Chemometrics and Intelligent Laboratory Systems 140 (2015) 117–125

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

# A novel adaptive algorithm with near-infrared spectroscopy and its application in online gasoline blending processes

Kaixun He [a,b], Feng Qian [a,*], Hui Cheng [a,**], Wenli Du [a]

[a] Key Laboratory of Advanced Control and Optimization for Chemical Processes (East China University of Science and Technology), Ministry of Education, Shanghai 200237, China
[b] School of Information Science and Engineering, East China University of Science and Technology, 200237, China

ARTICLE INFO

*Article history:*
Received 9 August 2014
Received in revised form 11 November 2014
Accepted 15 November 2014
Available online 22 November 2014

*Keywords:*
Near-infrared spectroscopy
Gasoline blending
Adaptive model

ABSTRACT

Near-infrared (NIR) spectroscopy has been widely used to estimate product qualities. Although numerous studies on NIR modeling methods have been conducted, few papers have reported the online application of an NIR spectrometer in the gasoline blending process. This study presents a novel adaptive modeling method to establish an NIR model for the gasoline blending process. This method is based on the local learning and recursive modeling framework. Based on the framework, the proposed method can adjust the model structure from two aspects: (i) in sampling intervals, the model is updated with a local learning strategy, and the weights of the training samples can be gradually adjusted; and (ii) when new reference samples become available, the new data pairs are selected and added to the training data set based on an effective evaluation mechanism. The high performance of the proposed algorithm was demonstrated through a spectroscopic data set from a real gasoline blending process. The research octane number (RON), as the most important properties of gasoline, was estimated. Several modeling methods such as recursive partial least squares (RPLS), partial least squares (PLS), and locally weighted PLS were utilized for comparison. The results show that the proposed approach produce more accurate results than the traditional RPLS and locally weighted PLS algorithms.



## 1. Introduction

Gasoline is one of the most important profitable products of oil refineries and accounts for 60% to 70% of the total profit [1–4]. Gasoline production involves two main segments, namely, component oil storage and produce as well as gasoline blending. Generally, gasoline blending, which is the final stage before product delivery, is highly complex and has huge profit margins. Hence, it has been considered as a key process for the successful operation of most petroleum refineries [5,6]. In the traditional gasoline blending process, strong focus should be given on the final product quality and production efficiency. The research octane number (RON), motor octane number (MON), Reid vapor pressure (RVP), density, olefin and aromatics content, benzene content, sulfur content, fractional composition, and boiling points are determined via laboratory analyses. However, these procedures are slow (time delay), expensive, laborious, and less frequent (about one or two samples per day) [7]. Consequently, these analyses may delay the operator's decision making, which could affect the dynamic process [8,9], and advanced process controls cannot be well-applied for the optimization of this blending process.

In recent decades, the petroleum refining industry has increasingly faced challenges such as stringent environmental constraints, unsteady crude oil supply, changing gasoline specifications mandated by new government regulations, and low-inventory flexibility [2]. To effectively use resources, improve gasoline quality, and shorten the delivery time, most gasoline blending components are blended in online blend units instead of being stored in intermediate tanks [10]. Therefore, the online gasoline blending process has gained considerable attention from the academia and industry. In the blending process, different components are pumped to a blender from intermediate storage tanks or upstream devices. After passing through the blender, all the components should be evenly blended, and the product would be piped to the final storage tanks or be directly delivered.

More importantly, to reduce re-blending frequency and "quality giveaway," an optimal control system used to update the blending recipes and control the flow stream of blending components is necessary. When the optimal system is online, blending recipes are recursively updated depending on the feedback information from process analytical tool, particularly the NIR spectrometer [4]. Evidently, the recipe updating strategy and optimal control mechanism of online blending process heavily rely on online data acquisition and analysis system. In the gasoline blending process, the sulfur content of gasoline is analyzed using a total sulfur analyzer. The measurement of the other properties, such as RON, MON, and RVP, heavily rely on the online NIR spectrometer. Particularly, the measurement of octane number (RON and MON) is an essential factor in the successful optimized operation of online gasoline blending. A consistent octane giveaway per 0.1 octane number

* Corresponding author. Tel.: +86 21 64252056.
** Corresponding author.
*E-mail addresses:* fqian@ecust.edu.cn (F. Qian), huihyva@ecust.edu.cn (H. Cheng).

http://dx.doi.org/10.1016/j.chemolab.2014.11.006
0169-7439/

can cost a refinery several millions of dollars per year [5]. Therefore, in the present study, the online modeling method with NIR for RON was investigated.

Compared with the traditional laboratory analyses, NIR analysis can provide estimates more rapidly. Therefore, it has been widely applied in various fields [11–14]. Its efficiency for laboratory and industrial applications has been proven [15]. However, an NIR spectrometer is a kind of secondary instrument in which the qualitative and quantitative analysis are based on multivariate statistics. Particularly, the principal component regression (PCR) and partial least squares (PLS) have been widely accepted as useful techniques in building NIR models [16,17]. PCR and PLS provide transparent correlation models and unique solutions for a given set of training data. These techniques have gained popularity because of their statistical background, easily interpretable model, and effective data co-linearity [18]. For several systems that exhibit a nonlinear relationship with NIR spectroscopy, nonlinear regression methods are desired. In response to this challenge, artificial neural networks [19–21], support vector machines [22], least-squares support vector machines, nonlinear part least squares [13], and Gaussian process regression [16] have been adopted. All these abovementioned linear or nonlinear algorithms are static modeling methods. These regression models are specifically trained using offline historical operating data and then used for prediction.

However, the static calibration models established using a historical data set might not be representative of the future process conditions because their estimation performance would deteriorate. In the online gasoline blending process, the NIR model performance may be jeopardized by the variability of the blending recipe, instrument aging or repair, and fluctuations in physical conditions (such as temperature and flow turbulence). In addition, the maintenance of the NIR model is crucial but laborious because the training data and modeling parameters should be carefully selected [23]. From a practical viewpoint, the long-term application of an NIR analysis system mainly relies on the maintenance and update mechanism of the NIR calibration models. To overcome such problems and to effectively update the NIR models, several recursive and adaptive modeling approaches have been developed [24]. These algorithms can be roughly classified into the following categories [25]: (1) recursive algorithms that expand the training data set by adding each new available sample and recursively updating the model coefficients and (2) just-in-time (JIT) modeling methods that recalculate the weight of each training sample or select local training data points based on query samples.

One of the most widely used methods for the first category is the recursive partial least-squares (RPLS). In this algorithm, the calibration model is updated when the new sample becomes available. Therefore, this approach can capture the new characteristics reflected by the new added data point without delay. The model update mechanism of RPLS substantially relies on the new available samples, and the model structure would change as long as the new samples are collected. However, if this new added sample was collected within a narrow range for a particular period, the model will excessively adapt. As a result, the calibration model will not function well in the sufficiently wide range of operating conditions. In addition, when an abrupt change in the process is observed in the sampling intervals, these changes are difficult to be captured by recursive methods, and the predictive result becomes questionable [4]. To overcome such problems, the JIT model was proposed. In this model, the model structure is not updated until an estimated value is requested. The main characteristic of the JIT model is that all of the local models are established during sample intervals, and its accuracy is highly dependent on the definition of similarity distance. In this approach, the new available sample's contribution to the calibration model may be significantly limited if the definition of the similarity is not appropriate.

As mentioned in the preceding section, the main characteristic of a significant adaptive model roughly includes the following two points: (1) in sampling intervals, the calibration model should be effectively updated, such that it can cope with abrupt changes during the process; and (2) when new samples become available, an effective mechanism should be present to fully use the new samples to update the calibration model. The newly available samples cannot be added to the training data set and directly used to update the calibration model. Therefore, a criterion to determine the representativeness of the newly added sample is desired.

To address the aforementioned problems in this study, a novel method for building adaptive NIR model based on RPLS and JIT modeling algorithms is proposed. Through this method, the model structure (referring to the weights of every training sample) can be gradually updated in the sampling intervals. By contrast, if a new data point becomes available, a weight index of the new sample is initially calculated according to the coming query sample, and the model is then updated with the weighted new sample. Based on the novel updating strategy, the proposed approach can obtain advantages of both JIT and RPLS. In addition, the approach guarantees that the model will not excessively adapt if the new sample is not appropriate. As a result, it can effectively update statistical models and achieve high-estimation performance. The estimation performance of the proposed algorithm is demonstrated for an NIR data set from a real gasoline blending process in China. The predicted results are compared with RPLS, PLS, LW-PLS, and several other nonlinear algorithms.

This article is organized as follows: in Section 2, the PLS regression and JIT modeling algorithms are briefly introduced; Section 3 details the theory of the proposed modeling algorithm; in Section 4, the case study of gasoline blending is presented; in Section 5, the benchmark are discussed; and in Section 6, the conclusions and future directions are outlined at the end of the article.

## 2. PLS regression and JIT strategy

Statistical regression algorithm is the heart of the NIR model. To establish a significant calibration model via NIR spectroscopy, considerable efforts have been directed toward the following problems [14]: (1) data preprocessing and wavelength selection have to be carefully performed in advance; (2) the samples used to build the initial training set should be selected and effectively organized; (3) an appropriate regression method is desired according to the industry process; and (4) the model updating strategy should be effective.

In this study, the proposed approach aims to improve the performance of the model updating mechanism; however, the other aspects were not discussed. In this section, the conventional adaptive methods (RPLS and JIT) for NIR spectroscopy analysis are briefly explained to derive the proposed modeling algorithm because the proposed novel algorithm are based on the RPLS and JIT methods.

### 2.1. RPLS

The recursive algorithms generally use the current model in the form of a covariance matrix [21]. When new data become available, a forgetting factor is usually used to down-weight the previous model. RPLS is extended from PLS, and it is a special case of such recursive algorithms.

PLS, which has been developed from PCR and canonical correlation analysis, was originally introduced by Wold [26]. PLS has been used in various industrial processes because it can effectively eliminate the effects of co-linearity. Given a pair of input and output matrices $\mathbf{X}$ and $\mathbf{Y}$, the $i$th sample is denoted as $\{\mathbf{x}_i = [x_1, x_2,\ldots, x_m]^T, \mathbf{y}_i = [y_1, y_2,\ldots, y_l]^T\}$, where $m$ and $l$ represent the dimensions of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

The goal of PLS is to project the scaled and mean-centered input data $\mathbf{X} \in \mathbb{R}^{n \times m}$ and output data $\mathbf{Y} \in \mathbb{R}^{n \times l}$ to separate the latent variables according to the following equations [27]:

$$\mathbf{X}=\mathbf{t}_1\mathbf{p}_1{}^{\mathbf{T}}+\mathbf{t}_2\mathbf{p}_2{}^{\mathbf{T}}+\ldots\mathbf{t}_h\mathbf{p}_h{}^{\mathbf{T}}+\mathbf{E}_h \quad (1)$$