Chemometrics and Intelligent Laboratory Systems 128 (2013) 144–152

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Target-oriented overall process optimization (TOPO) for reducing variability in the quality of herbal medicine products

Bing Xu [a], Zhaozhou Lin [a], Zhisheng Wu [a], Xinyuan Shi [a,b,*], Yanjiang Qiao [a,b,*], Gan Luo [a]

[a] Research Center of TCM Information Engineering, Beijing University of Chinese Medicine, Beijing 100029, PR China
[b] Engineering Research Center of Key Technologies for Chinese Medicine Production and New Drug Development, Ministry of Education of People's Republic of China, Beijing 100029, PR China

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 23 July 2013
Accepted 16 August 2013
Available online 30 August 2013

Keywords:
Target-oriented overall process optimization (TOPO)
Quality variability
PLS regression
Bayesian approach
Probability trajectory
Herbal medicine

ABSTRACT

This paper presents a new strategy, target-oriented overall process optimization (TOPO), which can be used to assure the consistent quality in herbal medicine products. The methodology of TOPO includes four parts, target definition, data pretreatment, process modeling and overall process optimization. The Bayesian approach is integrated into the optimization step. The mechanism of TOPO involves optimizing multiple units of the production system step by step, giving each unit optimal operating conditions consistent with the quality target. The effects of TOPO were assessed using the descriptive statistics of the Bayesian posterior predictive distribution and the final target achievement. The probability trajectory was adjusted to monitor and optimize the production process. The proposed TOPO strategy was successfully applied to a seven-unit manufacturing process used to produce Lonicerae Japonicae extract. Results demonstrated that TOPO could keep the production process in line with the predefined target and reduce the variability of the final products.

1. Introduction

Herbal medicines and their derivatives have been used extensively for thousands of years in many Asian countries, such as China, Japan and Korea. For example, in China's National Essential Drug List (2012 Edition), 203 out of 520 total recommended drugs were herbal medicine preparations. In Europe and North America, herbal medicines have seen increasing use over the past few decades, largely in the form of dietary supplements, functional foods or health products. It is estimated that nearly 80% of the world's people still rely on herbal medicines for health related benefits [1]. According to the World Health Organization (WHO), the global market for herbal remedies and supplements was about U.S. $83 billion in 2008, and it continues to grow exponentially [2].

These trends lead to the imperative requirements of quality control of herbal medicine products, because the quality of herbal products is linked directly to their efficacy and safety [3]. Generally, quality control of herbal products involves identification of the starting material, details of the manufacturing process, and standards for the finished product. Currently, there are many technologies that can be used to control the quality of these products, including chemical profile methods, biological methods, on-line analytical tools, integrated evaluation approaches, etc. [4–7]. However, these methods mainly refer to the analytical aspects of quality control, and very few studies have evaluated the engineering aspects.

The major difficulties and challenges in the quality control of herbal medicines lie in the variability of the herbal material, the degree of which depends on factors such as the location of growing, the time of harvest, preprocessing methods and storage conditions [8–10]. Natural variability may be introduced into the manufacturing process, causing fluctuations between different batches [11,12]. Under these circumstances, conventional analytical techniques can identify the variations, but they cannot maintain quality consistence across herbal products. For this reason, there is an urgent need to address the problem of variability from the production point of view, because the quality of herbal products is affected by the manufacturing processes to a large extent [13].

Nowadays, many technological systems have been adapted from the chemical and pharmaceutical industries, and used to modernize the ways in which herbal products are processed. These techniques include solvent extraction, macroporous resin column chromatography, high-speed counter-current chromatography and various dosage preparation methods [14–18]. The overall manufacturing process of herbal medicine often consists of multiple processing units, which could also be called multistage batch process [19]. Through the serial processing stages, the desired quality is transformed from the starting materials to the final products step by step. Downstream units are influenced by upstream units. All parameters of the manufacturing process more or

* Corresponding authors at: School of Chinese Materia Medica, Beijing University of Chinese Medicine, South of Wangjing Middle Ring Road, Chaoyang District, Beijing City, 100102, PR China. Tel.: +86 10 84738621; fax: +86 10 84738661.
E-mail addresses: shixinyuan01@163.com (X. Shi), yjqiao@263.com (Y. Qiao).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.08.008

less affect the final product [20]. Although the manufacture of herbal products is now subjected to the Good Manufacturing Practice (GMP) standards, there is a lack of strategies to coordinate the relationships between different production units and to optimize the process parameters, in order to assure consistent product quality. Nevertheless, developing such strategies is challenging, because of the complex kinetics and thermodynamics of these processes, and of the unclear mechanisms by which active ingredients are transferred.

In order to tackle the abovementioned difficulties, systematic methodologies including multistage batch process modeling, monitoring, control and optimization, are needed to effectively control and improve the product quality during the multistage manufacturing process [21–23]. And some new concepts, e.g. plant wide optimization (PWO), could also be applied to facilitate optimal operation conditions consistent with the quality objectives of the large scale production system [24–26]. Based on these thoughts, a new strategy named target-oriented overall process optimization (TOPO) is brought forward to address the problem of variability in the concentrations of active ingredients in herbal products. The rest part of this article is organized as follows. First, a brief introduction is made about the mathematical foundations of TOPO. Then, technical details of the TOPO are illustrated. The method of expanding PLS modeling proposed by A. Pomerantsev et al. [27] will be employed in the TOPO strategy together with the Bayesian optimization technique. After that, effects of the proposed TOPO strategy were tested in a seven-unit manufacturing process used to produce the Lonicerae Japonicae extract. Finally, a summary of this paper is provided.

2. Overview of TOPO

The target-oriented overall process optimization (TOPO) strategy proposed in this study integrates the expanding PLS regression method and the Bayesian approach together. The word "expanding" here means a series of PLS models are built at the end of each stage, where the quality variables can also be predicted [27,28]. For a multistage manufacturing system with many process variables, it is time consuming and even impossible to optimize all combinations of variables. Therefore, with the help of established series of PLS models, optimization operation is designed to start from the second stage to the last stage. For a certain stage to be optimized, all the historical data, as well as the measurements from its previous stages are utilized. The goal of TOPO is to consistently provide optimal assurance for the product quality meeting defined specifications.

2.1. Mathematical fundamentals of TOPO

2.1.1. Partial least square regression

Partial least square (PLS) regression is a popular chemometric tool and is widely applied in industrial research, development and production. In the presence of historical production data which are often none-designed and not orthogonal, or contain noisy and collinear process variables, PLS method deserves the property to grasp hidden relationships between process variables and quality variables.

The main purpose of PLS regression is to build a linear model relating the independent data $\mathbf{X}$ (size $m \times n$, m is the number of observations and n is number of variables) with the response data $\mathbf{Y}$ (size $m \times q$, q is the number of responses):

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{1}$$

where $\mathbf{B}$ (size $n \times q$) is the matrix of regression coefficients; $\mathbf{E}$ is a noise term and has the same dimension with $\mathbf{Y}$. The basic assumption of PLS method is that there is a small number of latent variables (LVs) [29], which are linear combination of the original $\mathbf{X}$ variables, and can capture most of information in the calibration data for predicting the responses. These latent variables are also known as $\mathbf{X}$-scores, by which the PLS model can be written as follows:

$$\mathbf{Y} = \mathbf{TV} + \mathbf{E} \tag{2}$$

where $\mathbf{T}$ (size $m \times p$, and p corresponds to the number of latent variables) is the matrix of $\mathbf{X}$-scores; $\mathbf{V}$ (size $p \times q$) is the matrix of inner regression coefficients for $\mathbf{T}$. PLS method produces the $\mathbf{T}$ through a weighting matrix $\mathbf{W}$ and loading matrix $\mathbf{P}$:

$$\mathbf{T} = \mathbf{XW}\left(\mathbf{P}^{\mathrm{T}}\mathbf{W}\right)^{-1} \tag{3}$$

where $\mathbf{W}$ is a $n \times p$ matrix and is computed to maximize the covariance between the scores and responses. $\mathbf{P}$ is a $n \times p$ matrix. Two popular algorithms can be employed to compute the scores matrix, i.e. the NIPALS algorithm and SIMPLS algorithm [30,31]. Once $\mathbf{T}$ is obtained, the inner regression coefficients $\mathbf{V}$ in Eq. (2) is estimated by regressing $\mathbf{Y}$ on $\mathbf{T}$ via ordinary least square regression (OLS) procedures:

$$\widehat{V} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y}. \tag{4}$$

Given a new sample vector $\mathbf{x}$ (size $n \times 1$), the $\mathbf{x}$ is firstly projected onto the latent space, generating a score vector $\mathbf{t}$ (size $p \times 1$):

$$\mathbf{t} = \left(\mathbf{W}^{\mathrm{T}}\mathbf{P}\right)^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{x} \tag{5}$$

Then, the corresponding response $\widehat{y}$ could be predicted according to Eq. (2):

$$\widehat{y} = \mathbf{t}^{\mathrm{T}}\widehat{V}. \tag{6}$$

Prediction can also be made directly from original variables of the sample $\mathbf{x}$ according to Eq. (1), where $\mathbf{B}$ is estimated as follows:

$$\widehat{B} = \mathbf{W}\left(\mathbf{P}^{\mathrm{T}}\mathbf{W}\right)^{-1}\mathbf{V}. \tag{7}$$

The number of latent variables determines the complexity of the model. Therefore, the test set validation method and cross validation method (e.g. leave one out, LOO) are usually introduced to help select the optimal number of LVs and to test the predictive ability during the model construction [32]. Some chemometric indicators, such as root mean square error of calibration (RMSEC), root mean square error of cross validation (RMSECV), root mean square error of prediction (RMSEP), ratio of performance to deviation (RPD), predicted residual error sum square (PRESS) and bias, are often used to assess the performance of the established model [33,34]. For example, the PRESS index is calculated as:

$$\mathrm{PRESS} = \sum_{i=1}^{m}\left(\widehat{y}_i - y_i\right)^2. \tag{8}$$

The PRESS value shows the sum of squares of deviation between the predicted and the true property of the validation sample i during cross-validation, and it decreases as the LVs increase. When the PRESS value tends to be constant, the optimum number of LVs is obtained.

2.1.2. Bayesian approach

Under the framework of Bayes' theorem, the Bayesian inference combines the prior knowledge about the model parameters with information from measured data [35]. In process optimization, the Bayesian approach provides a natural way of making inference on future response $\widetilde{y}$ from its posterior predictive distribution. Using the classical