specimens generated by synthetic lognormal end member GSDs with (A) 2, (B) 3, and (C) 4 end members. HALS-NMF refers to our new algorithm following Chen and Guillaume (2012), eigenvector rotation is the algorithm of Dietze et al. (2012), and simplex expansion is from Weltje (1997).

End members from the algorithm comparison

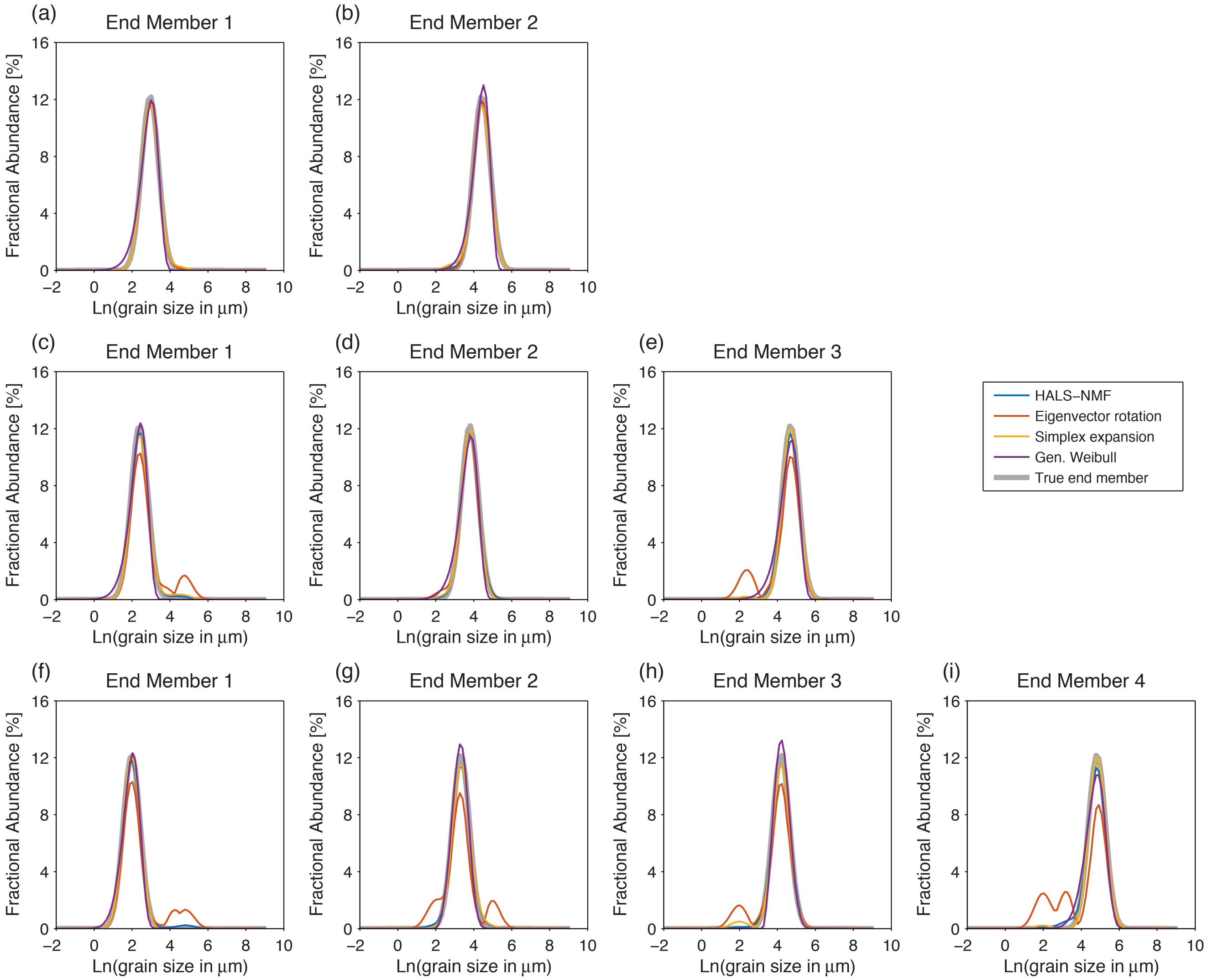

Figure 2. Comparison between the fitted and true end members for the various EMA methods tested. (a–b) The 2 end member data set, (c–e) the 3 end member data, set and (f–i) the 4 end member data set.

Abundances from the algorithm comparison

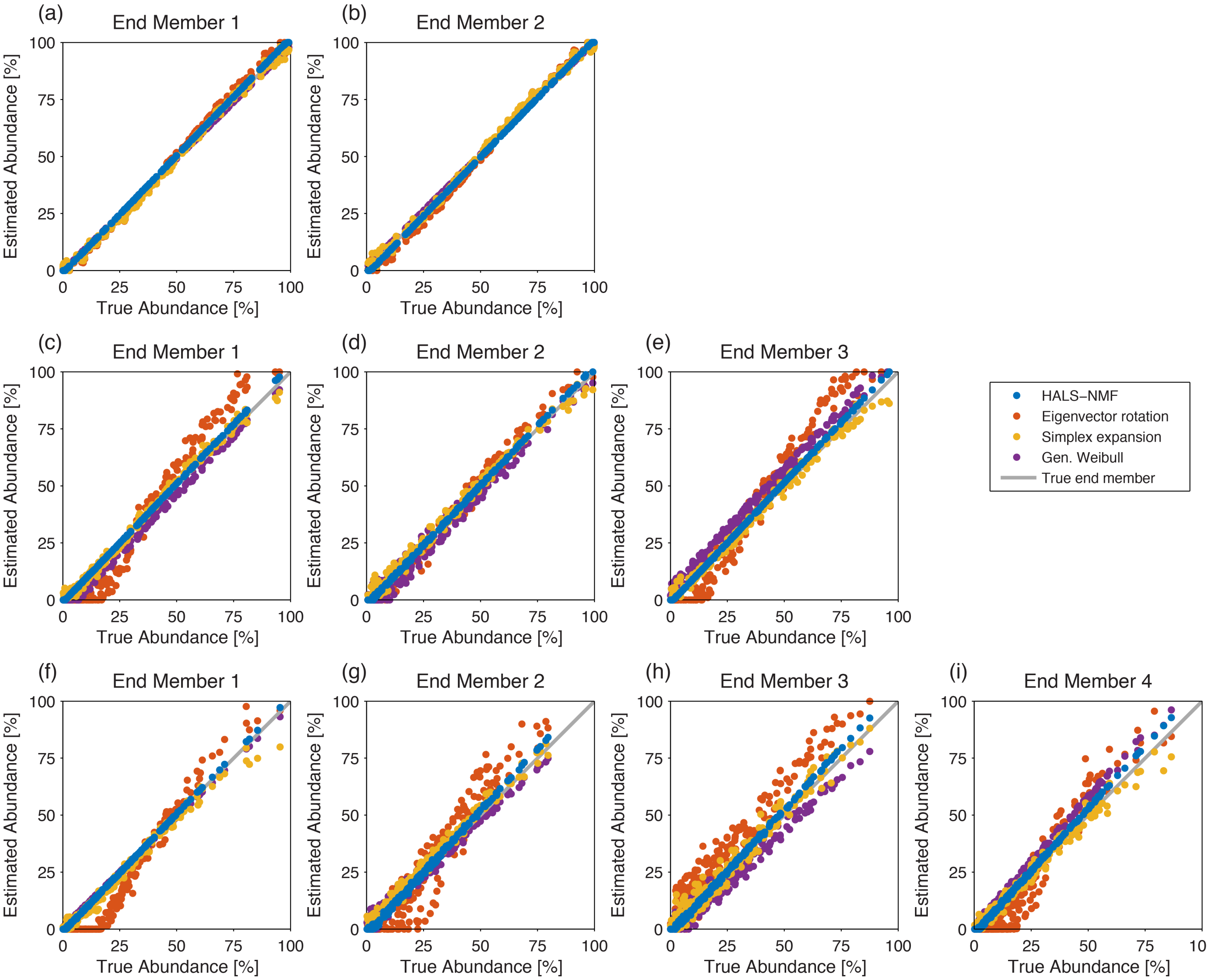

Figure 3. Comparison between the fitted and true abundances for the various EMA methods tested. (a–b) The 2 end member data set, (c–e) the 3 end member data, set and (f–i) the 4 end member data set.

Examples of plots generated by AnalySize

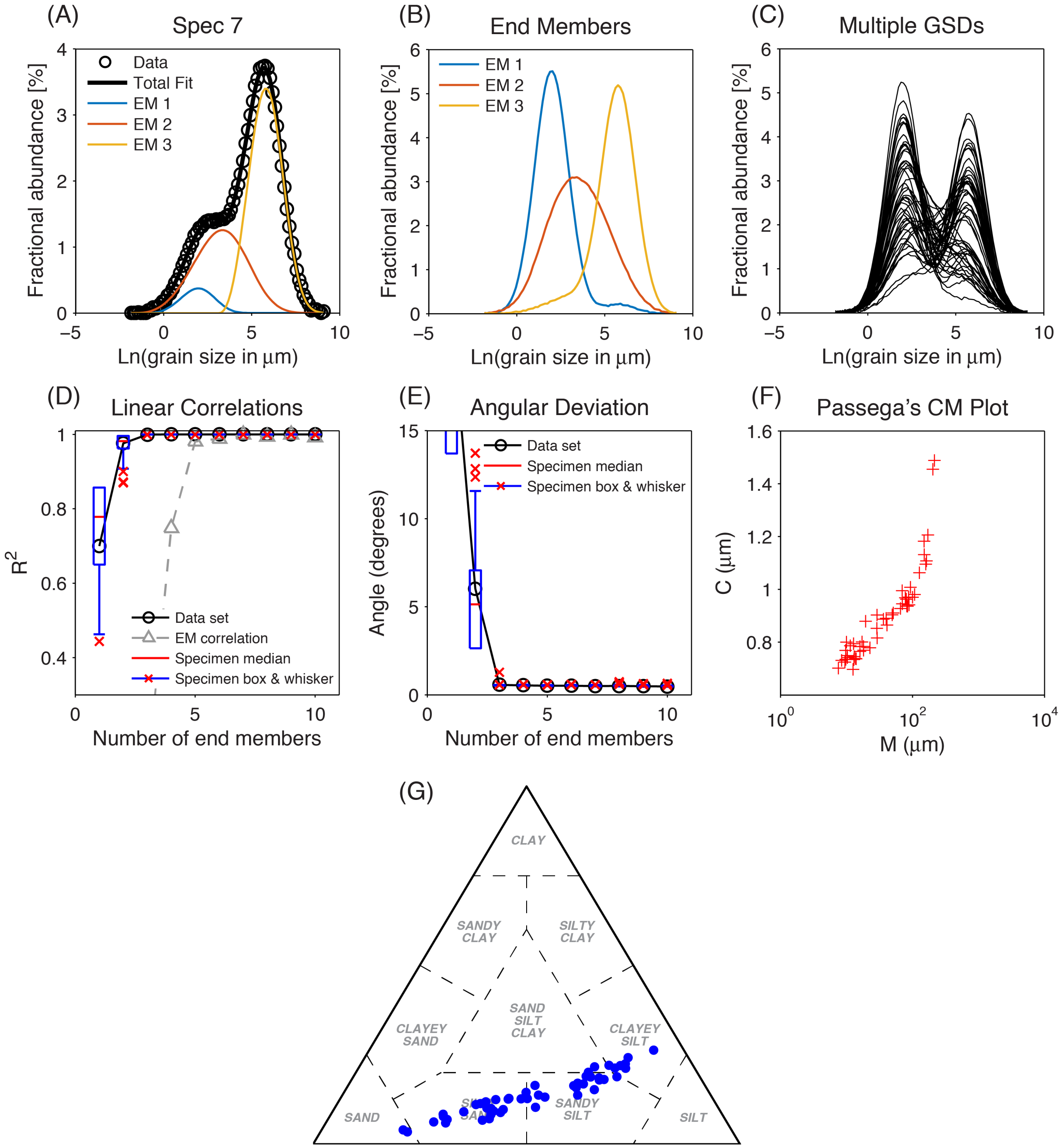

Figure 4. Examples of figures generated by AnalySize. Data are from a synthetics data set of 50 observations with 3 lognormally distributed end members with random abundances that sum to unity (these data are not used in the main paper). (A) A data plot with fitted end members. (B) An end member spectra plot. (C) A multispecimen spectra plot. (D) R^2 and (E) angular deviation goodness-of-fit statistics for various numbers of end members. (F) A CM plot. (G) A Shepard fine ternary diagram. Note that the position of some legends has been adjusted. In parts (D) and (E), the solid black lines and circles are the model misfit to the whole data set, the blue box and whiskers represent the values determined from individual specimens. The red bars are the median values, the blue boxes are the interquartile range, and the blue whiskers mark out the one-sided 95th percentiles (i.e., the 95% coverage interval). For R^2 this is the lower 95th percentile and for angular deviation it is the upper 95th percentile - these represent a measure of the lowest quality fits. The red crosses represent outlying specimens (i.e., specimens that lie outside of the 95% coverage interval). The grey dashed

line and triangles in (D) are the maximum squared linear correlation between the different fitted end members. This is a measure of the linear independence of the end members.

References

Bioucas-Dias, J.M., 2009. A variable splitting augmented Lagrangian approach to linear spectral unmixing. In: First IEEE GRSS workshop on Hyperspectral Image and Signal Processing: WHISPERS 2009, 1-4, doi: 10.1109/WHISPERS.2009.5289072.

Chen, W., Guillaume, M., 2012. HALS-based NMF with flexible constraints for hyperspectral unmixing. EURASIP J. Adv. Sig. Pr., 2012, 54, doi: 10.1186/1687-6180-2012-54.

Cichocki, A., Phan, A.H., Caiafa, C., 2008. Flexible HALS algorithms for sparse non-negative matrix/tensor factorization. In IEEE Workshop on Machine Learning for Signal Processing, 2008. MLSP 2008, 73-78, doi: 10.1109/MLSP.2008.4685458.

Dietze, E., Hartmann, K., Diekmann, B., Ijmker, J., Lehmkuhl, F., Opitz, S., Stauch, G., Wünnemann, B., Borchers, A., 2012. An end-member algorithm for deciphering modern detrital processes from lake sediments of Lake Donggi Cona, NE Tibetan Plateau, China. Sed. Geol., 243–244, 169-180, doi: 10.1016/j.sedgeo.2011.09.014.

Heinz, D.C., Chang, C.-I., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Trans. Geosci. Remote Sensing, 39, 529-545, doi: 10.1109/36.911111.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature, 401, 788-791, doi: 10.1038/44565.

Weltje, G.J., 1997. End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem. Math. Geol., 29, 503-549, doi: 10.1007/BF02775085.