Journal of Pharmaceutical and Biomedical Analysis 148 (2018) 214–223

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Combining near infrared spectroscopy with predictive model and expertise to monitor herb extraction processes

Tongchuan Suo [a,1], Haixia Wang [a,1], Xiaojie Shi [a], Linlin Feng [a], Jiayou Cai [a], Yu Duan [a], Huimin Bao [a], Xiaolin Wu [a], Yue Zhang [a], Heshui Yu [a,*], Zheng Li [a,b,*]

[a] College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, PR China
[b] Tianjin Key Laboratory of Modern Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, PR China

ARTICLE INFO

Article history:
Received 4 August 2017
Received in revised form 7 October 2017
Accepted 9 October 2017

Keywords:
Herb extraction process
Near infrared spectroscopy
Quality of process
Batch process monitoring
Partial least squared regression
Traditional Chinese medicine

ABSTRACT

Albeit extensively utilized, herb extraction process (HEP) is hard to be monitored because of its batch nature and the fluctuating quality of raw materials. Process analytical tools like near infrared spectroscopy (NIRS) can offer nondestructive examinations and collect abundant data of the process, which in principle contain the information about the quality of both the product and the process itself. However, extra effort is often required for the data mining of such process measurements, and extracting knowledge of the quality of process can be even harder. In this study, we take the extraction process of *licorice* as a typical HEP instance, and combine NIRS with classical partial least squared regression (PLSR) and expertise for its on-line monitoring. We show that our scheme effectively extracts information with clear physical meanings, through which we can even uncover the process fault that does not induce evident abnormalities in the product quality. Moreover, the constructed model can continuously evolve with more process data from daily operations, and the idea of the whole framework can be directly generalized to other HEP.

1. Introduction

Herb extraction process (HEP) is widely adopted in the production of health food, dietary supplements and medicine, such as the manufacturing of ginseng root tea and traditional Chinese medicine (TCM). This kind of process refers to utilizing certain solvent (water, ethanol, etc.) to extract the active pharmaceutical ingredients (API) from natural herbs in a batch fashion. Thus, it not only shares typical features of batch processes, including the presence of significant non-linearity, the absence of steady-state operation, etc. [1], but is also affected by the unavoidable property fluctuation of the raw materials due to the less controllable factors like cultivating location, climate, harvest time, etc. As such, in the spirit of the Quality-by-Design (QbD) initiative [2,3], many attempts have been taken to implement process analytical technologies (PAT) for the on-line monitoring of HEP, in order to improve the stability and consistency of the process output.

A typical PAT tool that is widely used in this respect is the near infrared spectroscopy (NIRS), which, combining with chemometrics and other mathematical approaches, has met extensive applications in the food and drug industries [4]. Relevant work can be roughly divided into two types. One is to utilize descriptive learning approaches to read the NIRS spectra, accompanying with certain statistical testings like Hotelling's T^2-test or Q-test to speculate whether the process is under its normal condition. For instance, Rosa et al. described the use of NIRS for the qualification of *Ginkgo biloba* extract, with the help of principle component analysis (PCA) and cluster methods [5]. Xiong and Qu adopted in-line NIRS measurements, combining with multivariate data analyses to monitor the batch-to-batch reproducibility of liquid–liquid extraction processes in TCM manufacturing [6]. Huang and Qu utilized NIRS in conjunction with multi-way PCA for the in-line monitoring of alcohol precipitation [7]. By constructing the control charts with Hotelling's T^2 and squared prediction error (SPE), they illustrated the capability of the methodology for real-time fault detection. While such descriptive learning requires little more input than the NIRS measurements and can be useful in alarming abnormal states (i.e., there is something wrong), the messages from the statistical testings are always abstract and limited such that little can be directly known about the problem (i.e., what is going wrong). The other type adopts predictive learning algorithms to extract information from the NIRS spectra. Although the model construction always requires the help from other analytical technologies (e.g., liquid chromatography), such methods can allow direct computation of, e.g., the API content from on-line NIRS measurements.

* Corresponding authors.
E-mail addresses: hsyu@tjutcm.edu.cn (H. Yu), lizheng@tjutcm.edu.cn (Z. Li).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jpba.2017.10.004
0731-7085/

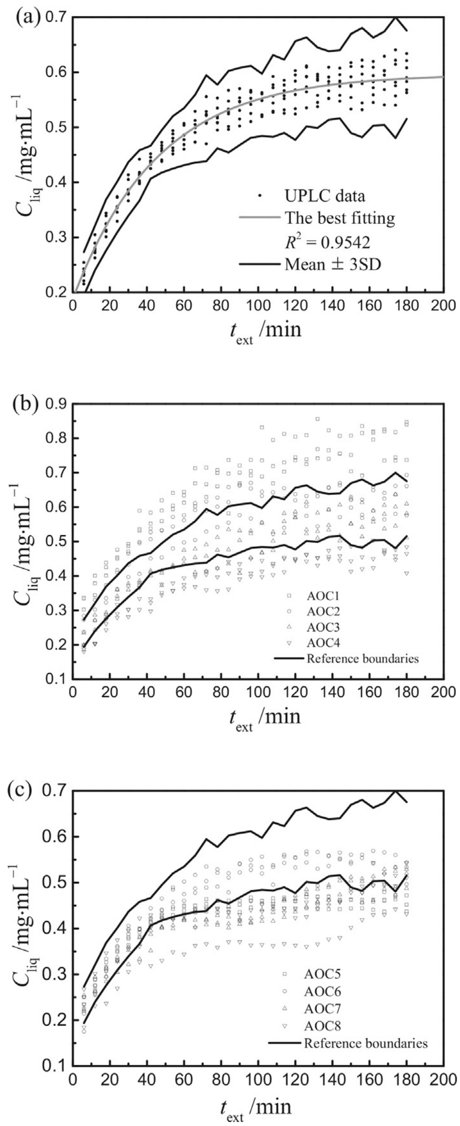

Fig. 1. UPLC measurements for (a) the NOC batches, (b) AOC1–AOC4, which have abnormal solvent consumptions, and (c) AOC5–AOC8, which undergo certain periods of power shutoff. The reference boundaries in panels (b) and (c) are borrowed from the "Mean ± 3SD" curves in panel (a).

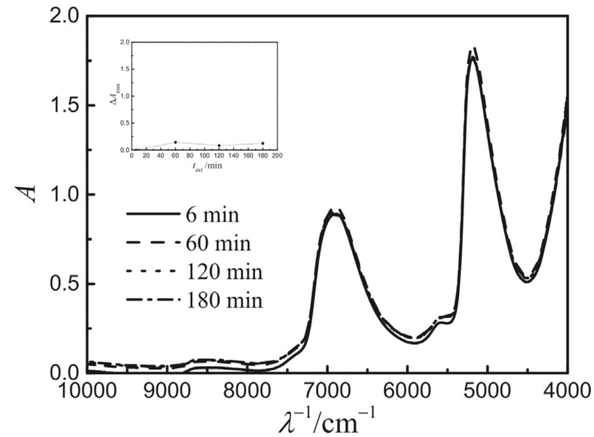

Fig. 2. Process measurements of a normal batch (NOC-2), taken by NIRS at the 6th, 60th, 120th and 180th minute. The inset shows the maximum difference between the data of the 6th minute and the other curves.

Among relevant investigations, Xu et al. studied NIRS applications in the ethanol precipitation process of *flos lonicerae japonicae* [8]. They built a partial least squared regression (PLSR) model to correlate the NIRS spectra with the high performance liquid chromatography (HPLC) data of chlorogenic acid content, and applied the simple interval calculation (SIC) theory for model updating among batches. Yang and coworkers investigated the on-line monitoring of the extraction process of *flos lonicerae japonicae* by using NIRS, and adopted synergy interval PLSR with genetic algorithm to predict the key quality parameters [9]. Vigni and Cocchi utilized NIRS and multivariate analysis to evaluate wheat flour doughs leavening, in which parallel factor analysis and n-way PLSR were adopted for the qualitative and quantitative modeling, respectively [10]. In addition, non-linear calibration models, such as kernel PLSR and artificial neural network, have been developed for the better processing of NIRS data [11].

In principle, once reliable models are constructed, PLSR and other predictive learning approaches reviewed above can give quantitative prediction of any critical quality attribute (CQA), the message from which is easy to interpret and helpful for the design of control strategy. Moreover, it is possible to further combine different descriptive and/or predictive algorithms to form more powerful PAT tools for process monitoring [12–16]. Nevertheless, it is worth noting that the information which can be directly obtained from the PAT tools is relatively preliminary, such as the instantaneous content of certain API at a given time step, while extracting knowledge from, e.g., the correlation between the measurements of various time steps, often requires more effort. Although the latter can also contain significant information about the process, it receives limited attention from the current application of PAT in HEP. This relates to the focus of the community on CQAs, which in HEP, can be several API monitored by NIRS. We can refer such situation as quality of product oriented monitoring, which is reasonable and is consistent with the QbD initiative. However, it is worth pointing out that in an industrial context, abnormalities during a process may not always lead to abnormal values in the CQAs. In practice, this may result from the inherent resistance of the process to the external disturbance, an example of which is presented later in this article. Nevertheless, it is still significant to pay attention to such "weak" abnormalities because they could introduce quality risk into the manufacturing and even imply, e.g., a process failure in the future. As such, in addition to the quality of product, we also need to pay attention to the quality of the process.

For HEP, the information about the process quality can be actually observed from the variation trend of the measurements from various time steps. This means that in principle, we may not need to take more measurements, but do need to put more effort into reading the measurement data. In this study, we take the extraction of *licorice* as an instance of HEP and utilize NIRS and ultra performance liquid chromatography (UPLC) for the measurements of the process states. In the mean time, classical PLSR is adopted for the information mining from the measurement data, with the help of some *a priori* knowledge about the extraction process. We further demonstrate that, albeit simple, such strategy works effectively for monitoring the extraction process, and provides physical and meaningful information on the quality of both the product and the process. The details of the method are presented in the next section, followed by the application results and relevant discussion.

2. Model and methods

2.1. Materials

Raw *licorice* root (batch number 1403008) was supplied by Qiyitang pharmaceutical Co., Ltd. (Hebei, China). Liquiritin (batch number 111610) standard substance was obtained from the