The Quantitative Analysis of Laser-Induced Breakdown Spectroscopy (LIBS) Using Partial Least Squares Regression Sub-models Based on Particle Swarm Optimization. Li Zhang[1], Zhongchen Wu[2]*, Zongcheng Ling[2], Xueqiang Cao[1], Jialun Li[1],Kaichen Guo[2]. [1]School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai 264209;[2]Institute of Space Science, Shandong University, Weihai 264209, China , (z.c.wu@sdu.edu.cn)

The Quantitative Analysis of Mars Science Laboratory Expanded Geochemical Database: Laser Induced Breakdown Spectroscopy(LIBS) is a powerful tool for geochemical applications because it is able to rapidly achieve multi-elemental information with little or without sample preparation[1].The largest LIBS public-availabe datasets maybe provided by the ChemCam LIBS instrument on Curiosity rover , which has obtained more than 300,000 spectra of rock and soil since landing at Gale Crater in 2012[2].

The accuracy of quantitative analysis is an important issue for Chemcam and Supercam[3].Therefore, many effective methods were applied to build quantitative modeling, such as Partial Least Squares Regression (PLSR)[5], Support Vector Regression(SVR)[6], and Artificial Neural Network (ANN) [7], etc.

The published LIBS dataset (totally more than 95,000 spectral data) of 408 geological standard samples of ChemCam calibration samples [2] was used for testing our method in this study.These data are very precious for modeling between material elemental abundant and plasma emission spectral information.

PLS Using Sub-models Method: PLSR is based on linear transition from a large number of original descriptors to a small number of orthogonal factors (latent variables) providing the optimal linear model in terms of predictive ability.

A significant difficulty with the PLS method in analyzing the ChemCam data is the diversity of the materials analyzed. This led to that a single PLS model can't balance the accuracy and extreme compositions. So it is always suitable for majority targets. Samuel M. Clegg, et al. proposed a PLS Sub-models method to overcome this problem[8].Fistly, each sub-model is trained respectively by virtue of its restricted compositional. For example: 'Full' (0-100 wt.%), 'Low' (0 – 50 wt.%), 'Middle' (30 – 70 wt.%), and 'High' (60 – 100 wt.%). This is shown in Fig.1. It can be seen that the sub-model ranges overlap each other, which is important to avoid discontinuities in the final combined results. When using this model, a 'Full' model is firstly taken as an initial approximation of the target composition. This is achieved using the result of the 'Full' model, and the result falls within which sub-model range, then suggests that model result should be used instead. If the 'Full' model result falls in one of the overlapping ranges, the two overlapping sub-models are combined using a linear weighted sum, such that the sub-models blend together smoothly.

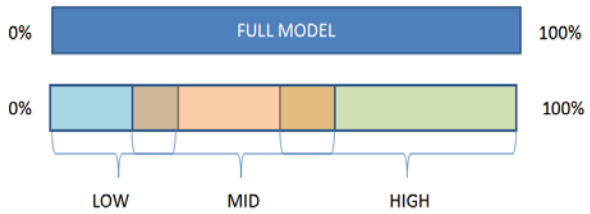

Fig.1 PLS-Sub-models Training

Choosing Sub-models parameters: The conventional choice of sub-model ranges is user-defined, and keep a balance between including a sufficient number of spectra to result in a reliable model.

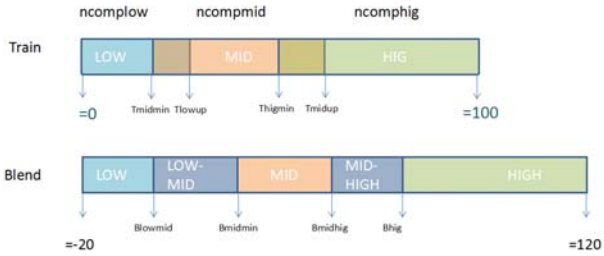

Fig2. The parameters need to be given in PLS-Sub-Model

As shown in Fig2, there are many parameters needed to be reset including 4 training boundaries, 4 blending boundaries, and 4 numbers of components. This increases the uncertainty of the model. Meanwhile, the content of different elements is quite different, for example the range of SiO_2 content is between 0% and 98%, and the range of TiO_2 content is between 0% and 5.81%. Obviously, their definition of the 'Low' 'Middle' 'High' are quite different according to various element species. Different elements models need to reselect and adjust parameters again, which is time-consumingand inaccuracy for modeling.

Parameter Optimization Using Particle Swarm Optimization (PSO) Method: Since the parameter settings affect the performance of predicting the composition. Therefore, selecting the optimal parameters is crucial when employing PLS Sub-models method. In this study, the PSO algorithm is utilized to select the best parameters of the sub-model. We use the Root Mean Square Error (RMSE) as the fitness function as follows :

$$V_{i,j}(k+1) = \omega_v * V_{ij}(k) + c_1 rand() * (P_{best_{i,j}}(k) - x_{i,j}(k)) + c_2 * rand() * (g_{best_j}(k) - x_{i,j}(k)) \quad (1)$$

$$x_{i,j}(k+1) = x_{i,j}(k) + \omega_x V_{i,j}(k+1) \quad (2)$$

Where $X_i = (X_{i1}, X_{i2}, \ldots, X_{iD})^T$ is the i^{th} particle from the initial population of the size of i=1,2,…,N and a dimension of j=1,2,…,D. $V_i = (V_{i1}, V_{i2}, \cdots, V_{iD})^T$ is the velocity of each particle X_i in the population. Rand() represents a random number between zero and one while the individual and global extreme values are represented by $P_{best_{i,j}}$ and g_{best_j}, respectively. The two values of c_1 and c_2 are usually within [0,2], we choose 1.6 and 1.5 respectively. And $\omega_v = 0.9, \omega_x = 0.6$ .The optimized flow chart is shown in Fig 3.

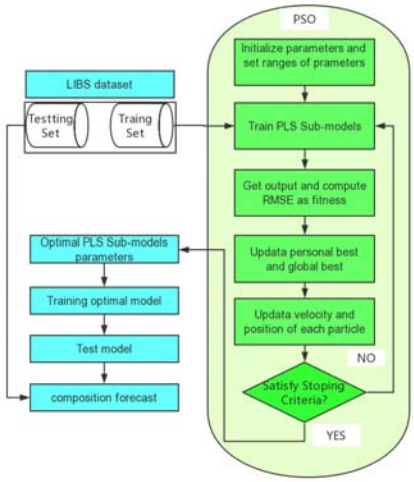

Fig.3 The PSO Optimized Parameters Flow Chart

Table1. Settings for Sub-models

		S_iO_2 (wt.%)		K_2O (wt.%)	
		Manual	PSO	Manual	PSO
TRAINING	LOW	0.0-50.0	0-36.0	0.0-2.0	0.0-0.5
	MID	30.0-70.0	20-56.6	——	0.235-4.0
	HIG	60.0-100.0	50.0-100.0	1.5-100	1.5-100.0
PREDICTIVE	LOW	-20.0-30.0	-20.0-25.0	-20-1.5	-20-1.0
	LOW-MID	30.0-50.0	25.0-30.0	——	1.0-1.4
	MID	50.0-60.0	30.0-41.4	——	——
	MID-HIG	60.0-70.0	41.4-61.8	1.5-2.0	1.4-2.0
	HIG	70.0-120.0	61.8-120.0	2.0-120.0	2.0-120.0

Table2. The test set results of RMSE

Composition	PLS	PLS-Submodel	PSO-PLS-Submodel
SiO_2	5.66	4.62	3.62
T_iO_2	0.51	0.46	0.46
AL_2O_3	2.79	2.26	2.31
FeO_T	3.34	2.21	2.30
MgO	1.43	1.19	1.07
CaO	1.80	1.89	1.08
Na_2O	0.60	0.57	0.44
K_2O	0.78	0.72	0.68

RESULTS: Table 1 is the parameters set by the conventional and our strategies. Only SiO_2 and K_2O modeling parameters are listed due to space limitations.The details will be given in subsequent articles. Fig.4 shows plots of the training set (red), test set (blue) predictions by the PSO blended sub-models against the known compositions of SiO_2. Perfect predictions would fall along the 1:1 line. As shown in Table2, the PSO blended sub-model RMSEP is lower than the full model and blended Sub-models PLS RMSEP for all elements except Al_2O_3 FeO_T, which maybe interfered by a few extreme composition.

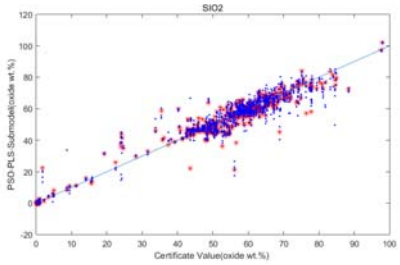

Fig.4 Plots of SiO_2 predictions vs. certificate values. *Perfect predictions would fall along the 1:1 line.*

Conclusion: We demonstrated that combine PLS Sub-models and PSO methods can improve the quantitative accuracy of LIBS spectral analysis. Some spectral with extreme composition situation should be follow up this issue.

Acknowledgements: This work was supported by Natural Science Foundation of China (41573056). Natural Science Foundation of Shandong Province(2016ZRE2703). We gratefully acknowledge ChemCam team provide the data set.

References： [1] EI Haddad et al.(2014),Spectrochimica Acta Part B,97, 57-64;[2] Clegg,S.M. et al.(2017),Spectrochimica Acta Part B,129,64-85;[3] Yang et al.(2015),pectrochimica Acta Part B,107, 45-55.;[4] Boucher et al.(2015),pectrochimica Acta Part B,107,1-10;[5]Takahashi et al.(2015), APPL OPTICS, 57,5872;[6]Yelameli et al.(2016),OCEAN 2016,1-4;[7] Li et al.(2017), APPL OPTICS, 56, 935;[8]Anderson et al.(2017),Spectrochimica Acta Part B,129, 49-57