On The Use of a Linear Spectral Unmixing Technique For Concentration Estimation of APXS Spectrum

B. Ayhan and C. Kwan
Signal Processing, Inc.
Rockville, Maryland, USA
chiman.kwan@signalpro.net

S. Vance
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, USA

Abstract— **Past estimates of chemical element concentration from investigations on the surface of Mars by Alpha-Particle X-ray Spectrometer (APXS) instrument have been conducted by converting the peak areas of the characteristic element lines into element concentrations using look-up calibration tables. In this work, we have investigated the feasibility of applying a linear spectral unmixing technique, NCLS (Nonnegatively Constrained Least Squares) to APXS spectra data for concentration estimation. A procedure for signature calibration with the NCLS technique is also introduced in this work. Estimates using NCLS are highly accurate in comparison to the applied benchmark technique, PLS (Partial Least Squares) in a leave-one-out testing framework that uses 11 geostandards. Future work will consider the performance comparison with a peak-area based concentration estimation method, and whether the fusion of the two methods can further increase the correct concentration estimation accuracy.**

I. INTRODUCTION

The Alpha-Particle X-ray Spectrometer (APXS) instrument on NASA's Mars rovers (Spirit, Opportunity and the newly landed Curiosity) can measure the chemical composition of rocks [1]. Chemical element concentration estimates using APXS spectra have mainly been achieved by converting the peak areas of the characteristic element lines into element concentrations using look-up calibration tables [2]. By contrast to peak-area based concentration estimation, in this work we have investigated a linear spectral unmixing technique, NCLS (Nonnegatively Constrained Least Squares) [3], for concentration estimation. A well-known concentration estimation technique in chemometrics, PLS (Partial Least Squares) [4], is also used as a benchmark technique in a leave-one-out testing framework [5] to compare the performance of the two techniques. APXS spectra of 11 geostandards have been used to evaluate the two concentration estimation techniques. Both APXS spectra and element concentrations in weight are available for 11 geostandards, and these 11 geostandards are used for calibration in the Mars Exploration Rovers (MER) [6]. A Leave-One-Out (LOO) testing framework is applied to assess the feasibility of NCLS for concentration estimation in comparison to PLS.

For concentration estimation, PLS requires a training spectral data set for which the element concentrations are known. NCLS also requires a chemical element signature library based on targets of known concentration. The Mars Science Laboratory (Curiosity) has a calibration target on board for APXS measurements, a polished slab of the basalt of Broken Tank (BT2) from near Socorro, NM, which has been well characterized for its mineral and bulk chemistry [7]. In this work, the uncalibrated signature library for NCLS has been formed with the use of the APXS calibration spectra, which include chemical compounds (high purity metal oxides) and 11 geostandards (complex rock samples) [6].

The contributions of this work are: 1) Demonstration of the use of a linear spectral unmixing technique, NCLS, for concentration estimation by assessing the relative performances of NCLS versus PLS in a leave-one-out testing framework, 2) Introduction of a signature calibration procedure to be used with NCLS, using calibration target measurements, 3) Demonstration of the use of NCLS outperforming PLS when the number of training samples is small.

Section II of the paper provides information about the 11 geostandards and their APXS spectra. Section III introduces the concentration estimation techniques and the calibration procedure for NCLS. Section IV provides brief information about the test methodology. Section V summarizes the results and analyses. Section VI provides the concluding comments and future work.

II. ABOUT APXS SPECTRA DATA

The APXS spectra of the 11 geostandards and their ground truth concentrations have been used in

this work. Combined with this dataset, APXS oxide spectra data have been used to extract the initial uncalibrated signature library for NCLS. These datasets have been used to calibrate APXS data from the Mars Exploration Rovers [6]. The concentrations of the 11 geostandards are from [6] and they are in the form of elemental concentrations in weight.

III. CONCENTRATION ESTIMATION TECHNIQUES

A. NCLS (Nonnegatively constrained least squares)

NCLS is based on linear spectral unmixing and requires the use of a spectral signature library of the chemical elements. Suppose the APXS spectra of which its concentration estimate is aimed to be found is denoted by $\mathbf{r}$, and the element signature library for the chemical elements is denoted by $\mathbf{M}$. NCLS assumes a linear mixture model and estimates $\boldsymbol{\alpha}$ from $\mathbf{r}$, $\mathbf{r} = \mathbf{M}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the concentration estimates vector and $\mathbf{r}$ is the APXS spectra of the test sample. The NCLS method guarantees that the estimated abundance fractions in $\boldsymbol{\alpha}$ are nonnegative, because there is a constraint about having nonnegative fraction values within the method. For detailed technical information about NCLS, one could refer to [3], [8].

The uncalibrated APXS signature library for the chemical elements was extracted from the APXS spectra of the oxide compounds obtained in laboratory. These oxide compounds are: Al_2O_3, $Ca_3(PO_4)_2$, $CaCO_3$, Fe_2O_3, FeO, K_2CO_3, KCl, MgO, $NaCl$, SiO_2, TiO_2 [6]. Because the oxide APXS spectra data set does not cover the majority of the chemical elements that are of interest, another APXS data set consisting of 11 geostandards (complex rock samples) is used for extracting uncalibrated element signatures. These laboratory APXS spectra for 11 geostandards are designated as: 1) Allende, 2) AN-G, 3) BE-N, 4) GXR-1, 5) I555, 6) JSd-2, 7) Mica-Fe, 8) Mica-Mg, 9) Milbillie, 10) Murchison, 11) SSK1.1 [6]. The uncalibrated APXS signature library can be seen in Fig. 1. The amplitudes of the element peaks are not important in the uncalibrated signature library since these signatures will be calibrated before they are used for concentration estimation.
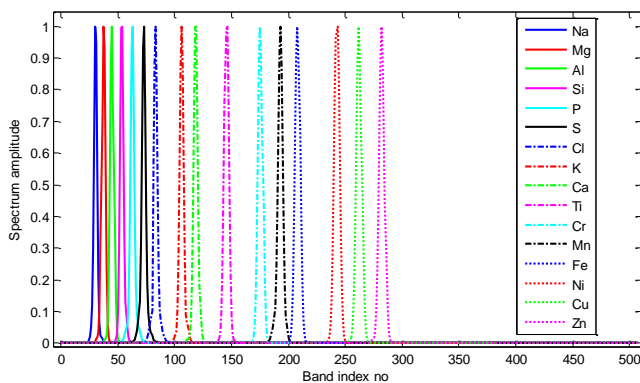

Fig. 1. Uncalibrated APXS element signature library.

The extracted chemical element signatures (uncalibrated) need to be calibrated with the use of a calibration target's APXS spectrum, of which its element concentration is known, before being used for concentration estimation of any given test APXS spectrum with NCLS. The block diagram in Fig. 3 summarizes chemical element signature calibration steps. For signature library calibration, there is a need for a calibration target (of which its element concentration is known in advance) and its APXS spectrum. In the calibration process, NCLS is applied to the calibration target spectrum using the uncalibrated signature library. This process results in concentration estimations for the calibration target; these estimations together with the ground truth concentration estimations of the calibration target are then used to determine the calibration coefficients for each chemical element in the signature library. The calibration coefficients are applied to the uncalibrated signature library resulting in the calibrated signature library. A signature library calibrated using the JSd-2 geostandard as the calibration target can be seen in Fig. 2.
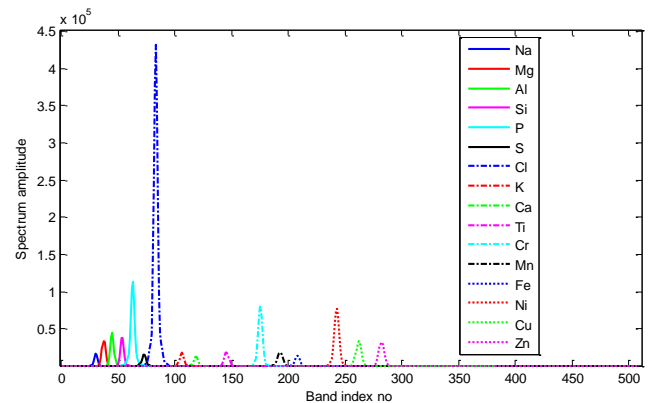

Fig. 2. Calibrated APXS element signature library (Jsd-2 is used as the calibration target)

The sum of the spectrum amplitudes of the calibration target spectrum is called as the calibration target spectrum normalization coefficient, and it is used to normalize any given test APXS spectrum before applying NCLS with the calibrated signature library for concentration estimation. Fig. 4 shows the block diagram for concentration estimation of a given test APXS spectrum using the calibrated signature library and the normalization coefficient.

B. PLS (Partial Least Squares)

PLS is used as a benchmark for comparing NCLS's performance to a well-known technique. Suppose the training set APXS spectra are denoted by $\mathbf{X}$ and the chemical element compositions of $\mathbf{X}$ are denoted by $\mathbf{Y}$. The PLS model is then based on predicting $\mathbf{Y}$ from $\mathbf{X}$, with $\mathbf{Y} = \mathbf{XB}$, via finding a parameter matrix $\mathbf{B}$ which will form the PLS model. With the computed PLS model, one can then estimate the chemical element concentrations of any given APXS test spectra. Suppose $\mathbf{r}$ is the APXS spectra of the test sample, then the concentration estimation of $\mathbf{r}$, $\mathbf{y}$, can be

estimated via, **y** = **rB**. For detailed information about PLS, one can refer to [4], [9], [10], [11].

IV. TEST METHODOLOGY

For the feasibility analysis of applying NCLS for concentration estimation, a leave-one out testing [5] methodology has been used. In this approach, each geostandard is used systematically for testing exactly once, whereas the remaining 10 geostandards are used for training models (in PLS) and for calibrating signature libraries (in NCLS). In PLS, six different numbers of latent variables have been used; these are: 4, 5, 6, 7, 8, 9. In NCLS, each of the training geostandard is considered as a calibration target and the signature library is calibrated accordingly. The calibrated signature library is then applied to the test geostandard spectrum after normalization. This has been repeated for all the geostandards in the training set and the average of the concentration estimates for the test geostandard spectrum is set as the final concentration estimate for the test spectrum. At the end of the leave-one-out testing, concentration estimates are obtained for each of the 11 geostandards with both NCLS and PLS (six set of PLS results each corresponding to a different number of latent variables). The next section presents and discusses the leave-one-out testing results.
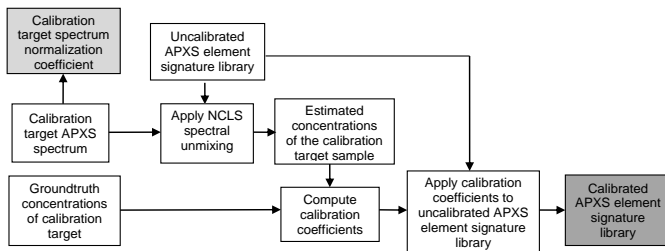

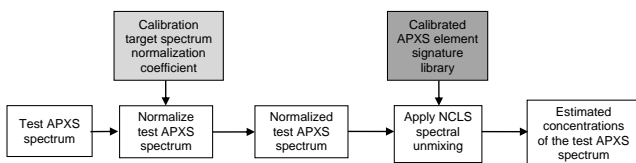Fig. 3. Block diagram for chemical element signature library calibration in NCLS.

Fig. 4. Block diagram for chemical element concentration estimation with NCLS.

V. RESULTS AND ANALYSES

In PLS, some of the chemical element concentration estimations provided negative values, whereas in NCLS the concentration estimates were always positive due to the nonnegativity constraint in NCLS. Since the negative concentration estimations of PLS are not physically meaningful, these values are set to zero and performance measures are computed accordingly. Root mean square error (RMSE) is used as a performance measure for concentration estimates as follows: Suppose $\mathbf{s}_i = \left(s_{i1}, s_{i2}, ..., s_{iL} \right)^T$ is the ground truth concentration vector for the ith

geostandard and $\mathbf{r}_i = \left(r_{i1}, r_{i2}, ..., r_{iL} \right)^T$ is the estimated concentration vector for that same geostandard, where i = 1,…, I, and I is the total number of spectra and j = 1,…, L, where L is the total number of chemical elements. The RMSE measure for the ith spectrum is computed as:

$$RMSE(\mathbf{s}_i, \mathbf{r}_i) = \sqrt{\frac{1}{L} \sum_{j=1}^{L} \left(s_{ij} - r_{ij} \right)^2} \qquad (1)$$

The RMSE errors have been computed from the estimated concentrations of PLS and NCLS and the ground truth concentrations. Table 1 shows the RMSE values of the 11 geostandards after leave-one-out testing. The shaded cells in Table 1 correspond to the best performance among the PLS versions (six different numbers of latent variables) and NCLS (lowest RMSE values). The last row of Table 1 is the sum of RMSE values for the 11 geostandards. NCLS performs significantly better than PLS in 10 of the 11 geostandards. PLS using 5 latent variables provides the lowest RMSE scores.

A second measure, normalized root mean square of prediction (NRMESP) [12] has been used with the resultant concentration estimates of the two techniques from the leave-one-out testing. This measure provides information about the accuracy of the concentration estimates with respect to each chemical element. Technical details about this measure can be found in [12]. Table 2 shows the NRMESP measures for NCLS and PLS (six different number of latent variables). Lower NRMESP measures indicate better concentration estimation accuracy for that chemical element.

Shaded cells in Table 2 correspond to the minimum NRMESP value of the two techniques. The last row of Table 2 shows the sum of the NRMESP values for each technique. NCLS outperforms PLS in 13 of the 16 chemical elements. The chemical elements for which PLS performs better are for the number of latent variables 8 and 9. In fact, these do not provide good performance according to the earlier RMSE measure. The resultant concentration estimates after leave-one-out testing for the 11 geostandards with the two techniques (together with the ground truth concentrations) can be found in Appendix. In Appendix, the concentrations for PLS belong to the case where the number of latent variables is set to 5.

Although the results indicate that NCLS performs significantly better than PLS, NCLS does not outperform PLS in general. The number of samples used in PLS model training in this work is very small in size (only 11 geostandards). A better estimation would most likely have been achieved with PLS if the number of training samples were large. An important strength of NCLS is in yielding significantly more accurate estimates than PLS for smaller training sets, without time-consuming configuration of parameters such as the number of latent variables.

Table 1. RMSE values with the Leave-One-Out testing framework.

	NCLS	PLS(#lv=4)	PLS(#lv=5)	PLS(#lv=6)	PLS(#lv=7)	PLS(#lv=8)	PLS(#lv=9)
Allende	1.308957	2.181138	1.908358	1.554327	1.952197	2.061299	3.731111
AN-G	1.112424	3.437689	2.977046	6.289246	5.984157	5.486227	6.308873
BE-N	0.835839	2.15699	1.84948	1.705092	2.041633	6.137549	8.552493
GXR-1	0.687447	10.80974	10.81389	11.18087	11.22621	11.71318	11.7294
I555	0.539791	3.969472	4.076977	4.466374	5.530412	8.779433	10.56828
JSd-2	0.59217	2.433258	2.190328	2.040929	3.477035	5.547032	5.471425
Mica-Fe	0.801412	3.11218	3.929877	4.562445	4.903586	4.598544	4.856168
Mica-Mg	0.948657	3.290848	3.444919	3.347243	3.878237	3.466385	5.665661
Milbillie	0.86112	1.483939	1.327666	1.568863	2.061543	2.153621	3.104602
Murchison	1.925253	1.871905	1.798293	1.782468	1.435887	2.040317	3.866523
SSK1.1	0.5724	0.767512	0.71388	0.947288	2.143147	1.333994	2.337756
Sum of RMSE	10.18547	35.51668	35.03071	39.44515	44.63405	53.31758	66.19229

Table 2. NRMSE values with the Leave-One-Out testing framework.

	NCLS	PLS(#lv=4)	PLS(#lv=5)	PLS(#lv=6)	PLS(#lv=7)	PLS(#lv=8)	PLS(#lv=9)
Na	0.289163	0.352479	0.2965848	0.3333296	0.3300627	1.031569	0.758729
Mg	0.099857	0.208306	0.1011529	0.08907	0.0778644	0.046152	0.036522
Al	0.063468	0.146596	0.1924815	0.241233	0.283751	0.441062	0.456126
Si	0.069266	0.313846	0.3014136	0.3658966	0.4050327	0.208219	0.222045
P	0.195836	0.337174	0.4496975	0.5047816	0.4905098	0.558173	0.202216
S	0.172679	0.753034	0.7771212	0.6250066	0.5721943	0.339972	0.253152
Cl	0.239664	0.481226	0.6092803	0.6112757	0.5134411	0.696498	0.576685
K	0.16961	0.542041	0.26461	0.1937247	0.1912699	0.17711	0.299008
Ca	0.136188	0.437116	0.1766919	0.189508	0.201446	0.262324	0.272383
Ti	0.255664	0.510243	0.6126394	0.4047759	0.3784524	0.338532	0.174261
Cr	0.304764	0.875252	0.8771324	0.8757776	0.9273263	1.238639	1.525121
Mn	0.195925	0.460519	0.5923527	0.7146581	0.8040564	1.344129	1.617702
Fe	0.111449	0.596628	0.6618463	0.741213	0.7830154	0.939533	1.115819
Ni	0.306763	0.734285	0.7495455	0.5664118	0.5680345	0.475341	0.36461
Cu	0.338691	0.243879	0.2549451	0.2490492	0.2523256	0.224712	0.283975
Zn	0.149246	0.558595	0.5470523	0.5882194	0.4551987	0.473706	0.5104
Sum of NRMSE	3.098233	7.551221	7.4645474	7.293931	7.2339812	8.795669	8.668753

VI. CONCLUSIONS

This work investigated the application of NCLS as a tool for concentration estimation from APXS spectra. Leave-one-out testing results demonstrate simple concentration estimation with NCLS when compared to the conventional technique, PLS. The small number of training samples decreases the accuracy of PLS, relative to NCLS. These comparisons have been conducted using quantitative measures RMSE and NRMESP. NCLS requires a chemical element signature library and calibration of the signature library by a calibration target. The signature calibration procedure for NCLS described in this work could be useful for APXS concentration estimates for current Mars rovers, which are equipped with on-board calibration targets.

Future work will consider the concentration performance comparison of NCLS with respect to a peak-area based concentration estimation method and whether the fusion of the two methods (NCLS and peak-area based method) can further increase the correct concentration estimation accuracy in comparison to using only of the two.

REFERENCES

[1] R. Gellert et al. "Notes on Alpha-Particle X-Ray Spectrometer (APXS) Data Reduction," http://pds-geosciences.wustl.edu/mer/mer2-m-apxs-2-edr-ops-v1/mer2ap_0xxx/calib/apxs_data_reduction.pdf

[2] R. Gellert et al, "Alpha Particle X-Ray Spectrometer (APXS): Results from Gusev crater and calibration report," Journal of Geophysical Research, Vol. 111, 2006.

[3] C. Kwan, B. Ayhan, G. Chen, C. Chang, J. Wang, and B. Ji, "A Novel Approach for Spectral Unmixing, Classification, and Concentration Estimation of Chemical and Biological Agents," IEEE Trans. Geoscience and Remote Sensing, vol. 44, no. 2, Feb. 2006.

[4] S. Wold, et al, "PLS regression: a basic tool of chemometrics," Chemometrics and Intelligent Laboratory Systems, Vol.58, 2001

[5] R. O. Duda, P.E. Hart and D. G. Stork, Pattern Classification, Second Edition, Wiley Interscience, New York, 2000.

[6] Mars Exploration Rover (MER) Project, APXS Calibration Plan, JPL Document Review CI#04-2301

[7] J. L. Campbell, et al., First measurements of the MSL APXS calibration target on Mars, 44th Lunar and Planetary Science Conference (2013).

[8] Hyperspectral Imaging, Techniques for Spectral Detection and Classification, Kluwer Academic, Chein-I Chang.

[9] M. Barker and W. Rayens, Partial least squares for discrimination. *J. of Chemometrics*. 2003; 17: 166-173.

[10] L. Hoegaerts, J. Suykens, J. Vandewalle, B. De Moor, Kernel PLS variants for regresssion. *ESANN'2003 proceedings-European Symposium on Artificial Neural Networks*. pp. 203-208, April 23-25, 2003.

[11] Haenlein, Michael and Andreas M. Kaplan (2004). A Beginner's Guide to Partial Least Squares Analysis, Understanding Statistics, 3(4), 283–297.

[12] N. Yang, et al. "Multivariate analysis of laser induced breakdown spectroscopy of soil samples," Soil Science, Volume 175, Number 9, September 2010.

APPENDIX

The following are the plots showing the concentration estimations from NCLS and PLS (number of latent variables = 5) for the 11 geostandards together with the groundtruth concentrations in the leave-one-out testing framework. Because a logarithmic scale is used in the plots, the value 0 is not shown, so these missing values in the plots can be considered as having a value of 0. It can be seen that NCLS has better matches than PLS.
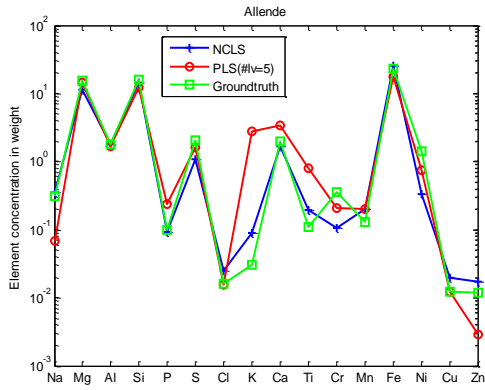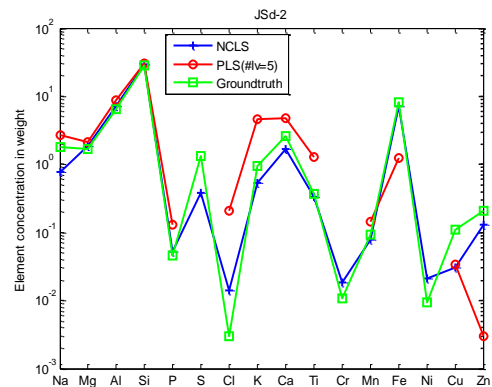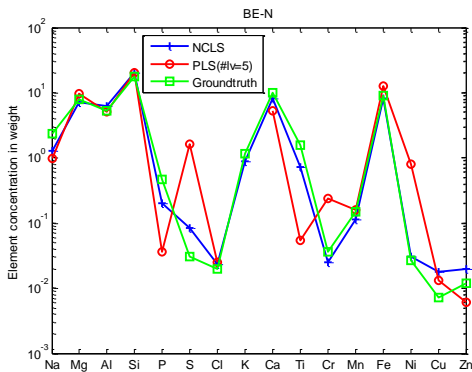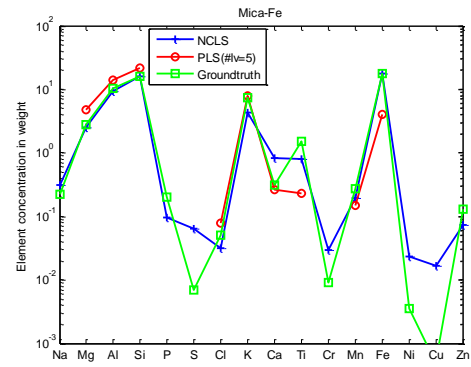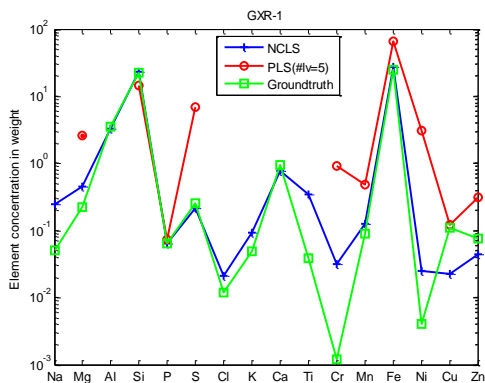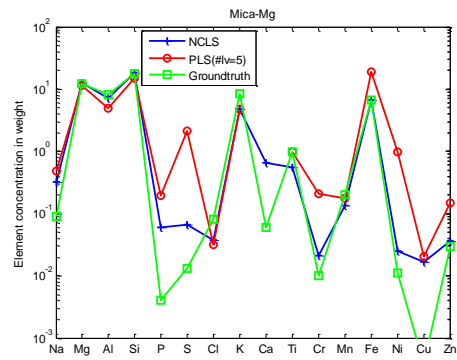
(a) Allende
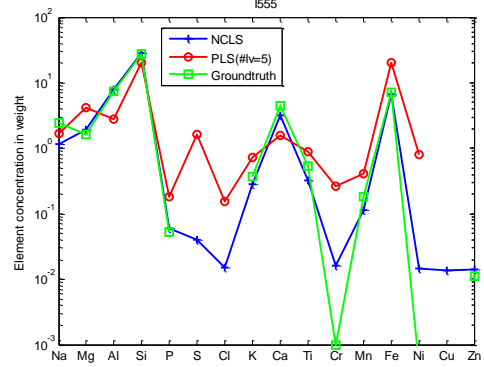
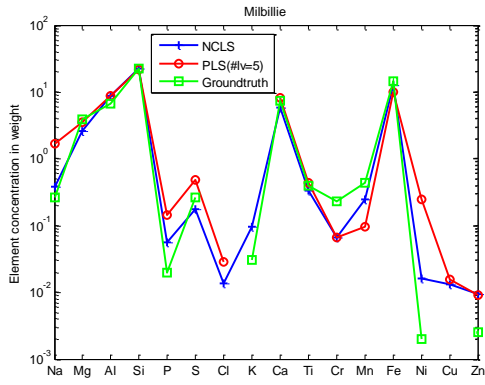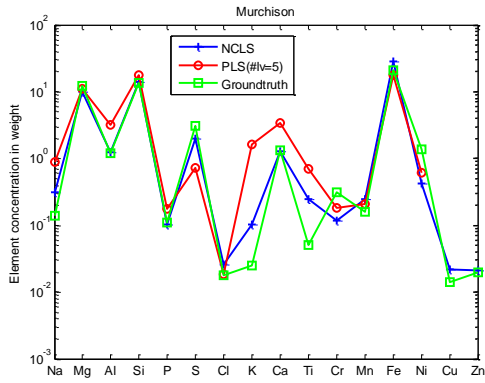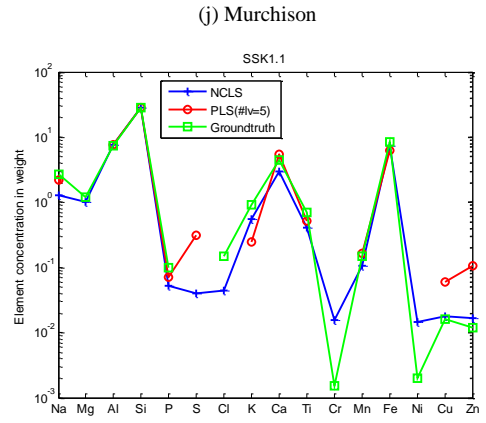(e) I555
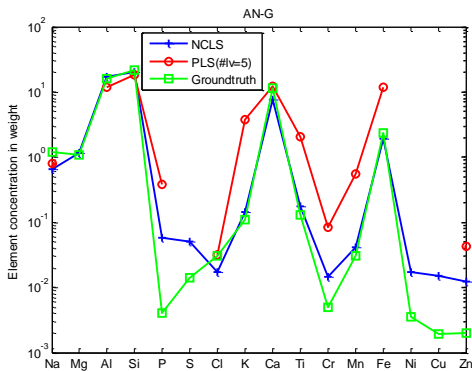
(b) AN-G

(f) JSd-2

(c) BE-N

(g) Mica-Fe

(d) GXR-1

(h) Mica-Mg

(i) Milbillie

(j) Murchison

(k) SSK1.1