QUANTIFICATION OF IRON IN MARTIAN BASALT USING LIBS DATA WITH PENALIZED SHRUNKEN REGRESSION METHODS.

Hongchun Bai, Zongcheng Ling*, Changqing Liu, Shangke Tian, Shandong Key Laboratory of Optical Astronomy and Solar-Terrestrial Environment, Institute of Space Sciences, Shandong University, Weihai Shandong 264209, China (zcling@sdu.edu.cn).

Introduction: Mars is a red planet with basaltic crust rich in iron element. Laser induced breakdown spectroscopy (LIBS) is a quick element analysis technique which has been used by Curiosity rover in Mars exploration. LIBS spectra of iron have abundant peaks (mainly at 240~340 nm, 380~480 nm and 490~570 nm). How to better utilize those abundant peaks for precise quantitative analysis of iron is a key question for LIBS application in Mars exploration.

Root-mean-square error (RMSE) value of FeO_T in old PLS1 and ICA model is larger than other major element except SiO_2 [1]. Prediction error of iron decreased when FeO_T was predicted by the model which choose different latent variables at various range (wt.%) of iron (PLS1-SM) [2], which suggest the significance of input data for the quantitative model of iron. David et al. [3] cautiously processed input data of their calibration model of iron based on the correction coefficient between spectra and abundance of iron. However, the relationship between iron and intensity of each channel may be complex due to element substitution and paragenetic association. With the purpose of better quantitative analysis of iron, we investigated LIBS peaks overlay between iron and other major elements, feature selection of iron, and compared the results of different quantitative iron models for basaltic samples.

Methods: There are three penalized shrunken regression methods (ridge regression, lasso regression and elastic net) [4], all of which aim to decrease the size of channels for LIBS spectra. The penalty of ridge regression is L2 penalty, of lasso regression is L1 penalty and of elastic net is combination of L1 and L2 penalty (a*L1+b*L2). Briefly, L1 is the sum of absolute values of coefficients and L2 is the sum of squared coefficients. The amount of penalization for those three regularization techniques were calculated by 5-fold cross validation; L1_ratio (a/(a+b)) in elastic net is a default number (0.5). PLS1-SM also was used to analyze iron in basalt in addition to three penalized shrunken regression models. LIBS spectra of 21 basalts or standards mixed basalts were selected from ChemCam 408 standards [1].

Results and discussions: Analyses of peak overlay between iron and other major elements indicate that there are a large number of overlapping iron peaks. Iron lines strongly overlap with Ti and Al. The number of iron lines is 54 (39) when interval of peak center between Fe and Ti (Al) less than 0.1 nm, and the number of overlapped iron lines is 108 (for Ti) and 48 (for Al) when the threshold of interval of peak center is 0.2 nm. Effect of peak overlay affect the accuracy of quantitative analyses iron. Feature selection could help to diminish the effect of peak overlay by choose un-overlapped lines. The key function of feature selection is the removal of redundant and irrelevant features; redundant features are strongly correlated with each other. We selected the 17 highest scoring channels of basalt using SelectKBest in order to check the redundancy of channels after feature selection. Those channels could be classified into six peaks (302.1, 299.5, 259.9, 748, 311.1 and 248.4 nm).

LIBS spectral features were selected by lasso and elastic net (Fig. 1). Features number of elastic net is larger than lasso due to the difference of both penalties. Though the number and intensity of the coefficients for both models are diverse, there are some features at same wavelength. For example, features selected by both models all have a negative coefficient at 334.9 nm and a positive coefficient at 259.9 nm which is one of the highest scoring channels for iron in basalt. Feature at 259.9 nm is a peak of Fe II and that at 334.9 nm is a peak of Ti II. Peaks at 259.9 and 334.9 nm are not disturbed by peaks of other elements even Mg have peaks (258.6 and 260.7 nm) around 259.9 nm. Area of peak 259.9 nm has a positive relation with Fe (wt.%) in basalt and the data were best fit with a second-order polynomial. Area of peak at 334.9 nm negatively correlated with Fe (wt.%) in basalt.

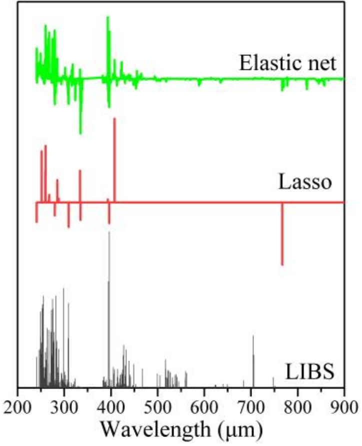

Figure 1. LIBS spectra of Fe [5] and coefficient of features selected by lasso and elastic net. Vertical axis is scaled for clarity.

LIBS spectra of basalts or mixed basalt selected from ChemCam 408 standards were classified into two sets, one is test set (five standards) and remaining is training set. Five models (elastic net, lasso, ridge, PLS1 and PLS1-SM) were used to predict iron within standards. The Root-mean-square error of prediction (RMSEP) of each model is 1.73 (for elastic net), 1.26 (for lasso), 2.03 (for ridge), 3.87 (for PLS1) and 1.69 (for PLS-SM), and them are comparable except PLS1. Those penalized shrunken models were then used to investigate LIBS spectra measured on Mars.

The models were trained with all laboratory basalt spectra, and then predicted the Fe (wt.%) in four CCCTs. LIBS spectra of three CCCTs (NAu2HiS, NAu2MedS and NAu2LowS) which have basalt matrix and anther CCCT (basaltic Shergottite) were download from Planetary Data System (https://pds-geosciences.wustl.edu/msl/msl-m-chemcam-libs-4_5-rdr-v1/mslccm_1xxx/data/). RMSEPs of four CCCTs are shown in Fig. 2. RMSEPs of Fe in those CCCTS which have basalt matrix when using elastic net and ridge are less than Ref.[1] and when using lasso are comparable with Ref.[1]. Same matrix (basalt) also contributes to the smaller RMSEP besides quantitative analyses models.

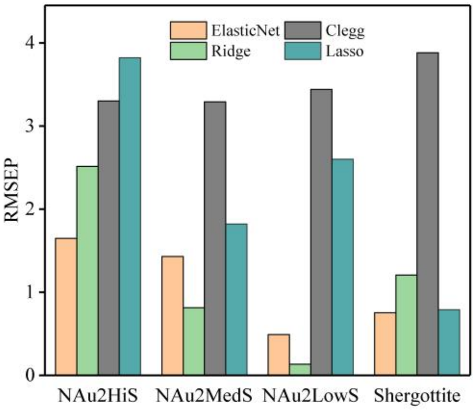

Figure 2. RMSEPs of three models (Elastic net, Ridge, Lasso and Clegg et al., 2017) for four CCCTs.

Those different models were used to analyses the iron within igneous rocks (Fig. 3) in Gale Crater. Targets Kodak and Ravalli were both measured by LIBS and APXS. Abundance of FeO_T for Kodak predicted by the ChemCam team are largely varying and are far away from the value of APXS. Our predictions are between the values of APXS and MOC, and the variations between points is much less. For another target (Ravalli), the values measured by APXS and the predictions of ChemCam team are more consistent compared to those of ridge and elastic net. Differences of measured area and deep between APXS and LIBS are likely responsible for the difference of FeO_T. Predictions of FeO_T for two basaltic targets, Ashuanipi and La_Reine, are larger and more variable when using the model of ChemCam team and lasso compared to ridge and elastic net. Predictions of different points on same target using ridge and elastic net are more robust than lasso and combination of PLS-SM and ICA. Predict performances of lasso and the combined model are similar (larger and varying). In a word, penalized shrunken regression methods are generally much more efficient regression models in LIBS study particularly for special rock or mineral group rather than all calibration targets. More detailed analyses of those models especially elastic net will be done based on more LIBS standards [6].

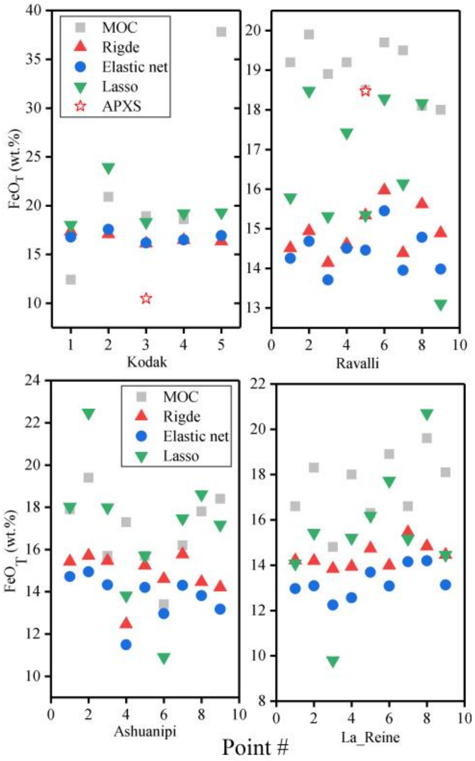

Figure 3. Predictions of iron with different models.

Acknowledgments: We thank the fundings from the National Natural Science Foundation (U1931211, 41972322, 11941001), the Natural Science Foundation of Shandong Province (ZR2019MD008). This work is also supported by the Pre-research project on Civil Aerospace Technologies No. D020102 funded by China National Space Administration (CNSA).

References:

[1] Clegg S. M. et al. (2017) *SAB*, *129*, 64–85. [2] Anderson R. B. et al. (2017) *SAB*, *129*, 49–57. [3] David G. et al. (2020) *JGR*, *125*. [4] Ogutu J. O. et al. (2012) BMC *Proc 6*, S10. [5] Cousin A. et al. (2011) *SAB*, *66*, 805–814. [6] Liu et al. (2021) 52nd LPSC Abstract #1743.