Spectrochimica Acta Part A 86 (2012) 631–636

Contents lists available at SciVerse ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Short communication

Visualizing excipient composition and homogeneity of Compound Liquorice Tablets by near-infrared chemical imaging

Zhisheng Wu[a,b], Ou Tao[a,b], Wei Cheng[a,b], Lu Yu[c], Xinyuan Shi[a,b,*], Yanjiang Qiao[a,b,*]

[a] Beijing University of Chinese Medicine, Beijing, 100102, China
[b] The Key Laboratory of TCM-Information Engineering of State Administration of Traditional Chinese Medicine, Beijing, 100102, China
[c] PerkinElmer Instrument Co., Ltd., Beijing, 100022, China

ARTICLE INFO

Article history:
Received 26 August 2011
Received in revised form 10 October 2011
Accepted 14 October 2011

Keywords:
Near-infrared chemical imaging
Spatial distribution
Uniformity
Compound Liquorice Tablets

ABSTRACT

This study demonstrated that near-infrared chemical imaging (NIR-CI) was a promising technology for visualizing the spatial distribution and homogeneity of Compound Liquorice Tablets. The starch distribution (indirectly, plant extraction) could be spatially determined using basic analysis of correlation between analytes (BACRA) method. The correlation coefficients between starch spectrum and spectrum of each sample were greater than 0.95. Depending on the accurate determination of starch distribution, a method to determine homogeneous distribution was proposed by histogram graph. The result demonstrated that starch distribution in sample 3 was relatively heterogeneous according to four statistical parameters. Furthermore, the agglomerates domain in each tablet was detected using score image layers of principal component analysis (PCA) method. Finally, a novel method named Standard Deviation of Macropixel Texture (SDMT) was introduced to detect agglomerates and heterogeneity based on binary image. Every binary image was divided into different sizes length of macropixel and the number of zero values in each macropixel was counted to calculate standard deviation. Additionally, a curve fitting graph was plotted on the relationship between standard deviation and the size length of macropixel. The result demonstrated the inter-tablet heterogeneity of both starch and total compounds distribution, simultaneously, the similarity of starch distribution and the inconsistency of total compounds distribution among intra-tablet were signified according to the value of slope and intercept parameters in the curve.

1. Introduction

Since 2002, the Food and Drug Administration's Process Analytical Technology (PAT) initiative has encouraged pharmaceutical manufacturers to adopt new technologies for timely measurement of critical product and process attributes in the manufacturing environment [1]. Near-infrared chemical imaging (NIR-CI) is a novel technique for acquisition of qualitative and quantitative information about both spatial and chemical characteristics of individual components at the micro scale [2–10]. It is well known that a formulation with identical ingredients can produce varying therapeutic performance depending upon how the ingredients are spatially distributed in the final matrix. Although conventional methods such as high performance liquid chromatography (HPLC) can provide accurate content on active pharmaceutical ingredients (API), they cannot provide information about the spatial distribution of individual components, making it difficult or impossible to trace the source of failures or anomalies.

Chinese Medicine Tablet (CMT) has its own characteristic compared with synthetic drug tablets. One of the most important differences is the distribution homogeneity of CMT including complex plant extraction (e.g. fluid extraction and plant powder) and excipient. More attention should be paid to the component distribution because complex compounds distribution will seriously affect the dissolution rate and therapeutic performance of CMT. However, no paper has reported on NIR-CI to understand and characterize the compound distribution of CMT. Compound Liquorice Tablets (CLT) is a antitussive expectorant in Chinese Pharmacopoeia, which can cover in a mucosal inflammation of the throat to play antitussive effect [11,12]. The formulation of CLT is composed of *Radix glycyrrhizae*, *Papaver somniferum* L., sodium benzoate, anise oil, camphor and starch. In this paper, CLT was taken as an example, and NIR-CI served as an ideal tool to visualize the excipient distribution and homogeneity.

Data cube analysis about NIR-CI can be accomplished at several different levels to produce the contrast of heterogeneous samples. Many excellent reviews covered general aspects of chemometrics tools about NIR-CI (e.g., principal component analysis (PCA), basic analysis of correlation between analytes (BACRA), multivariate curve resolution-alternating least squares (MCR-ALS)) [2,13–16]. Besides, some authors reported image-processing tools to analyze

* Corresponding authors at: Beijing University of Chinese Medicine, Beijing, 100102, China. Tel.: +86 10 84738621; fax: +86 10 84738621.
E-mail addresses: shixinyuan01@163.com (X. Shi), yjqiao@263.net (Y. Qiao).

1386-1425/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.saa.2011.10.030

data cube (e.g., histogram graph, macropixel, fractals) [17]. The objective of this paper was to demonstrate a novel application about NIR-CI, the suitability of chemometrics tools (e.g. BACRA, PCA) and a new algorithm about an image-processing tool named Standard Deviation of Macropixel Texture (SDMT).

2. Standard Deviation of Macropixel Texture method

Standard Deviation of Macropixel Texture was introduced as a measurement of texture information based on binary image. Firstly, an important procedure was binary-coding process. NIR image was composed of different gray level values, which in fact were not of uniform size. Therefore, net unit was adopted to divide the whole image, and a gray level threshold value k was fixed before net unit was divided. Generally, the gray level was formed of 0–255 colors. The set of image f gray level was defined to be Z where,

$$Z \equiv [Z_l, Z_t]. \quad (1)$$

when the f gray level of the image was greater than the threshold value k, the pixel was defined as 1; when the f was less than k, the pixel was defined as 0. Namely, the mathematical expression was:

$$f_k(x, y) = \begin{cases} 1 & f(x, y) \geq k, \\ 0 & f(x, y) < k. \end{cases} \quad (2)$$

Thus, the binary image was created. Threshold value k was determined according to many methods. In this paper, a procedure was introduced to calculate threshold value k by global optimization method [18]. Secondly, every interception binary image of 600×600 pixels was divided into different sizes length of macropixel and the size length ε ($\varepsilon = 2, 3, 4, 5, 6, \ldots, 200, 300, 600$) was the 600 divisor except 1. Thus, the number of zero values in each macropixel was counted to calculate standard deviation. Finally, a curve fitting graph was plotted on the relationship between standard deviation and the size length of macropixel. Slope and intercept parameters were important indicators about the heterogeneity of compounds distribution.

3. Results and discussion

3.1. Determination of starch and plant extraction distribution domain by BACRA method

Different pretreatment methods were investigated to obtain high quality of spectra. Fig. 1 showed NIR spectra extracted from imaging graph of starch tablet. The correlation coefficients between starch spectrum and each absorbance spectrum of four samples were calculated. The result was listed in Table 1. Then, taking the highest correlation coefficient parameter as criteria, the combination of smoothing average with a window size of 9 points and normalize provided the best pretreatment results. Based on the preprocessed spectra, the correlation-coefficient images showed in Fig. 2. The correlation-coefficient maps of starch distinguished the zones from the highest to lowest correlation coefficient according to the spatial distribution of analytes in the sample. The red domain indicated the existence of starch material because the correlation coefficients were above 0.95. Simultaneously, the situation of low correlation coefficients suggested the distribution of plant extraction. Therefore, the spatial distribution of starch and plant extraction could be elucidated by BACRA method.

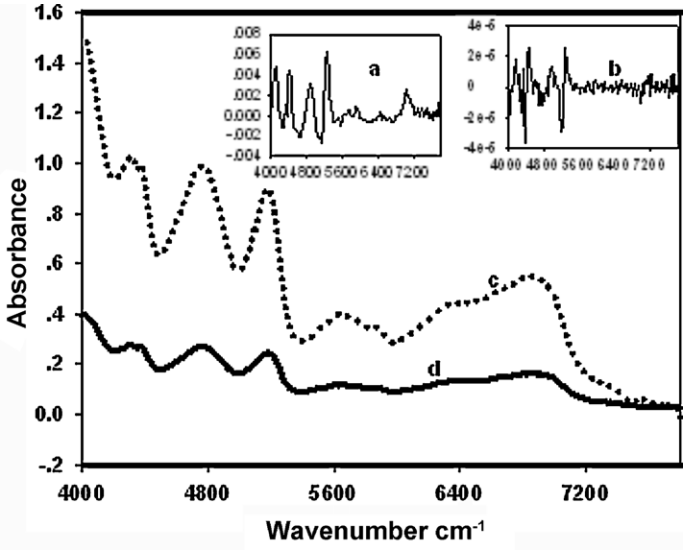

Fig. 1. Starch spectra extracted from imaging graph of starch tablet. (a) Spectrum of normalize + smooth + 1st, (b) spectrum of smooth + 2nd, (c) spectrum of smooth + normalize, and (d) raw spectrum.

3.2. Homogeneity analysis of compound distribution by histogram graph

Fig. 2 showed the agglomerates such as blue domains. However, "human eyes" discrimination is fuzzy. For a technique to be useful in quality control, the results should be objective, reproducible, and quantitative. An approach for deriving numerical information was to represent image data as histograms, enabling statistical and objective analysis. Fig. 3 showed the histograms of correlation-coefficient maps. Histograms were bar graphs of correlation coefficient value at each pixel. When more pixels had the same correlation coefficient, the bar for that x-value was higher. From the histograms, the graph fitted a Gaussian distribution, offering a reliable qualitative parameter for the compound distribution.

Furthermore, each histogram showed a different peak shape. Table 2 depicted statistical analysis and provided a quantitative representation. The parameters described were mean, standard deviation, skewness and kurtosis. The mean was a measure about the center of the distribution. In this case, mean was relatively stable in all pixels. For standard deviation, the value of sample 3 was large, but it did not fully demonstrate the prevalence of high or low abundance domains because skewness and kurtosis parameters characterized the abundance domains. The skewness was a measurement of asymmetry. The larger the skewness was, the longer the tail on the right side was. A negative skewness indicated that the tail was on the left side. The skewness of sample 3 was −2.17. It illuminated that the heterogeneity of plant extraction distribution in sample 3 was worse than other samples. Simultaneously, the kurtosis was a measure of the flatness or "peakedness". High kurtosis was the result of infrequent extreme deviations. The kurtosis of sample 3 was 17.85. Obviously, the result elucidated that the component distribution of sample 3 was relatively heterogeneous. The above four parameters provided a quantitative presentation, and made it possible to create reliable, reproducible results.

3.3. Homogeneity analysis of the compound distribution using PCA scores layers

In chemical imaging, spectral components were used only for identification purpose, and the interesting information was spatial component. For this reason, PCA was used to describe the

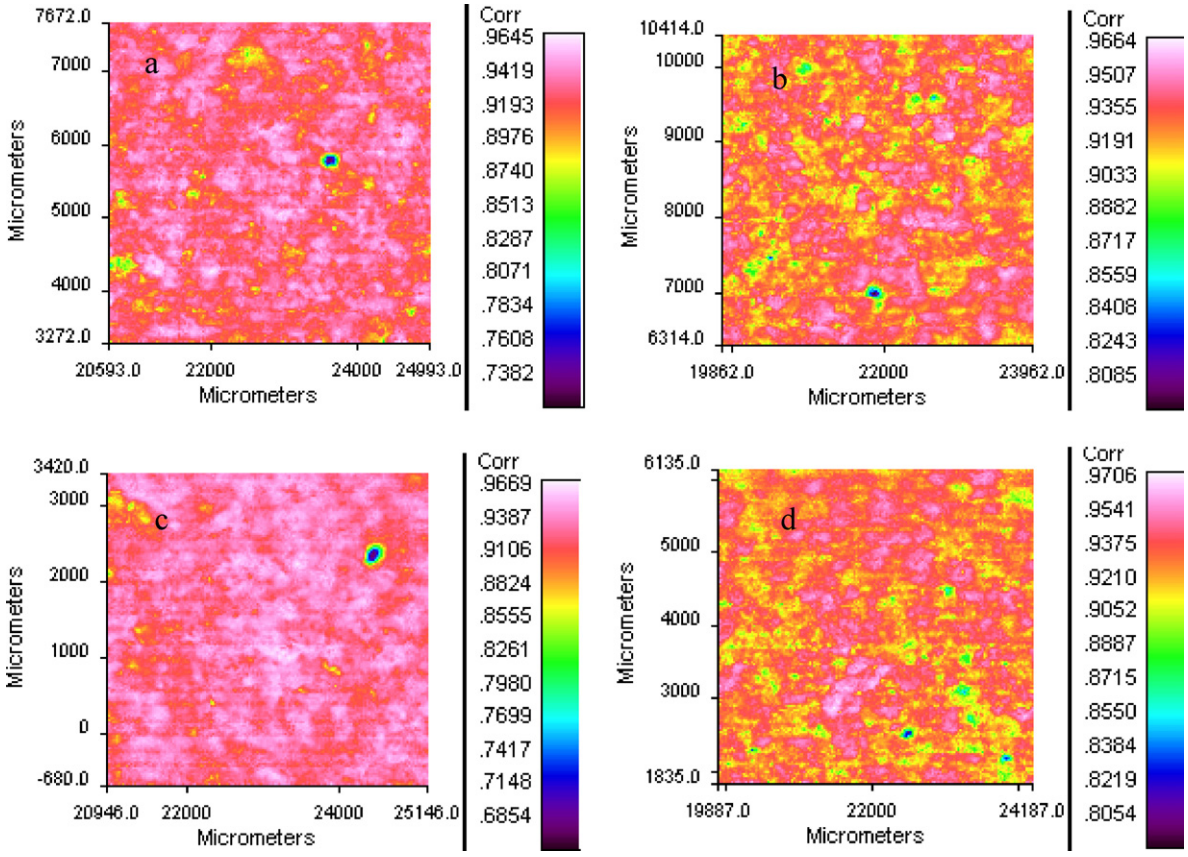

Fig. 2. BACRA analysis of CLT image. (a) Sample 1, (b) sample 2, (c) sample 3, and (d) sample 4.

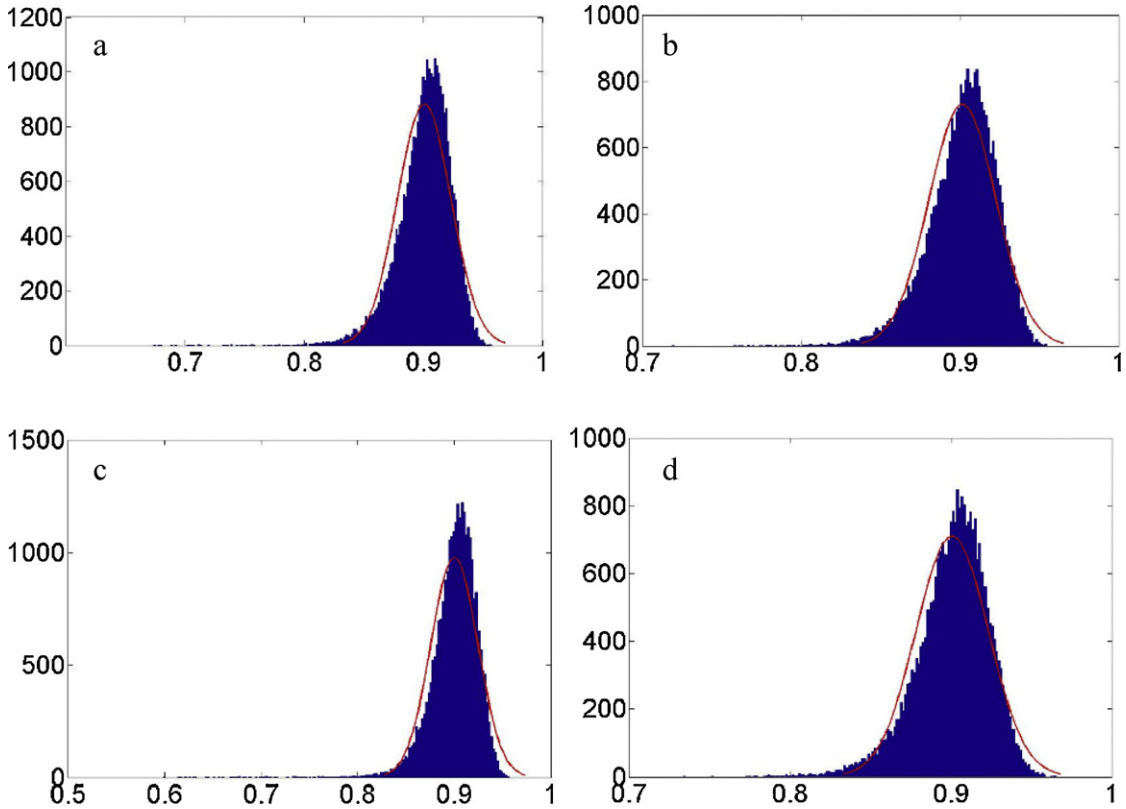

Fig. 3. Histograms of correlation-coefficient maps for starch. *X*-axis represents correlation coefficient value and *y*-axis represents the number of pixels. (a) Sample 1, (b) sample 2, (c) sample 3, and (d) sample 4.

Table 1
The result of the highest correlation coefficient between starch spectrum and the spectrum of each pixel developed with different pretreated methods.

Pretreated methods	Sample 1	Sample 2	Sample 3	Sample 4
Raw spectra	0.9622	0.9638	0.9630	0.9683
Smooth + normalize	0.9645	0.9664	0.9669	0.9706
Smooth + 2nd	0.8619	0.8700	0.8690	0.8598
Normalize + smooth + 1st	0.9475	0.9476	0.9489	0.9469

Table 2
Statistical analysis of the image data shown in Fig. 3(a–d).

Statistical parameters	Sample 1	Sample 2	Sample 3	Sample 4
Standard deviation	0.0225	0.0211	0.0243	0.0224
Mean	0.900	0.901	0.900	0.901
Kurtosis	8.618	4.878	17.854	4.496
Skewness	−1.320	−0.855	−2.166	−0.796

important spectral variations with a reduced number of components to improve the image contrast. Factors were removed based on spatial but not spectral information. From Fig. 4, three score image layers were extracted, and the comparison of each sample could be easily seen. The green pixels presented large agglomerates, showing in sample 1, sample 3 and sample 4. The same phenomenon existed in blue domains. Fig. 2 also illuminated the same interesting domains as Fig. 4. This method demonstrated the feasibility of rapid detection of interesting domains (or agglomerates) in complicated CLT. Simultaneously, it should be noted that score image layer did not refer to separate the ingredients. Score image layer was used to rapidly discover which layer was chemically distinct entities and then extracted them.

3.4. Homogeneity analysis by Standard Deviation of Macropixel Texture

The result was further depicted by binary image, showed in Fig. 5. The aggregation of starch distribution in each sample existed in binary image. Fig. 6 showed the result by SDMT method. For starch distribution, the slopes of each fitted line were 2.283, 2.261, 2.268 and 2.228, respectively. For total compounds distribution, it was 0.661, 0.628, 0.800 and 1.225, respectively. The result illuminated the inter-tablet heterogeneity of both starch and total compounds distribution since the slope closing to 0 confirmed the homogeneity of compound distribution. Furthermore, the relative stable slope and intercept parameters in Fig. 6(a) signified

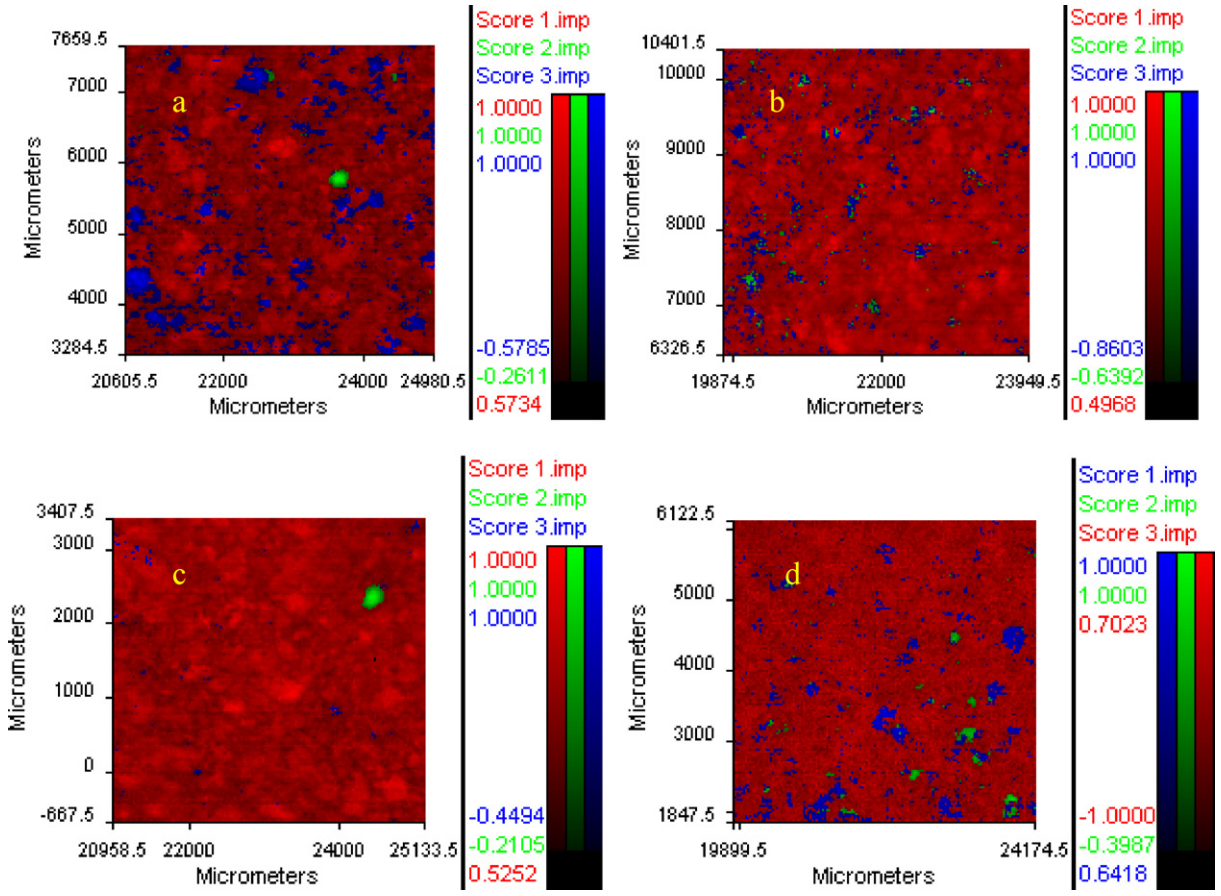

Fig. 4. RGB image of three score image layers by PCA method. (a) Sample 1, (b) sample 2, (c) sample 3, and (d) sample 4.

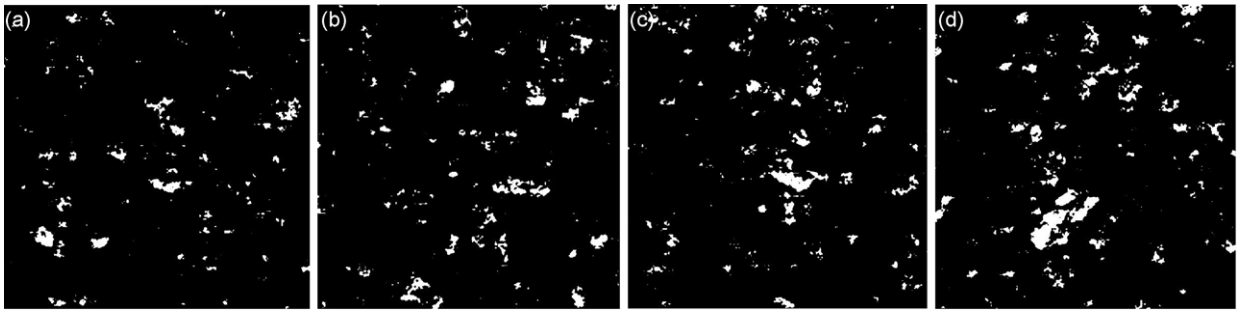

Fig. 5. Binary image of CLT. White region represents the cluster of starch, and black region represents the cluster of plant extraction. (a) Sample 1, (b) sample 2, (c) sample 3, and (d) sample 4.

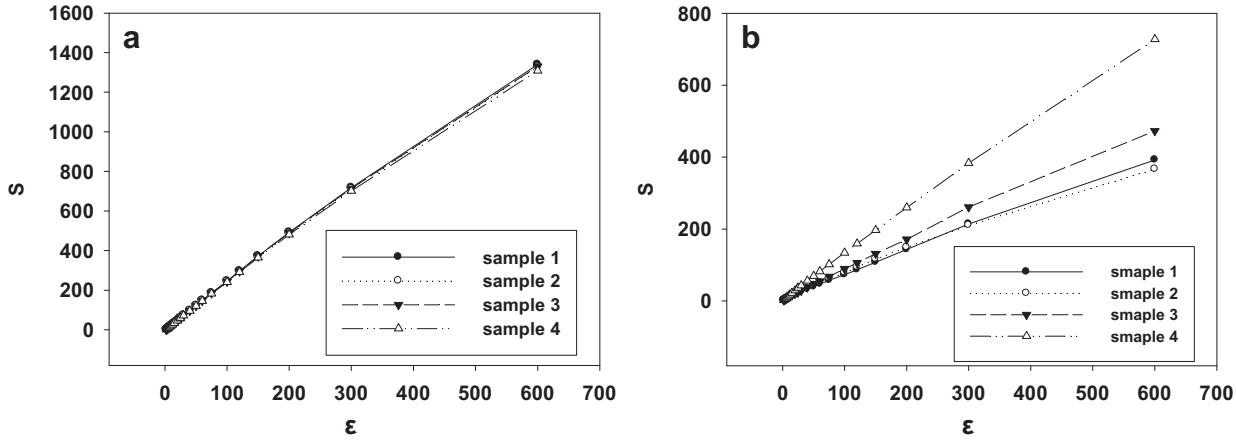

Fig. 6. The result of the SDMT method applied to each CLT image. S represents the standard deviation; ε represents the size length of macropixel. (a) For the starch distribution, the regression equations were expressed by $S = 2.283\varepsilon + 5.808$ (sample 1), $S = 2.261\varepsilon + 5.689$ (sample 2), $S = 2.268\varepsilon + 5.966$ (sample 3) and $S = 2.228\varepsilon + 5.933$ (sample 4), respectively. (b) For the total compound distribution, the regression equations were expressed by $S = 0.661\varepsilon + 4.960$ (sample 1), $S = 0.628\varepsilon + 8.484$ (sample 2), $S = 0.800\varepsilon + 4.968$ (sample 3) and $S = 1.225\varepsilon + 5.331$ (sample 4), respectively.

the intra-tablet similarity of starch distribution. However, four lines were significantly separated in Fig. 6(b), suggesting the intra-tablet inconsistency of total compounds distribution. Finally, the larger the slope was, the more heterogeneous the distribution. Compared with starch distribution in each sample, total compounds distribution presented smaller agglomerate domain (data cannot be shown). This observation elucidated why the slope of total compounds distribution was smaller than starch's. These results explained the potential content uniformity of CLT by SDMT method.

4. Conclusion

NIR-CI is a newly emerging PAT tool. The spatial starch distribution (indirectly, plant extraction) in CLT was visualized using BACRA method. Furthermore, histogram analysis method was applied to assess the homogeneity of the tablets. The result demonstrated that spatial starch distribution was relatively heterogeneous in sample 3 according to four statistical parameters. Additionally, the large agglomerates (heterogeneity) were also detected in each tablet by PCA score image layers. Finally, SDMT method was introduced to characteristic the compound distribution of CLT based on binary image. In conclusion, NIR-CI has a potential of being adopted for spatial distribution and homogeneity of CLT. In addition, chemometrics tools (e.g. BACRA, PCA) and image-processing method (histogram graph, binary image, SDMT) were putted forward for Chinese herbal medicine in the field of quality control by NIR-CI.

Acknowledgments

This work was supported by the Major Special Project Foundation of China (Nos.: 2010ZX09502-002; 2011ZX09201-201-24). Authors also thank Dr. Ahsana Dar Farooq for her assistance in language.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.saa.2011.10.030.

References

[1] http://www.fda.gov/cder/OPS/PAT.htm.
[2] C. Gendrin, Y. Roggo, C. Collet, J. Pharm. Biomed. Anal. 48 (2008) 533–553.
[3] E.N. Lewis, L.H. Kidder, E. Lee, Innov. Pharm. Technol. 17 (2006) 107–111.
[4] G. Reich, Adv. Drug Deliv. Rev. 57 (2005) 1109–1143.
[5] C. Cairos, J.M. Amigo, R. Watt, J. Coello, S. Maspoch, Talanta 79 (2009) 657–664.
[6] C. Ravn, E. Skibsted, R. Bro, J. Pharm. Biomed. Anal. 48 (2008) 554–561.
[7] W.Y. Li, A. Woldu, R. Kelly, J. McCool, R. Bruce, H. Rasmussen, J. Cunningham, D. Winstead, Int. J. Pharm. 350 (2008) 369–373.
[8] N.J.Y. Roggo, A. Edmond, P. Chalus, M. Ulmschneider, Eur. J. Pharm. Biopharm. 61 (2005) 100–110.
[9] F. Clarke, Vib. Spectrosc. 34 (2004) 25–35.
[10] M.B. Jordi Cruza, J.M. Amigo, M. Blanco, Talanta 80 (2009) 473–478.
[11] G.X. Sun, Y. Wang, Y.Q. Sun, K.S. Bi, Anal. Sci. 19 (2003) 1395–1399.

[12] Y.F. Zhuang, S.C. Zhang, J.S. Yu, H.X. Ju, Anal. Bioanal. Chem. 375 (2003) 281–286.
[13] H.B. Qu, X.X. Liu, Y.Y. Cheng, Chem. J. Chinese Univ. 25 (2004) 39–43.
[14] C. Gendrin, Y. Roggo, C. Spiegel, C. Collet, Eur. J. Pharm. Biopharm. 68 (2008) 828–837.
[15] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, S.E.J. Bell, Eur. J. Pharm. Biopharm. 69 (2008) 10–22.
[16] J.M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco, TRAC Trends Anal. Chem. 27 (2008) 696–713.
[17] M.L. Hamad, C.D. Ellison, M.A. Khan, R.C. Lyon, J. Pharm. Sci. 96 (2007) 3390–3401.
[18] R.C. Gonzalez, R.E. Woods, Digital Image Process, 2nd ed., Prentice Hall, Upper Saddle River, 2002, pp. 305–307.