Electronic Supplementary Information of

"Improvement in analytical performance of underwater LIBS signal by exploiting the plasma image information"

Qingyang Li, Ye Tian, Boyang Xue, Nan Li, Wangquan Ye, Yuan Lu and Ronger Zheng*

Optics and Optoelectronics Laboratory, Ocean University of China,

Qingdao 266100, China

*Corresponding author: rzheng@ouc.edu.cn

1 Determination of the detection delay time

The detection delay time used in this work was determined according to the signal-to-noise ratio (SNR) of the spectral line of analytical elements. The spectra (average of 500 laser pulses) and SNRs of different analytical lines under different detection delay time (50–500 ns) were shown in Fig. S1 and Fig. S2 respectively.

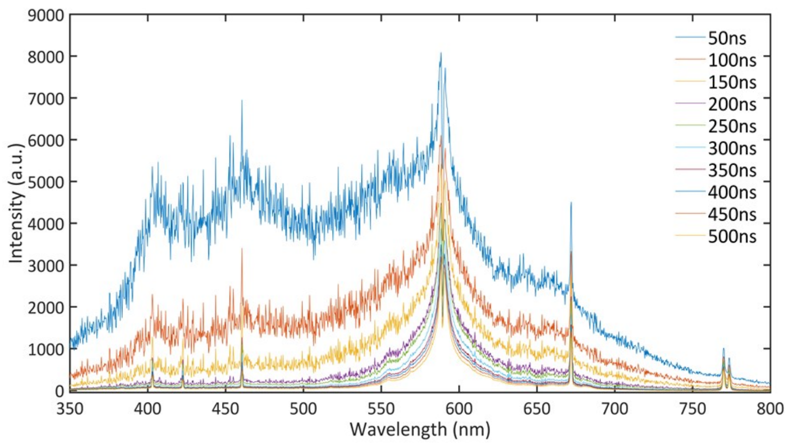

Fig. S1 Typical LIBS spectra at different detection delay time from 50 to 500 ns.

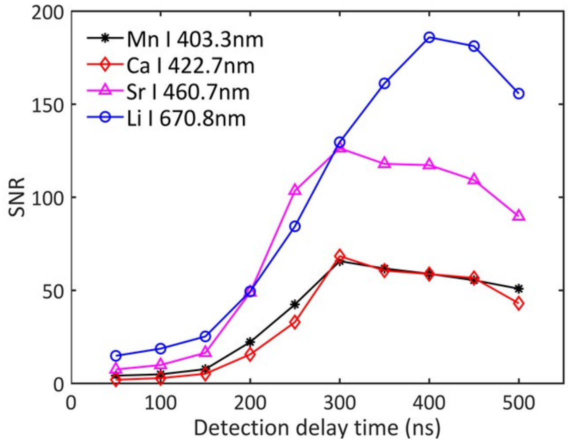

Fig. S2 SNRs of different analytical lines as a function of detection delay time.

According to Fig. S2, the SNRs of different element reached the maximum at different delay time. The SNRs of spectral lines of Mn I 403.3 nm, Ca I 422.7 nm and Sr I 460.7 nm reach their maximum at 300 ns after the laser pulse while the SNR of Li I 670.8 nm reaches a maximum at 400 ns. However, due to the detection gate width was set to cover the entire lifetime of the plasma, the earliest time that the maximum SNRs correspond to was commonly selected as the detection delay time. In this way, the best SNRs of all the analytical spectral lines would be included during the recording time. In order to realize the simultaneous detection, the delay time of image collection was set the same as the spectra. Under the different experimental conditions or different detected elements, the detection delay time may vary because of the different temporal evolution. In this work, the detection delay time was set at 300 ns.

2 Calculation of plasma morphology

2.1 Plasma area calculation

For each plasma image, the “plasma area” was defined as the region of which the gray value is more than 1/e times of the maximum intensity. As the pixel size of ICCD is 13μm/pixel and the image magnification is 2.96, the value of plasma area can be calculated as follows:

$$Area = N \times (13/2.96)^2 \tag{1}$$

where *N* is the number of pixels within the plasma area.

2.2 Plasma flatness calculation

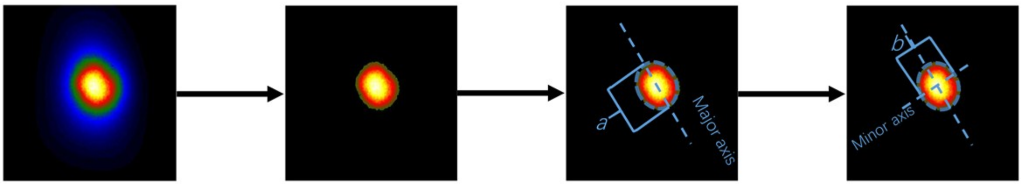

Fig. S3 Schematic diagram of plasma flatness calculation

As Fig.S3 shows, the plasma was considered as an approximate ellipse, therefore, the "plasma flatness" could be defined by *b/a*, where *b* is the length of the minor axis of the ellipse while *a* is the length of the major axis. Principal component analysis (PCA) was used to find the direction with the largest variance of the projection data and the direction was considered as the direction of the major axis. For a certain dataset *D(X,Y)* (location of the pixels within the plasma area), the principal component P_{ji} can be calculated as follows:

$$P_{ji} = a_{j1} \times X_i + a_{j2} \times Y_i = A_j D \quad (2)$$

where *i* is the number of the data point, *j* is the number of the principle component, $A_j(a_{j1},a_{j2})$ is the direction of the j_{th} principal component.

Then the variance of the principal component *Var(P)* can be calculated by Eq. (3):

$$Var(P) = Var(AD) = (AD)(AD)' = ADD'A' = V \quad (3)$$

As the principal component *P* is irrelevant and independent with each other, the *V* should be a diagonal matrix. Cause *A* is the orthogonal matrix, the Eq. (3) can be transformed as follows:

$$DD'A' = \Sigma A' = A'V \quad (4)$$

Where Σ is the covariance matrix of *D.*

The A' is the feature vectors of Σ. By solving the Eq. (4), the eigenvalues and eigenvectors will be obtained. The eigenvector corresponding with the largest eigenvalue is the direction of first principal component and the projection data will have the largest variance. The largest distance between 2 points in the projection data will be the length of the major axis *a*. The direction perpendicular to the major axis was taken as the direction of minor axis and the length of the minor axis *b* of the ellipse could be obtained.

2.3 Plasma central position calculation

As for the fluctuations of the central position, the position of the maximum intensity in the averaged image of all data was used as the origin of coordinates (x_0,y_0), and then the distance *d* could be calculated as Eq. (5) to indicate the central position of plasma for every single image:

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2} \tag{5}$$

where *(x,y)* is the location of the maximum intensity in each plasma image.

3 Paramater selection in PLSR model

In PLSR, the number of PLS components (expressed as k) is usually determined by cross validation, however, the k determined by this method can be too large, resulting in overfitting problems. Therefore, we used the ratio of the cumulative sum of the variances explained by the PLS components to determine the optimum k value. As Fig.S4 shows, when the number of PLS component reaches 7 (the ratio reaches 98%), the ratio increases slowly. Therefore, 7 PLS components were used, which means the number of the iterations is 7. By this way, we can fully extract the information contained in the plasma image and keep a good generalization ability of the model at the same time.

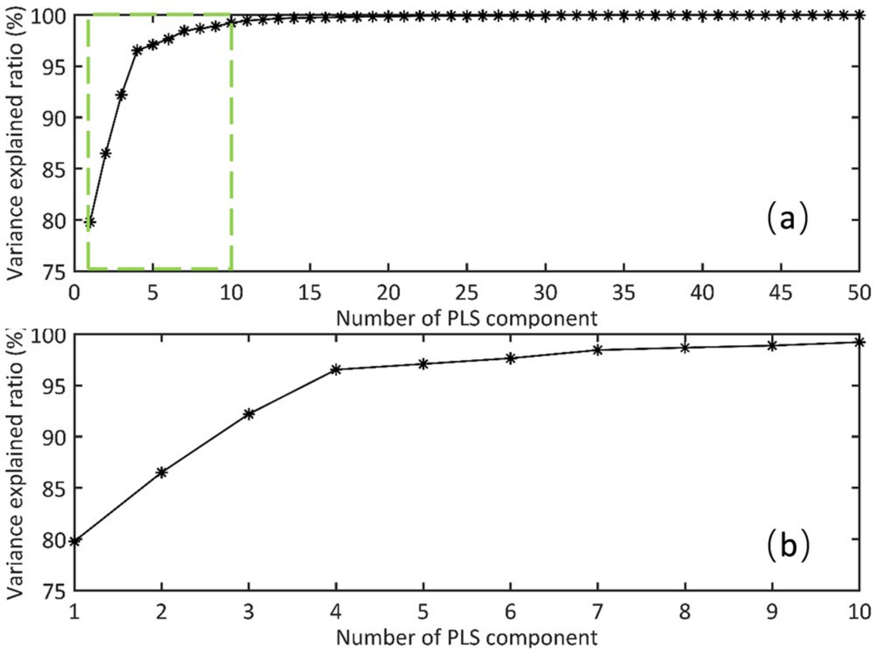

Fig. S4 Variances ratios explained by (a) top 1-50 PLS components (b) top 1-10 PLS components of plasma image (Average of the Li, Mn, Sr, Ca model)