Improved Independent Component Regression Modeling

Chunhui Zhao, Furong Gao, Tao Liu, and Fuli Wang

Abstract— The conventional independent component regression (ICR), as an exclusive two-step implementation algorithm, has the risk similar to principal component regression (PCR). That is, the extracted independent components (ICs) are not guaranteed to be informative with respect to quality prediction and interpretation. Moreover, it inherits some inconveniences of conventional ICA. In this paper, first, the drawbacks of original ICR are analyzed. Then a modified ICR (M-ICR) modeling algorithm is developed. To enhance the causal relationship between the extracted ICs and quality variables, a dual-objective optimization solution is constructed in the first-step feature extraction modeling. It simultaneously considers two-fold statistical requirements, the independence and quality-correlation. Moreover, their different roles in calibration modeling can be quantitatively evaluated by flexibly adjusting the sub-optimization objective weights. The practicability and performance of M-ICR are illustrated and discussed in simulation experiment.

I. INTRODUCTION

As one important issue of statistical analysis, multivariate calibration methods[1-4] have been used to well-establish a quantitative relationship between process measurement (X) and quality property (Y). Accurate qualitative and quantitative calibration analysis may help avoiding heavy and costly chemical measurements. For practical applications, calibration modeling and analysis can be accomplished with familiar conventional chemometrical techniques, such as multiple linear regression (MLR), principal component regression (PCR), partial least squares (PLS), and so on. Among them, latent variable (LV) based methods in the form of PCR and PLS have a dominating role, in which, usually fewer uncorrelated LVs are first defined to comprehensively represent the important aspect of correlated observed variables and then used to build a quantitative relationship relevant to the concerned quality properties.

Manuscript received March 6, 2009. This work was supported in part by Hong Kong Research Grant Council under project number (613107) and National Natural Science Foundation of China (No. 60774068).

Chunhui Zhao, Furong Gao, and Tao Liu are with the Department of Chemical and Biomolecular Engineering, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong (phone: 852-2358-7139; fax: 852-2358-0054; e-mail addresses: huihui.zhao@gmail.com (C. Zhao); kefgao@ust.hk (F. Gao); liurouter@ieee.org (T. Liu)).

Fuli Wang is with College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning Province, P. R. China (e-mail addresses: flwang@mail.neu.edu.cn (F. Wang)).

However, neither the PLS nor PCR algorithm will exactly recover a true underlying linear LV model in general.[5] Although they may yield good predictions and sometimes offer interesting insights about the key underlying players, their chemical interpretations are very difficult. Therefore, the calibration analysis is not satisfactory enough as reported in many practical applications. Great efforts have been and are being made focusing on extracting latent features with more interesting statistical characteristics to serve as descriptors and thus to improve the calibration modeling performance.

In recent years, independent component analysis (ICA)[6] has drawn considerable attention and found potential applications as a further development of principal component analysis (PCA). It attempts to recover statistical independent signal sources given only observations that are assumed to be linear mixtures of the original signal sources when the Gaussian distribution assumption is invalid. In probability theory, independence is much stronger than orthogonality. Those extracted ICs may thus provide more informative statistical explanations and more chemically interpretable characteristics and better reflect the intrinsic properties of measurement data, which shows its superiority over PCA-based methods. Motivated by the appealing role of ICA, independent component regression (ICR)[5,7,8] was naturally developed as a combination of a first-step ICA implementation on process measurement and a second-step quantitative regression modeling. Gustafsson[5] has made some meaningful investigations to reveal the superiority of ICR, which reported that compared with PCR and PLS, ICR can retrieve statistically independent and, therefore, chemically interpretable LVs in multivariate regression. However, since conventional ICR is implemented in such a two-step statistical analysis that the ICA feature extraction and calibration modeling are exclusively performed, it has the risk similar to principal component regression (PCR). That is, those ICs decomposed from the observed mixed signals are not ensured desirable and close causal relationship with quality properties. Because process observations often contain major sources of variations that are of little or no predictive feature, it may be observed that the extracted first several ICs are capturing most of process variations but not necessarily explaining the quality properties. It is not difficult to imagine that some useful (predictive and interpretative) information may end up in the discarded ICs whereas some quality-uninformative process systematic variations may be introduced into the retained ICs for the subsequent calibration analysis. To make up the deficiency, generally more ICs have

to be employed for comprehensive quality description. It, however, increases the model complexity although increasing descriptor information in a regression model will improve the fitting to the training reference data. In particular, it will often cause a substantial reduction in the generalized predictive ability of the model concerning the new sample. Despite the reported progress in ICR algorithms over the past few years, the aforementioned highlight that a number of issues still remain.

I. METHODOLOGY

The performance of ICR greatly depends upon how well the features are prepared in the LV modeling step. As analyzed previously, in the conventional ICR method, the LV model is extracted using ICA exclusively based on the process measurement ($\mathbf{X}$) to capture the major process systematic variations. This estimation takes no account of the roles of LVs in quality prediction and thus is not optimized for quality prediction. It should be recognized that some ICs may be undesirably quality-uninformative and thus make no contribution to quality description although they are strongly independent with each other. If those ICs are directly used to interpret quality variations, it is not difficult to imagine that more ICs may be required in the regression model to capture quality property, which may result in both complex model structure and overfitting defect. Therefore, it is natural that feature extraction should be paid more attention if one desires further improvement of ICR calibration analysis.

The proposed alternative algorithm described below, a modified ICR (M-ICR), tries to make the extraction of ICs more direct and comprehensible for regression modeling. First, the mathematical basis of the modified ICA feature extraction algorithm is formulated as a dual-objective optimization problem, in which, the synthetical objective function is designed as a weighted combination of quality-related characteristics and statistical independency. Thus, the extraction of LVs can be feasibly adjusted and then the desired regressors can be obtained for quality interpretation and prediction. Then, the M-ICR modeling procedure is accordingly developed to derive the regression relationship.

A. Description of dual-objective optimization solution

By choosing different optimization criteria and objective functions, different decomposing algorithms are available to derive different latent features. To improve the extracted ICs for calibration modeling, the latent factors derived from process measurement are expected to simultaneously satisfying the following two objectives:

(1) They should be as independent of each other as possible, i.e., the maximization the non-Gaussianity, which are obtained at certain optima of $E\{G(\mathbf{w}^T \mathbf{z})\}$. According to Hyvärinen and Oja,[6] the following choices of $G(\cdot)$ have proved very useful:

$$G_1(u) = \frac{1}{a_1} \log \cosh(a_1 u) \quad G_2(u) = -\exp\left(-\frac{u^2}{2}\right) \quad (1)$$

where constant a_1 satisfies $1 \leq a_1 \leq 2$. $G_1(u)$ is a good general-purpose contrast function and will be used in our M-ICA iteration algorithm.

(2) They should remain as close as possible relationship with quality vector, i.e., maximization of the absolute value of covariance between them: $\max(E((\mathbf{w}^T \mathbf{z})y))^2$ (where, y is one sample of the normalized quality value). Considering that both $\mathbf{w}^T \mathbf{z}$ and y have unit variance, the covariance actually is equal to the correlation coefficient. Moreover, in the case of multivariate quality variables, synthetically considering the multiple correlations, the optimization index can thus be figured out as $\max \frac{1}{J_y} (E(\mathbf{w}^T \mathbf{z} \mathbf{y})^T E(\mathbf{w}^T \mathbf{z} \mathbf{y}))$ (where, $\mathbf{y}(J_y \times 1)$ is the normalized J_y -dimensional quality variable).

Considering that both independency and quality-related statistical characteristics are desired, a synthetical optimization objective should be designed. Note that in objective (1), the maximization of non-Gaussianity boils down to maximizing or minimizing $E\{G(\mathbf{w}^T \mathbf{z})\}$, where the type of searched extrema (maximization or minimization) depends on the sign of $E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(v)\}$ (v is a standardized Gaussian variable). The positive corresponds to the super-Gaussian and the negative to the sub-Gaussian. Therefore, the two sub-optimization objectives couldn't be simply added. Instead of $E\{G(\mathbf{w}^T \mathbf{z})\}$, the mathematical form, $\max(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(v)\})^2$, is adopted, which, thus, does not need to distinguish the super- or sub- Gaussian characteristics.

Making a tradeoff between the above two maximization objectives, an integrated dual-objective optimization problem can be finally designed as:

$$\max \begin{pmatrix} \alpha \cdot \frac{1}{J_y} (E(\mathbf{w}^T \mathbf{z} \mathbf{y})^T E(\mathbf{w}^T \mathbf{z} \mathbf{y})) \\ + \beta \cdot (E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(v)\})^2 \end{pmatrix} \quad (2)$$

$$\text{subject to the constraint: } \mathbf{w}^T \mathbf{w} = 1 \quad (3)$$

where, α and β are respectively the weight coefficients attached to each sub-optimization objective. Here they are defined satisfying $\alpha + \beta = 1$ for simplicity.

The modified ICA estimator is in pursuit of both non-Gaussianity and close relationship with quality. On the basis of the original fast fixed-point ICA algorithm (FastICA)[6] by Hyvärinen et al., here the developed optimization solution, termed modified FastICA algorithm (MFASTICA), is formulated to achieve the above objective as follows using an approximative Newton iteration scheme.

Using Lagrange operator, the initial objective function can be expressed to be non-constraint extremum problem:

$$F(\mathbf{w}, \lambda) = \alpha \cdot \frac{1}{J_y} \left(E(\mathbf{w}^T \mathbf{z} \mathbf{y})^T E(\mathbf{w}^T \mathbf{z} \mathbf{y}) \right) + \beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right)^2 + \lambda(\mathbf{w}^T \mathbf{w} - 1) \quad (4)$$

where λ is a constant scalar.

Respectively calculate the derivative of $F(\mathbf{w}, \lambda)$ to $\mathbf{w}$ and λ and set both of them to be zero, the following equations can obtain:

$$\nabla F_{\mathbf{w}} = \frac{\partial F}{\partial \mathbf{w}} = 2\alpha \cdot \frac{1}{J_y} E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} + 2\lambda \mathbf{w} = 0 \quad (5)$$

$$\nabla F_{\lambda} = \mathbf{w}^T \mathbf{w} - 1 = 0 \quad (6)$$

where, g is the first-order derivative of G .

Multiply both sides of Eq. 5 by $\mathbf{w}^T$:

$$2\alpha \cdot \frac{1}{J_y} \mathbf{w}^T E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) \mathbf{w}^T E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} + 2\lambda \mathbf{w}^T \mathbf{w} = 0 \quad (7)$$

According to the relationship shown in Eq. 7, λ can be specifically calculated as:

$$\lambda = -\alpha \cdot \frac{1}{J_y} \mathbf{w}^T E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) - \beta \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) \mathbf{w}^T E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \quad (8)$$

Based on Kahn-Tucker conditions[9], it is known that the optima of objective function shown in Eq. 2 under the constraint $\mathbf{w}^T \mathbf{w} = 1$ may be achieved at points shown in Eq. 5. Here, to improve the convergence speed, Newton's method is adopted to solve this problem, which makes use of the second-order derivative of objective function and thus takes the changing trend of grade into account.

First, denoting the function on the left-hand side of Eq. 5 by Ψ , its Jacobian matrix, $J\Psi(\mathbf{w})$, can be obtain by:

$$J\Psi(\mathbf{w}) = 2\alpha \cdot \frac{1}{J_y} E(\mathbf{z} \mathbf{y}^T) E(\mathbf{y} \mathbf{z}^T) + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) E\{\mathbf{z} \mathbf{z}^T g'(\mathbf{w}^T \mathbf{z})\} + 2\beta \cdot E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \left(E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \right)^T + 2\lambda \mathbf{I} \quad (9)$$

Substituting Eq. 8 into Eq. 9, the following expression of Jacobian matrix will be derived:

$$J\Psi(\mathbf{w}) = 2\alpha \cdot \frac{1}{J_y} E(\mathbf{z} \mathbf{y}^T) E(\mathbf{y} \mathbf{z}^T) + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) E\{\mathbf{z} \mathbf{z}^T g'(\mathbf{w}^T \mathbf{z})\} + 2\beta \cdot E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \left(E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \right)^T - \left[\begin{array}{c} 2\alpha \cdot \frac{1}{J_y} \mathbf{w}^T E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) \\ + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) \mathbf{w}^T E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \end{array} \right] \mathbf{I} \quad (10)$$

To simplify the inversion of this matrix, the second term was approximated in the original FastICA algorithm since the data has been whitened. Here, it is simply inherited to get the following approximation:

$$E\{\mathbf{z} \mathbf{z}^T g'(\mathbf{w}^T \mathbf{z})\} \approx E\{\mathbf{z} \mathbf{z}^T\} E\{g'(\mathbf{w}^T \mathbf{z})\} = E\{g'(\mathbf{w}^T \mathbf{z})\} \mathbf{I} \quad (11)$$

where, g' is the second-order derivative of G .

So the Jacobian matrix can be approximately expressed as a simple symmetrical form:

$$J\Psi(\mathbf{w}) = 2\alpha \cdot \frac{1}{J_y} E(\mathbf{z} \mathbf{y}^T) E(\mathbf{y} \mathbf{z}^T) + 2\beta \cdot E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \left(E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \right)^T - c \cdot \mathbf{I} \quad (12)$$

where, c is a constant, which is defined to denote the following expression:

$$c = 2\alpha \cdot \frac{1}{J_y} \mathbf{w}^T E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) \mathbf{w}^T E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} - 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) E\{g'(\mathbf{w}^T \mathbf{z})\} \quad (13)$$

Moreover, substituting Eq. 8 into Eq. 5, the derivative of $F(\mathbf{w}, \lambda)$ to $\mathbf{w}$, $\nabla F_{\mathbf{w}}$, can then be expressed as:

$$\nabla F_{\mathbf{w}} = 2\alpha \cdot \frac{1}{J_y} E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) + 2\beta \cdot \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} - 2 \left[\begin{array}{c} \alpha \cdot \frac{1}{J_y} \mathbf{w}^T E(\mathbf{z} \mathbf{y}^T) E(\mathbf{w}^T \mathbf{z} \mathbf{y}) \\ + \beta \left(E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\} \right) \mathbf{w}^T E\{\mathbf{z} g(\mathbf{w}^T \mathbf{z})\} \end{array} \right] \mathbf{w} \quad (14)$$

The Newton iteration direction $\mathbf{d}_n$ is actually derived by solving the following relationship equation:

$$J\Psi(\mathbf{w}) \cdot \mathbf{d}_n = -\nabla F_{\mathbf{w}} \quad (15)$$

Therefore, the solution of Newton direction greatly depends upon whether the Jacobian matrix is invertible. When it is not satisfied, some modified strategy should be adopted to modify it and make it invertible. An approximative Newton iteration direction can thus be calculated.

Moreover, considering that Newton method can't keep its convergence ability when the searching region is far from the optima of objective function, one-dimensional searching algorithm can be introduced after obtaining the Newton iteration direction $\mathbf{d}_n$. It is used to define the searching step γ , which is carried out by solving $\max F(\mathbf{w} + \gamma \cdot \mathbf{d}_n)$, where function $F(\cdot)$ is shown in Eq. 4. Adopting one-dimensional search will make the method have the global convergence ability. If $\gamma = 1$, it converges to the conventional Newton iteration method.

Finally, $\mathbf{w}$ is derived based on the following approximative Newton iteration:

$$\mathbf{w} \leftarrow \mathbf{w} + \gamma \cdot \mathbf{d}_n \quad (16)$$

$$\text{i.e., } \mathbf{w} \leftarrow \mathbf{w} - \gamma \cdot (J\Psi(\mathbf{w})_m)^{-1} \nabla F_{\mathbf{w}}$$

where, $J\Psi(\mathbf{w})_m$ is the modified invertible Jacobian matrix and ∇F_w is the first-order derivative of $F(\mathbf{w}, \lambda)$ to $\mathbf{w}$ as shown in Eq. 14.

B. M-ICR calibration modeling

In this subsection, the M-ICR modeling procedure will be comprehensively described on the basis of the designed MFastICA algorithm to derive the quantitative regression relationship between the M-ICs and quality variables. For any optimization problem, its solution depends on two aspects: the initial setting and the searching algorithm, which respectively determine the starting point and the searching trace. Therefore, to achieve the desired objectives shown in Eq. 2, both aspects should be paid attention to.

First of all, those LVs from PLS should be good initial estimates of ICs from the particular viewpoint of quality-concerned calibration analysis:

$$\mathbf{t} = \mathbf{R}^T \mathbf{x} \quad (17)$$

where, $\mathbf{t}(d \times 1)$ is the retained d -dimensional latent factor vector by decomposing $\mathbf{x}(J \times 1)$ using PLS projection pursuit.

$\mathbf{R}(J \times d)$ is PLS weights matrix. $E\{\mathbf{t}\mathbf{t}^T\} = \mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_d\}$, in which, $\lambda_i (i = 1, 2, \dots, d)$ is the variance of retained PLS LVs. If some LVs have variance closely approximating zero, they can be removed by excluding their corresponding weights vectors in $\mathbf{R}$ to avoid the singular problem.

Then Eq. 17 can be further normalized as:

$$\mathbf{z} = \mathbf{\Lambda}^{-1/2} \mathbf{t} = \mathbf{\Lambda}^{-1/2} \mathbf{R}^T \mathbf{x} = \mathbf{Q} \mathbf{x} \quad (18)$$

where $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$ ($\mathbf{I}(d \times d)$ is identity matrix), and $\mathbf{Q} = \mathbf{\Lambda}^{-1/2} \mathbf{R}^T = \mathbf{\Lambda}^{-1/2} (\mathbf{W}_{\text{pls}}^T \mathbf{P}_{\text{pls}} \mathbf{W}_{\text{pls}})^{-1}$ (where $\mathbf{W}_{\text{pls}}$ is PLS weights matrix and $\mathbf{P}_{\text{pls}}$ is PLS loading matrix) is the whitening matrix. The scaled latent scores $\mathbf{z}$ are regarded as the preprocessed data of process observations after centering and whitening. Note that despite its scaled variance, $\mathbf{z}$ have been ordered according to the covariance information between $\mathbf{t}$ and quality. In particular, they have been removed statistical dependency up to second order. Therefore they can be a good initialization of further ICA iterative procedure. Starting from them, a consistent solution can be obtained. Correspondingly, the initialized demixing matrix is set as:

$$\mathbf{W}(n \times d) = [\mathbf{I}_n : \mathbf{0}] \quad (19)$$

where, n denoted the retained number of M-ICs. $\mathbf{I}_n$ is the n -dimensional identity matrix and $\mathbf{0}(n \times (d - n))$ is zero matrix.

Denoting $\mathbf{w}_i$ as the i th row of $\mathbf{W}$, the modified feature extraction tries to obtain the i th latent component with maximized synthetical measure of non-Gaussianity and quality-related correlation. The detailed procedure is given below:

- Determine the number of independent components n , set the counter $p = 1$.
- Set $\mathbf{w}_p$, the p th row of $\mathbf{W}$, as the initial demixing

vector.

- Let $\mathbf{w} \leftarrow \mathbf{w} - \gamma \cdot (J\Psi(\mathbf{w})_m)^{-1} \nabla F_w$.

This step is an approximative Newton iteration procedure in pursuit of the maximization of dual-objective optimization problem shown in Eq. 2.

- Orthogonalization and Normalization:

$$\mathbf{w}_p = \mathbf{w}_p - \sum_{j=1}^{p-1} \mathbf{w}_p^T \mathbf{w}_j \mathbf{w}_j; \quad \mathbf{w}_p = \frac{\mathbf{w}_p}{\|\mathbf{w}_p\|}.$$

- If $\mathbf{w}_p$ is not converged (here the convergence means that the dot-product of the old and new values of $\mathbf{w}$ is almost equal to 1), go back to step (c).
- Set $p = p + 1$. If $p \leq n$, return to step (b) until all desired ICs are retained and output the updated demixing matrix $\mathbf{W}(n \times d)$.

Once demixing matrix is figured out, the ICs from the initial process observations can be derived using the following expression:

$$\hat{\mathbf{s}} = \mathbf{W} \mathbf{z} = \mathbf{W} \mathbf{Q} \mathbf{x} = \mathbf{W} \mathbf{\Lambda}^{-1/2} \mathbf{R}^T \mathbf{x} = \mathbf{W}_x \mathbf{x} \quad (20)$$

In the general formulation, the first-step feature extraction based on MFastICA can be considered a variant of projection pursuit. It is developed in statistics for finding more “interesting” projection directions, in which the non-Gaussianity and quality-related measures are both called projection pursuit “indices”. These projection directions are derived based on an approximative Newton iteration algorithm.

After obtaining the desired M-ICs, regression modeling between them and quality variables can then be performed using simple least-squares algebra:

$$\begin{aligned} \mathbf{\Theta} &= (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{Y} \\ \hat{\mathbf{Y}} &= \mathbf{S} \mathbf{\Theta} \end{aligned} \quad (21)$$

where, ICR readily solves the collinearity problem of typical MLR calculation by guaranteeing an invertible matrix $\mathbf{S}^T \mathbf{S}$ because of the mutual orthonormality of the ICs. Actually, $\mathbf{S}^T \mathbf{S}$ is a diagonal matrix with identical diagonal elements.

It is possible to convert the regression relationship $\mathbf{\Theta}$ to $\mathbf{\Theta}_{x \rightarrow y}$, which is directly related to process measurement data $\mathbf{X}(N \times J)$:

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{X} \mathbf{W}_x^T \mathbf{\Theta} = \mathbf{X} \mathbf{\Theta}_{x \rightarrow y} \\ \mathbf{\Theta}_{x \rightarrow y} &= \mathbf{W}_x^T \mathbf{\Theta} \end{aligned} \quad (22)$$

Since those M-ICs are extracted taking the quality-relevant information into consideration to some extent, therefore, it is conceivable to get more appealing performance of quality interpretation and prediction.

II. SIMULATIONS AND DISCUSSIONS

To illustrate the performance of the modified ICR algorithm, it is applied to a numerical example system similar to that used by Lee et al.[10] Consider three source variables with the following distribution:

$$s_1(k) = 2\cos(0.08k)\sin(0.06k)$$

$$s_2(k) = \text{sign}[\sin(0.3k) + 3\cos(0.1k)] \quad (23)$$

$$s_3(k) = \text{uniformly distributed noises in the range } [-1, 1]$$

On the basis of the three source signals, process data and quality data are respectively generated by different mixing relationship: $\mathbf{x}^T = \mathbf{s}^T \mathbf{A}$ and $\mathbf{y}^T = \mathbf{s}^T \mathbf{M}$, in which, the mixing matrices $\mathbf{A}$ and $\mathbf{M}$ are defined as follows:

$$\mathbf{A} = \begin{bmatrix} 0.86 & -0.55 & 0.17 & -0.33 & 0.89 \\ 0.79 & 0.65 & 0.32 & 0.12 & -0.97 \\ 0.67 & 0.46 & -0.28 & 0.27 & -0.74 \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} 2 \\ 0 \\ 7 \end{bmatrix} \quad (24)$$

From the above mixing formulation, it can be seen that the mixing coefficient corresponding to the second IC is zero, which means that the quality attribute has no relationship with the second source signal and the variation of the second IC will not contribute to quality prediction. 1000 samples of mixture are generated for modeling building and random noises with variance 0.02 are added to the data. The source signals and the mixture data are respectively shown in Figs. 1 (a)~(c).

(a)

(b)

(c)

Figure 1 (a) Source data (b) Mixed process data (c) Quality data

The feature extraction results using the conventional FastICA, modified ICA[10] and the proposed method (M-ICR) are respectively shown in Figs. 2 (a)~(c). From them, the ICA solutions shown in Figs. 2 (a) and (b) reveal that both FastICA and modified ICA may recover original sources existing in process data, which, however, just provide an independent representation and reconstruction of the process data with no consideration of their effects on quality prediction and interpretation. For example, in Fig. 2 (b), the first three ICs will be chosen as dominant ICs according to the work of Lee et al. However, it is known that the second IC has no relationship with quality, whose participation in regression modeling will introduce undesirable variation and deteriorate the causal relationship between process data and quality. On the other hand, since those extracted ICs may not close relate with quality information, more ICs are needed to describe and fit quality variations, which, however, may suffer from the risk of introducing process systematic noises and complicating model structure. In conclusion, the prediction performance with the wrongly selected dominant ICs can be poor. Comparatively, using the proposed method, only two ICs which really and significantly contribute to quality are identified. Therefore, it gives an opportunity to analyze and pay attention to those sources which are really responsible for quality interpretation. In this way, it directly reduces the model structure by decreasing the number of latent variables and makes the interpretation easier.

(a)

(b)

(c)

Figure 2 (a) Original ICA solutions (b) Modified ICA solutions (c) Proposed M-ICR solutions

Since quality-related criterion is accommodated in the feature extraction, M-ICs are sifted out to be more specific for calibration analysis and thus provide possibility for the improvement of quality prediction and analysis. Some quality prediction results are illustratively shown in Fig. 3. The MSE of quality prediction is 0.5727, which is an acceptable value.

Figure 3 Quality prediction result

III. CONCLUSION

In this paper, an improved ICR (M-ICR) modeling algorithm is formulated. By integrating the statistical independency and quality-related requirement, it, thus, results

in more appealing ICs with enhanced causal relationship with qualities. Meanwhile, it indicates parsimonious model structure by excluding those quality-irrelevant ICs in regression analysis. Simulation examples have shown the potential of the proposed method for the improvement of multivariate calibration modeling and analysis. Considerable further research is thus suggested.

REFERENCES

- [1] H. Martens, T. Naes, Multivariate calibration (2nd edn), Wiley: Chichester, 1994.
- [2] R. G. Brereton, Introduction to multivariate calibration in analytical chemistry, *Analyst*, 2000;125:2125-2154.
- [3] D. G. Kleinbaum, L. L. Kupper, K. E. Muller, A. Nizam, Applied Regression Analysis and Other Multivariable Methods(3rd edn), China Machine Press: Beijing, 2003.
- [4] S. Jönsson, L. A. Eriksson, B. van Bavel, Multivariate characterisation and quantitative structure-property relationship modeling of nitroaromatic compounds, *Anal. Chim. Acta.* 2008;621:155-162.
- [5] M. G. Gustafsson, Independent component analysis yields chemically interpretable latent variables in multivariate regression, *J. Chem. Inf. Model.* 2005;45:1244-1255.
- [6] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Networks* 2000;13:411-430.
- [7] J. Chen, X. Z. Wang, A new approach to near-infrared spectral data analysis using independent component analysis, *J. Chem. Inf. Comput. Sci.* 2001;41:992-1001.
- [8] H. Kaneko, M. Arakawa, K. Funatsu, Development of a new regression analysis method using independent component analysis, *J. Chem. Inf. Model.* 2008;48:534-541.
- [9] D. G. Luenberger, Optimization by vector space methods, Wiley: New York, 1969.
- [10] J.-M. Lee, S. J. Qin, I.-B. Lee, Fault detection and diagnosis based on modified independent component analysis, *AIChE J.* 2006;52:3501-3514.