An Improved Independent Component Regression Modeling and Quantitative Calibration Procedure

Chunhui Zhao, Furong Gao, and Fuli Wang
Dept. of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

DOI 10.1002/aic.12079
Published online November 2, 2009 in Wiley InterScience (www.interscience.wiley.com).

An improved independent component regression (M-ICR) algorithm is proposed by constructing joint latent variable (LV) based regressors, and a quantitative statistical analysis procedure is designed using a bootstrap technique for model validation and performance evaluation. First, the drawbacks of the conventional regression modeling algorithms are analyzed. Then the proposed M-ICR algorithm is formulated for regressor design. It constructs a dual-objective optimization criterion function, simultaneously incorporating quality-relevance and independence into the feature extraction procedure. This ties together the ideas of partial-least squares (PLS), and independent component regression (ICR) under the same mathematical umbrella. By adjusting the controllable suboptimization objective weights, it adds insight into the different roles of quality-relevant and independent characteristics in calibration modeling, and, thus, provides possibilities to combine the advantages of PLS and ICR. Furthermore, a quantitative statistical analysis procedure based on a bootstrapping technique is designed to identify the effects of LVs, determine a better model rank and overcome ill-conditioning caused by model over-parameterization. A confidence interval on quality prediction is also approximated. The performance of the proposed method is demonstrated using both numerical and real world data. © 2009 American Institute of Chemical Engineers AIChE J, 56: 1519–1535, 2010

Keywords: independent component regression, quantitative calibration analysis, dual-objective optimization, statistical independence, quality-related characteristics

Introduction

Over the last few years, data-based statistical modeling and analysis methods have drawn much attention and found successful applications in practice.[1–16] They are an attractive option for handling various problems in many fields of chemical engineering that are ''data rich but information poor''. As one important area of statistical analysis, multivariate calibration,[9–16] has been widely used to establish quantitative relationships between process measurements (**X**), and quality properties (**Y**). Accurate qualitative and quantitative calibration analysis may help avoid cumbersome and costly chemical measurements. In practice, calibration modeling and analysis can often be accomplished with familiar, conventional statistical techniques such as multiple linear regressions (MLR), principal component regression (PCR), PLS, and so on. Among them, LV-based methods in the form of PCR and PLS have a dominating role, in which, correlation patterns among the variables can be modeled and underlying features are used to build and explain a quantitative relationship relevant to the quality properties concerned.

Correspondence concerning this article should be addressed to F. Gao at keczhao@ust.hk
Current Address of Fuli Wang: College of Information, Northeastern University, Shenyang, Liaoning Province, P. R. China

However, as Gustafsson has pointed out,[17] the confusing aspect of LV-based regression methods is understanding which type of LVs best model the observed data, since there are many ways to define uncorrelated LVs. Gustafsson used a simple numerical example inspired by the prediction of molecular activity from absorption spectra to show that neither PLS nor PCR will exactly recover a true underlying linear LV model in general. Although they may yield good predictions and sometimes offer interesting insights about the key underlying relationships, their chemical interpretation may be very difficult. Great effort has since been devoted to extracting latent features with more interesting statistical characteristics and chemically interpretable meanings to serve as descriptors in order to improve the performance of regression modeling.

In recent years, the theory of independent component analysis (ICA),[18–20] a further development of principal component analysis (PCA), has shown that to obtain a unique LV model, it is not in general enough to consider only up to second-order statistics (correlations, covariances) describing the measured variables. In ICA, the extracted components are assumed to be mutually independent statistically instead of merely uncorrelated. It attempts to recover statistically independent signal sources given only observations that are assumed to be linear mixtures of the original signals. To do so, ICA makes full use of higher-order statistics and applies criteria related to information theory and entropy. In probability theory, independence is a high-order statistic which guarantees uncorrelatedness so that it is a much stronger condition than uncorrelatedness that is only up to second-order moment. Any independent components (ICs) which can be extracted may, thus, provide more informative statistical information, be easier to interpret in chemical terms, and better reflect the intrinsic properties of the measured data. It has drawn considerable attention and found potential applications in process monitoring.[21–25] Li and Wang[21,22] used ICA to untangle variable dependencies, reduce the dimensionality of data on their monitored variables and identify trends in a processing operation. Kano et al.[23] demonstrated online process monitoring by charting the individual ICs as time series, which showed ICA's superiority over PCA-based methods. Lee et al.[24] constructed three ICA-based monitoring statistics to reveal multivariate correlations and determined their confidence limits in process monitoring. Moreover, ICA can generate meaningful representations of non-Gaussian data[25–27] and outperform PCA in revealing the LVs underlying noisy measurements.[28]

Motivated by the success of ICA, ICR[17,29–32] was developed as another alternative to classical LV-based regression methods (PLS and PCR). In ICR the model's LVs are designed to be as statistically independent as possible. PCR and PLS, in contrast, are designed to find LVs that are statistically uncorrelated, but not necessarily statistically independent. Substantial previous work[29–32] has demonstrated that ICR can better recover true LVs,[17] and, thus, facilitate meaningful interpretation. Chen and Wang[29] have applied ICA to near-infrared (NIR) spectra, and shown that it was able to identify the pure spectra of unknown constituent components from the spectrum of their mixture and estimate their concentrations. Shao et al.[31] built independent component regression models of NIR spectra and compared them with PCR-based models. They found that although the quantitative prediction ability was completely equivalent, the ICR approach was more powerful in qualitatively describing the intrinsic properties of the NIR spectra. Kaneko et al.[32] have demonstrated that a combination of ICA and MLR could extract effective components from explanatory variables and construct a regression model with high-predictive accuracy. Gustafsson[17] has made some useful comparisons between different regression algorithms. He demonstrated that while PLS could give more accurate quality predictions, the advantage of ICR lay in its ability to retrieve a more meaningful LV model.

Even when an LV model structured by ICR seems to have a chemical meaning, ICs decomposed from mixed signals are not guaranteed to be closely correlated with quality properties, since ICR extracts features without reference to their effects on quality. In calibration modeling, this creates a risk similar to that in principal component regression (PCR). That is, because process observations often contain major sources of variation that are of little or no predictive value, the first several ICs extracted may be capturing most of the process variation, but not necessarily explaining quality properties. It is not difficult to imagine that some useful predictive and interpretative information may end up in ICs which have been discarded, while systematic variations with little quality significance may be retained for calibration analysis. To make up for this deficiency, more ICs generally must be employed for a comprehensive quality description. This, of course, increases model complexity, although the additional information will normally improve the model's fit to the training reference data. Moreover, it will often cause a substantial reduction in the models' general predictive ability with new data.

The dominating role of LV extraction in determining the success of subsequent regression modeling makes it important to further improve LV modeling, and to obtain LVs better suited to the specific purpose of regression analysis, referring to both their statistical meaning and their quality-related characteristics. Since both PLS and ICR have their respective merits, a natural idea is to combine their advantages for LV extraction. In this study, such a method is designed based on a conventional ICR algorithm, but described as modified ICR (M-ICR). M-ICR tries to make the extraction of LVs more direct and more comprehensible for regression modeling. The method can be regarded as a joint LV-based multivariate regression modeling framework. The construction rule for new regressors encompasses and ties together the ideas of PLS and ICR under the same mathematical umbrella by constructing a generalized function for the objective criterion.

The following two problems of particular interest are addressed in this work:

1. Constructing a dual-objective mathematical optimization solution for better LV modeling. Higher-order statistical independence and quality-related requirements are considered simultaneously during the feature extraction procedure by constructing a synthetic dual-objective mathematical optimization solution. The corresponding search optimization algorithm is then developed. The optima are, thus, pursued maintaining a balance between statistical independence and quality-relevant characteristics, a balance which can readily

be controlled and adjusted by setting suboptimization weight parameters. Furthermore, by analyzing the effects of the two kinds of characteristics on regression modeling, proper weight parameters can be intentionally selected to get better LVs, and the advantages of PLS and ICR can be better combined.

2. Performing bootstrap-based quantitative calibration analysis for model validation and performance evaluation. Multiple ICR solutions are derived using a bootstrap technique, which represent different input spaces and can be readily aggregated. This can reduce the overdependence of model's performance on reference samples and enhance the model's reliability and robustness. Moreover, such bootstrap-based results provide the possibility of further quantitative calibration analysis. On the one hand, the contribution of each IC to the quality description can be inspected, and the results can be used to rank the extracted ICs and provide a reference for selecting the powerful ones for regression modeling. At the same time, a confidence interval can be developed for the quality prediction to help evaluate the M-ICR model's performance.

This article first presents the original ICR algorithm and analyzes its defects. The proposed M-ICR algorithm is then described, LV extraction demonstrated and a regression model constructed. A quantitative calibration analysis is then developed for model validation and performance evaluation. Simulation examples are presented to illustrate the performance of the proposed method. Discussion is conducted based on the results, highlighting the suitability of the proposed method, its advantages, but also analyzing its limitations. Finally, conclusions are drawn in the last section. Future directions and possible improvements for the proposed method are also pointed out.

Conventional ICR Algorithm

As a two-step calibration method, the basic idea of ICR is to first derive a LV model from the measured process data in terms of higher-order statistics in which the LVs are as statistically independent as possible. Such a representation is reported to be able to recover the real LVs, and, thus, capture the true structure of the measured data.[17,31] The LVs are then used to build a regression model based on least-squares algorithm. Clearly, this combination of ICA and MLR is conceptually very similar to PCR. The only difference is that ICs obtained by ICA are used in the regression instead of PCs obtained by PCA.

In the ICA algorithm, it is assumed that the measured variables $x_1, x_2,\ldots,x_j$ can be described as linear combination of R ($R \leq J$) independent components $s, s_2,\ldots, s_R$. The basic thrust of ICA is to estimate both the latent components $\mathbf{s}$, and the demixing relationship $\mathbf{W}_x$ ($R \times J$) from the process measurements $\mathbf{x}$ without any related prior knowledge, a process termed blind separation. The decomposition relationship can be expressed as follows

$$\mathbf{s} = \mathbf{W}_\mathbf{x}\mathbf{x} \tag{1}$$

where $\mathbf{s}(R \times J)$ denotes the independent component vector, which has unit variance $E(\mathbf{s}\mathbf{s}^\mathrm{T}) = \mathbf{I}$.

Before ICA feature extraction, it is assumed that the process data are pre-whitened, which is generally achieved by PCA, so that its components are uncorrelated and their variances all equal to one. This whitening transformation can be expressed as

$$\mathbf{z} = \Lambda^{-1/2}\mathbf{U}^\mathrm{T}\mathbf{x} = \mathbf{Q}\mathbf{x} \tag{2}$$

where $\mathbf{Q} = \Lambda^{-1/2}\,\mathbf{U}^\mathrm{T}$ is the whitening matrix, in which $\mathbf{U}$ (an orthogonal matrix of eigenvectors), and Λ (a diagonal matrix of eigenvalues) are generated from the eigenvalue decomposition of the covariance matrix $E(\mathbf{x}\mathbf{x}\mathrm{T})$.

Naturally, $\mathbf{s}$ from whitened vector $\mathbf{z}$ can be estimated by

$$\mathbf{s} = \mathbf{W}\mathbf{z} = \mathbf{W}\mathbf{Q}\mathbf{x} \tag{3}$$

where $\mathbf{W}$ is an orthogonal matrix, given that $E(\mathbf{s}\mathbf{s}^\mathrm{T}) = \mathrm{W}E(\mathbf{z}\mathbf{z}^\mathrm{T})\mathrm{W}^\mathrm{T} = \mathbf{W}\mathbf{W}^\mathrm{T} = \mathbf{I}$.

After this first-step analysis, the ICs are prepared as regressors for the second-step calibration analysis. A linear regression relating the extracted ICs $\mathbf{S}(N \times R)$ (where N is the number of observations), with the response variables $1 \leq a_1 \leq 2$ (where J_y is the number of predicted variables) can be readily calculated through a simple least-squares calculation

$$\begin{aligned} \Theta &= \left(\mathbf{S}^\mathrm{T}\mathbf{S}\right)^{-1}\mathbf{S}^\mathrm{T}\mathbf{Y} \\ \hat{\mathbf{Y}} &= \mathbf{S}\Theta \end{aligned} \tag{4}$$

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

It is possible to convert the regression relationship Θ to $\Theta_{\mathrm{X}\rightarrow\mathrm{Y}}$, which is directly related to the process measurement data $\mathbf{X}(N \times J)$

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{X}\mathbf{W}_\mathbf{x}^\mathrm{T}\Theta = \mathbf{X}\Theta_{\mathbf{X}\rightarrow\mathbf{Y}} \\ \Theta_{\mathbf{X}\rightarrow\mathbf{Y}} &= \mathbf{W}_\mathbf{x}^\mathrm{T}\Theta \end{aligned} \tag{5}$$

Proposed M-ICR Algorithm

Drawbacks of the conventional ICR algorithm

Although many investigators have reported that ICs extracted from operating data are able to capture the essential structure of the process and are chemically meaningful, conventional ICR-based regression nevertheless has some drawbacks.

First, in two-step calibration analysis, the performance of ICR greatly depends on how well the features are prepared in the LV extraction procedure. The LV model is extracted from the process measurements ($\mathbf{X}$) to capture the major process systematic variations using ICA exclusively. This estimation takes no account of the roles of the LVs in quality prediction, and, thus, is not optimized for quality prediction. Some ICs may be uninformative about quality and make no contribution to the quality description even though they are strongly independent of each other. If those ICs are used

directly to interpret quality variations, it is not difficult to imagine that more ICs may be required in the regression model to describe quality results properly, which may result in both a more complex model and over-parameterization defects.

In addition, the quality of the ICs extracted is also affected to some extent since conventional ICA bears the following disadvantages and ambiguities.[33] PCA can rank the extracted PCs using variance information, but in ICA the significance order of feature extraction is not controlled. Moreover, there is no explicit standard for determining how many ICs should be retained. Furthermore, it can lead to different ICA solutions even when ICA is run multiple times on the same data. These ambiguities essentially result from the fact that ICA cannot determine the order of the independent components because both $\mathbf{s}$ and $\mathbf{W_x}$ are unknown, and the variances of the extracted ICs are all defined to be unity. Therefore, the order of the terms can be freely changed in the decomposition relationship $s_i = \sum_{j=1}^{J} w_{x,i,j} x_j (i = 1, 2, \ldots, \mathrm{R})$, and any of the ICs can be regarded as the first. Random initialization of the demixing matrix $\mathbf{W_x}$ in the whitened space can then lead to different optimization solutions when applying ICA multiple times to the same data. In PCA only the first several PCs need to be derived, and they do not interact with each other, but in ICA, generally all potential ICs should be extracted,[33] and the retained ICs are then chosen from among them. It, thus, generates a high-computational load but, more importantly, inevitable ambiguities in the following calibration analysis. Lee et al.[33] have reported an improved ICA algorithm which can produce a unique and repeatable ICA solution by fixing the initialization instead of the random initialization of conventional ICA. Their method assumes that the space spanned by the major PCs is a good starting point for extracting ICs, because ICA can be viewed as a natural way to "fine-tune" PCA (centering and whitening) with an iterative search optimization procedure. By maintaining the variance information about the PCs during ICA feature extraction, Lee's method more successfully captures critical process variation features and gives a better visualization of the process data for multivariate statistical process monitoring (MSPM). However, for the specific objective of regression modeling and quality prediction, it is not proper to attribute significance to the ICs based on the variance of the PCs, since it is not directly related to quality attributes.

Clearly, regressor extraction is a crucial step in any calibration analysis. In the search for better regressors, questions of choice arise. PLS and conventional ICR algorithm each has its advantages, and it would be better if their merits could be combined. A dual-objective optimization solution promises to achieve this. In the proposed method, first, the mathematical basis of the combined objective function is formulated. It incorporates a weighted combination of quality-related characteristics and statistical independence during LV extraction and gives insight into the nature of the extracted LVs and their effects on the following regression. The adjustable suboptimization parameters in the continuum [0,1] play an important role, which can control and regulate the nature of the LVs extracted and determine the balance between PLS and ICR. Some remarks will be provided for the setting of optimization weights to help further comprehension. Regression relationship is then formulated relating the extracted LVs with the qualities.

A dual-objective optimization

In designing different optimization criteria and objective functions, different decomposing relationships are available to derive different latent features. Since both independence and quality-related statistical characteristics are desired, a synthetical optimization objective should be designed. The latent factors derived from process measurement are expected to simultaneously satisfy the following two objectives:

1. They should be as independent of each other as possible, which is achieved at certain optima of $\mathrm{E}\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}$. According to Hyvärinen and Oja,[19] the following choices of $\mathrm{G}(\cdot)$ have proved very useful

$$\mathrm{G}_1(u) = \frac{1}{a_1}\log\cos\mathrm{h}(a_1 u) \quad \text{and} \quad \mathrm{G}_2(u) = -\exp\left(-\frac{u^2}{2}\right) \tag{6}$$

where constant a_1 satisfies $1 \le a_1 \le 2$. $\mathrm{G}_1(u)$ is a good general-purpose contrast function,[18] and it will be used in the M-ICA iteration algorithm.

2. They should remain in as close a relationship as possible with the quality vector. In other words, they should maximize the absolute value of the covariance between them and the quality vector $\max(\mathrm{E}((\mathbf{w}^{\mathrm{T}}\mathbf{z})y))^2$ (where y is one sample of the normalized quality value). Since both $\mathbf{w}^{\mathrm{T}}\mathbf{z}$ and y have unit variance, the covariance actually is equal to the correlation coefficient. Moreover, in the case of multivariate quality variables, considering the multiple correlations together, the optimization index can be represented as $\max \frac{1}{J_y}\left(\mathrm{E}(\mathbf{w}^{\mathrm{T}}\mathbf{z}\mathbf{y})^{\mathrm{T}}\mathrm{E}(\mathbf{w}^{\mathrm{T}}\mathbf{z}\mathbf{y})\right)$, where $\boldsymbol{y}(\boldsymbol{J}_y \times 1)$ is the normalized $\boldsymbol{J}_y$ -dimensional quality vector.

Note that in terms of the first objective, maximizing independence boils down to maximizing or minimizing $\mathrm{E}\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}$, where the choice of maximization or minimization depends on the sign of $\mathrm{E}\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\} - \mathrm{E}\{\mathrm{G}(\upsilon)\}$. (Here υ is a standardized Gaussian variable.) The positive corresponds to the super-Gaussian and the negative to the sub-Gaussian. Therefore, the two optimization objectives could not be simply added. Instead of $\mathrm{E}\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}$, the mathematical form $\max(\mathrm{E}\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\} - \mathrm{E}\{\mathrm{G}(\upsilon)\})^2$ is adopted, which does not need to specify the super- or sub- Gaussian characteristics.

Making a tradeoff between the two maximization objectives and encapsulating them into an integrated dual-objective optimization model, the following mathematical function can be designed for LV extraction

$$\max\left(\alpha \cdot \frac{1}{J_y}\left(\mathrm{E}(\mathbf{w}^{\mathrm{T}}\mathbf{z}\mathbf{y})^{\mathrm{T}}\mathrm{E}(\mathbf{w}^{\mathrm{T}}\mathbf{z}\mathbf{y}) + \beta\left(\mathrm{E}\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\} - \mathrm{E}\{\mathrm{G}(\upsilon)\}\right)^2\right)\right) \tag{7}$$

$$\text{subject to the constraint} : \mathbf{w}^{\mathrm{T}}\mathbf{w} = 1 \tag{8}$$

where α and β, and are, respectively, the weight coefficients attached to each suboptimization objective. Here they are defined to satisfy $\alpha + \beta = 1$ for simplicity, as will be further

clarified in the discussion of optimization parameter setting which follows.

Based on the aforementioned formulation, a modified ICA estimator can be calculated by solving a mathematical dual-objective optimization problem in pursuit of both statistical independence and a close relationship with quality. On the basis of Hyvärinen's original fast fixed-point ICA algorithm (FastICA),[18,19] the following optimization solution, termed the modified FastICA algorithm (MFastICA), is mathematically formulated using an approximative Newton iteration scheme (the detailed deduction procedure is shown in Appendix A)

$$\begin{aligned} &\mathbf{w} \leftarrow \mathbf{w} + \gamma \cdot \mathbf{d}_n \\ &\text{i.e., } \mathbf{w} \leftarrow \mathbf{w} - \gamma \cdot \left(\mathrm{J}\Psi(\mathbf{w})_m\right)^{-1} \nabla \mathrm{F}_{\mathbf{w}} \end{aligned} \tag{9}$$

where $\mathrm{J}\Psi(\mathbf{w})_m$ is the modified Jacobian matrix (explained in Appendix B), which is invertible, and $\bigtriangledown \mathrm{F}_{\mathrm{w}}$ is the first-order derivative of $\mathrm{F}(\mathbf{w}, \lambda)$, with respect to $\mathbf{w}$ (see Eq. A11 in Appendix A).

Optimization parameter setting

Multiobjective optimization sometimes fails to find an absolutely optimal solution which simultaneously optimizes all the subobjectives. This usually results from a conflict between subobjectives. This is one essential difference compared with single-objective optimization. However, it is almost always possible to get a relatively optimal solution by making a compromise among the subobjectives involved. To obtain a relative optimum, two key issues should be taken into account: (1) each suboptimization objective function may have a different scale; and (2) they may converge at different speeds.[35] To avoid the adverse case in which one subobjective is overwhelmed by the other, a uniform quantity-scale balance between the two subobjectives should be determined during the initialization stage and applied during the iterative searching.

In this study, the weights are set as follows. First, the weight parameters α and β are initially and intentionally decomposed into three parts, respectively, $\alpha_{\mathrm{imp}} \cdot \alpha_{\mathrm{scale}} \cdot \alpha_{\mathrm{adj}}$ and $\beta_{\mathrm{imp}} \cdot \beta_{\mathrm{scale}} \cdot \beta_{\mathrm{adj}}$. α_{imp} and β_{imp}, are here called significance factors, and they represent the relative importance subjectively attached to each suboptimization objective. α_{scale} and β_{scale}, called scale factors, are used to unify the scales of the two subobjectives. α_{adj} and β_{adj}, here called adjustable factors, are adjusted to balance the speeds of convergence to the suboptimization objectives, which can be regulated in real time during the search procedure. Using the grade of each objective function to describe its speed of convergence, the adjustable factors can be defined as follows

$$\begin{aligned} \alpha_{\mathrm{adj}} &= 1 \Big/ \left\| \nabla F_{1,\mathbf{w}} \right\|^2 \\ \beta_{\mathrm{adj}} &= 1 \Big/ \left\| \nabla \mathrm{F}_{2,\mathbf{w}} \right\|^2 \end{aligned} \tag{10}$$

where the denominators are the squared norm of the first-order derivatives of the two suboptimization objectives, both of which are calculated in a way similar to the calculation of $\bigtriangledown \mathrm{F}_{\mathrm{w}}$ in Appendix A

$$\begin{aligned} \nabla F_{1,\mathbf{w}} &= 2\frac{1}{J_y}\mathrm{E}\left(\mathbf{z}\mathbf{y}^{\mathrm{T}}\right)\mathrm{E}\left(\mathbf{w}^{\mathrm{T}}\mathbf{z}\mathbf{y}\right) - 2\frac{1}{J_y}\mathbf{w}^{\mathrm{T}}\mathrm{E}\left(\mathbf{z}\mathbf{y}^{\mathrm{T}}\right)\mathrm{E}\left(\mathbf{w}^{\mathrm{T}}\mathbf{z}\mathbf{y}\right) \\ \nabla \mathrm{F}_{2,\mathbf{w}} &= 2\left(\mathrm{E}\left\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\right\} - \mathrm{E}\{\mathrm{G}(\upsilon)\}\right)\mathrm{E}\{\mathbf{z}g(\mathbf{w}^{\mathrm{T}}\mathbf{z})\} \\ &\quad - 2\left(\mathrm{E}\left\{\mathrm{G}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\right\} - \mathrm{E}\{\mathrm{G}(\upsilon)\}\right)\mathbf{w}^{\mathrm{T}}\mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}\mathbf{w} \end{aligned} \tag{11}$$

Therefore, α_{adj} and β_{adj} will be adjusted online during the optimization procedure. If one suboptimization objective changes faster, i.e., its squared grade norm is larger, the corresponding adjustable factor will be set smaller, and vice versa, making the corresponding suboptimization objectives change in parallel.

The grades will approximate zero when the iteration results are close to the optima, so the two parameters α and β, are restricted as follows

$$\begin{cases} \alpha = \alpha_{\mathrm{imp}} \cdot \alpha_{\mathrm{scale}} \cdot \alpha_{\mathrm{adj}} & \left\| \nabla F_{1,\mathbf{w}} \right\| \geq \mu \\ \beta = \beta_{\mathrm{imp}} \cdot \beta_{\mathrm{scale}} \cdot \beta_{\mathrm{adj}} & \left\| \nabla F_{2,\mathbf{w}} \right\| \geq \mu \\ \alpha = \alpha_{\mathrm{imp}} \cdot \alpha_{\mathrm{scale}} & \left\| \nabla F_{1,\mathbf{w}} \right\| \leq \mu \\ \beta = \beta_{\mathrm{imp}} \cdot \beta_{\mathrm{scale}} & \left\| \nabla F_{2,\mathbf{w}} \right\| \leq \mu \end{cases} \tag{12}$$

where μ is a predefined small threshold, which actually envelops a very narrow neighborhood around the optimum. Whenever the optimization search enters the neighborhood, the corresponding grade approaches zero, so the significance and scale factors alone describe the weight parameters.

For convenience, further formulate the two parameters satisfying $\alpha + \beta = 1$ using a simple ratio algorithm. For example, when $\left\| \nabla F_{1,\mathbf{w}} \right\| \geq \mu$ and $\left\| \nabla F_{2,\mathbf{w}} \right\| \leq \mu$, the weights can finally be calculated as

$$\alpha = \frac{\alpha_{\mathrm{imp}} \cdot \alpha_{\mathrm{scale}} \cdot \alpha_{\mathrm{adj}}}{\alpha_{\mathrm{imp}} \cdot \alpha_{\mathrm{scale}} \cdot \alpha_{\mathrm{adj}} + \beta_{\mathrm{imp}} \cdot \beta_{\mathrm{scale}}}, \quad \beta = 1 - \alpha \tag{13}$$

So the final weight parameters α and β arise from considering the effects of the three factors together.

The optimization objective expressed in Eq. 7 implies that the extracted ICs can be controlled by adjusting the weight parameters attached to each subobjective. If $\alpha = 0$, only the statistical independence index is being optimized, and the MFastICA algorithm will converge to the common ICA solution, which only considers higher-order statistical independence. If $\beta = 0$, only the quality-related optimization objective holds, and the method derives LVs closely related with quality. Keeping both α and β nonzero results in a compromise between statistical independence and quality-related characteristics, and a combination of PLS and ICR algorithms to some extent. By setting different weight parameters step-by-step, it is feasible to interrogate in detail the respective effects of the two characteristics on multivariate calibration modeling, from which one can choose the desired balance. The problem of selecting the subobjective weights is not something for which specific rules can easily be given. It greatly depends on experience with the process and the specific objective of the study—which characteristics are more important in the current practical application. When no prior information exists, as a rule of thumb, the

weight parameters can be determined by cross-validation and trial and error.

M-ICR calibration modeling

In any optimization problem, the solution depends on the initial settings and the searching algorithm, which, respectively, determine the starting point and the search path. Both aspects are important, therefore, in achieving the objectives specified in Eq. 7. In the proposed method, normalized LVs extracted from PLS are first set as the initial conditions for further ICA iterative optimization. Such LVs should be good initial estimates of ICs from a quality viewpoint

$$\mathbf{t} = \mathbf{R}^{\mathrm{T}}\mathbf{x} \tag{14}$$

where $\mathbf{t}(d \times 1)$ is the d-dimensional latent factor vector retained after decomposing $\mathbf{x}(J \times 1)$ using PLS. $\mathbf{R}(J \times d)$ is the PLS weights matrix. $\mathrm{E}\{\mathbf{t}\mathbf{t}^{\mathrm{T}}\} = \Lambda = \mathrm{diag}\{\lambda_1,\ldots,\lambda_d\}$, in which λ_i $(i = 1, 2, \ldots, d)$ is the variance of the LVs retained from PLS. The number of LVs in the whitening procedure should be kept as large as possible,[33] since they will provide information essential for constructing ICs through linear combination during the subsequent IC extraction. So only those LVs with variance values closely approximating zero should be excluded to avoid singularity problems. Equation 14 can then be further normalized as

$$\mathbf{z} = \Lambda^{-1/2}\mathbf{t} = \Lambda^{-1/2}\mathbf{R}^{\mathrm{T}}\mathbf{x} = \mathbf{Q}\mathbf{x} \tag{15}$$

where $E\{\mathbf{z}\mathbf{z}^{\mathrm{T}}\} = \mathbf{I}$ ($\mathbf{I}(d \times d)$ is an identity matrix), and $\mathbf{Q} = \Lambda^{-1/2}\mathbf{R}^{\mathrm{T}} = \Lambda^{-1/2}\left(\mathbf{W}_{\mathrm{pls}}(\mathbf{P}_{\mathrm{pls}}^{\mathbf{T}}\mathbf{W}_{\mathrm{pls}})^{-1}\right)^{\mathbf{T}}$ (where $\mathbf{W}_{\mathrm{pls}}$ is the weights matrix and $\mathbf{P}_{\mathrm{pls}}$ the loading matrix from PLS) is the whitening matrix. The scaled latent scores $\mathbf{z}$ are the preprocessed process data after centering and whitening. Note that despite its scaled variance, $\mathbf{z}$ has been ordered according to the covariance between $\mathbf{t}$ and quality. In particular, statistical dependence up to the second-order has been removed. This gives a good initialization for further ICA iterative approximation. From there a consistent LV model structure, specifically the demixing matrix $\mathbf{W}$, can be obtained. The detailed iterative modeling procedure is given in Appendix C. Once the demixing matrix has been specified, the ICs from the initial process observations can be derived using the following expression

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{z} = \mathbf{W}\mathbf{Q}\mathbf{x} = \mathbf{W}\Lambda^{-1/2}\mathbf{R}^{\mathrm{T}}\mathbf{x} = \mathbf{W}_{\mathbf{x}}\mathbf{x} \tag{16}$$

After obtaining the desired ICs, their regression relationship with the quality variables can be set up using simple least-squares algebra, as shown in Eq. 4.

Bootstrap-based Quantitative Calibration Analysis

The preceding section has shown how LVs can be retrieved and how a quantitative regression relationship can be formulated. Furthermore, two problems require attention: the proper ranking of the regressors accepted and the method's robustness in application.

Previous ICR modeling research has not properly addressed model validation despite its fundamental importance. Too many regressiors increase model complexity and introduce redundant noise, but too few may lose some important quality information. Lee et al.[33] have summarized some statistical criteria suitable for determining the number of ICs for process monitoring. Liu et al.[26] developed a negentropy-based test procedure to rank ICs according to their degree of non-Gaussianity, which, in turn, could be used to categorize the importance of the ICs and thus to assess how many components should be retained. However, both groups pursued the retrieval of the major non-Gaussian information or the reconstruction the recorded variables. Their methods are not suitable for the specific purpose of regression modeling, since neither directly addresses quality prediction and interpretation. Westad and Kermit[36] presented a cross validation scheme to verify ICA modeling results, but it called for procrustes rotation[37] to rotate the ICA components, since the order of an ICA solution is arbitrary due to the ambiguity previously discussed. This demands particular caution and makes cross validation difficult.

On the other hand, statistical analysis results are to some extent sensitive to the samples used for modeling as a result of various factors including contamination by often significant levels of experimental noise. The calibration modeling results may be, thus, biased and different if employing different modeling samples. So, how to make full use of the samples to create a model as robust as possible while avoiding overfitting? At the same time, another problem in the practical application of the ICR method is the lack of confidence bounds for evaluating its performance in predicting quality. An attractive approach to improving calibration robustness is to develop a set of models and combine them. Zhang et al.[38–40] have made a series of investigations focusing on enhancing the accuracy and robustness of inferential models built from a limited amount of reference data. In their methods, the available data were re-sampled using a bootstrap technique[41,42] to generate several sets of training data. Multiple neural network models were then developed for each set, and finally they were combined to form a bootstrap aggregated inferential model. Moreover, a confidence interval for neural network model predictions was also obtained by estimating their standard errors using bootstrap re-sampling technique, which could give the process operator extra information about the predictions.[38]

Such a bootstrap technique is applied in this study in an attempt to improve the reliability and robustness of the calibration models. Through training with various bootstrap re-sampled datasets, multiple ICR models are derived which model different regions of the input space. Their combination is intended to reduce overdependence of the model's performance on the particular reference samples used. The bootstrap-derived ICR models are then used to develop a statistical calibration analysis to validate the combined model and evaluate its performance. The method inspects the specific significance of each IC in the regression analysis, identifies the correct rank, and sets up the final M-ICR calibration model for quality prediction and interpretation. Moreover, a confidence interval for the quality predictions is computed which can be used to evaluate the combined model's performance.

Bootstrap-based model validation

In order to find the regressors most significant for quality prediction, the key is to define a proper analysis criterion. Here, it is defined by analogy with the reliability index used in uninformative variable elimination (UVE).[43–45] The criterion is termed the contribution index and defined as follows

$$\mathrm{Ct}_{j_y j_s} = \left|\mathrm{mean}(\Theta_{j_y j_s})\right| / \mathrm{std}(\Theta_{j_y j_s}),\ j_s = 1, 2, \ldots, d, j_y = 1, 2, \ldots, J_y \tag{17}$$

where $\Theta_{j_y j_s}$ ($N_b \times 1$) is composed of the regression coefficients of the j_sth IC when explaining $\mathbf{Y}_{j_y}$ collected from N_b bootstrap datasets. mean($\Theta_{j_y j_s}$) and std($\Theta_{j_y j_s}$), respectively, estimate the mean and standard deviation from the vector $\Theta_{j_y j_s}$ ($N_b \times 1$).

Briefly, bootstrap-based model validation involves six steps:

1. The initial dataset $\{\mathbf{X}, \mathbf{Y}\}$ is randomly reordered N_b times to generate the corresponding number of bootstrap data sets in which the data are the same but ordered differently.

2. Each bootstrap dataset is separated into a calibration subset of n_{tr} and a validation subset of n_{te} items. This results in N_b calibration datasets $\{\mathbf{X}_{tr}, \mathbf{Y}_{tr}\}_i$, and validation datasets $\{\mathbf{X}_{te}, \mathbf{Y}_{te}\}_i$, ($i = 1,2,\ldots,N_b$) covering different data samples and representing different regions of the input and output spaces.

3. The proposed M-ICR algorithm is applied to each calibration dataset $\{\mathbf{X}_{tr}, \mathbf{Y}_{tr}\}_i$ to generate the regression coefficient matrix Θ^i ($d \times J_y$) ($i = 1,2,\ldots,N_b$) relating the ICs and quality.

4. The contribution power $\mathrm{Ct}_{j_y j_s}$, of each IC describing its role in explaining each dependent variable $\mathbf{Y}_{j_y}$ is then calculated using Eq. 17. Then, to select the most powerful descriptors for interpreting the multivariate quality variables, the combined contribution power of each descriptor variable Ct_{j_s} is calculated by summing the contribution values ($\mathrm{Ct}_{j,js}$), corresponding to all the quality variables concerned. This produces a descending order of priority for the ICs as indicated by their decreasing Ct_{j_s}.

5. The mean-squared prediction error (MSE) for each bootstrap validation subset is then calculated as $\mathrm{MSE}^i = \frac{1}{n_{te} J_y} \sum_m^{n_{te}} \sum_j^{J_y} (y^i_{m,j} - \hat{y}^i_{m,j})^2$ (where subscripts m and j denote the validation sample and quality variable, respectively; $\hat{y}$ is the quality prediction, and y is the real quality measurement). These MSE^i index values from all the bootstrap validation subsets are then summed to quantitatively evaluate these bootstrap models' prediction performance.

6. Removing the IC with the least contribution and excluding its regression coefficient in each bootstrap model Θ^i results in new quality predictions and a new MSE index. This new MSE value is compared with the previous one. If the prediction errors have decreased, the exclusion of the IC from the regression model is accepted. The process is then repeated until no further decrease in prediction errors result, and the model from the previous iteration step is accepted as the final one. (When candidate ICs are numerous, to simplify the calculations involved in evaluating all the possible regressors one by one, an interval can be used instead of individual ICs and these intervals searched step by step.).

These steps eventually sort out d_s ICs as the descriptors of the bootstrap M-ICR model. They form the predigested regression matrix $\Theta^i(d_s \times J_y)$ ($i = 1,2,\ldots,N_b$), which is most informative with respect to multivariate quality variables. Moreover, the ICA decomposition model for each bootstrap dataset can also be corrected as $\mathbf{W}^i_{\mathbf{x}}(d_s \times J)$, by deleting the row corresponding to each IC removed. So the direct regression relationship between process variables and quality can be defined as $\Theta^i_{\mathrm{X}\to\mathrm{Y}} = \mathbf{W}^{i\mathrm{T}}_{\mathbf{x}} \Theta^i$.

Here it should be pointed out that the following preconditions can guarantee the success of bootstrap-based model validation:

1. On the basis of consistent and repeatable ICA decomposition results, the independent regressors from various bootstrap-based ICR solutions can be listed in a homologous order and can be readily stacked.

2. The regression coefficient can be regarded as the magnitude of the contribution of each IC to quality because the regressors (ICs) are mutually independent. It is appealing that there is then no need to rebuild the regression model when excluding each IC. Only removing the row in the coefficient matrix $\Theta^i(d \times J_y)$ corresponding to the deleted IC is required.

Performance evaluation

This modeling procedure determines the proper rank of all the M-ICR models used in the bootstrapping, and the final aggregated M-ICR model is designed by stacking those bootstrap models. As Zhang[39] has pointed out, since these bootstrap M-ICR models reveal similar calibration relationships, they are highly correlated, and the appropriate aggregating weights can be obtained through PCR. In this study, the stacking weights attached to each bootstrap M-ICR model are defined as follows:

1. Process measurement $\mathbf{X}$ is inserted into all N_b bootstrap M-ICR models to get the bootstrap quality prediction matrix $\hat{\mathbf{Y}}^i(N \times J_y)(i = 1, 2, \ldots, N_b)$. The quality attributes $\hat{\mathbf{y}}^i(N \times 1)$ are separated and collected in J_y matrices $\left[\hat{\mathbf{y}}^1, \hat{\mathbf{y}}^2, \ldots, \hat{\mathbf{y}}^{N_b}\right]_m (m = 1, 2, \ldots, J_y)$.

2. For each quality attribute, the appropriate stacking weight vector attached to each bootstrap M-ICR model, $[w^1, w^2, \ldots, w^{N_b}]_m$ is then derived by applying PCR to $\left[\hat{\mathbf{y}}^1, \hat{\mathbf{y}}^2, \ldots, \hat{\mathbf{y}}^{N_b}\right]_m$, and the real quality attribute vector $\boldsymbol{y}(N \times 1)$.

3. The final M-ICR model corresponding to each quality variable $\Theta_{\mathbf{x},m}(d_s \times 1)$, is obtained as a weighted linear combination of the N_b bootstrap M-ICR models: $\Theta_{\mathbf{x},m} = \frac{1}{N_b} \sum_i^{N_b} w^i \Theta^i_{\mathbf{x},m}$, where $\Theta^i_{\mathbf{x},m}(d_s \times 1)$ is the regression coefficient between process variables and each quality variable for the ith bootstrap dataset.

4. The M-ICR model for multivariate quality variables is formulated as $\Theta_{\mathbf{x}}(d_s \times J_y)$ by placing all $\Theta_{\mathbf{x},m}(d_s \times 1)$ side by side.

For each quality variable at each sampling time, the bootstrap quality predictions can be used to derive a simple confidence region to evaluate the quality prediction performance, which is defined as follows:

$$\bar{\hat{y}}_{m,n} = \frac{1}{N_b}\sum_{i=1}^{N_b}\hat{y}^i_{m,n}$$

$$std(\hat{y}_{m,n}) = \left\{\frac{1}{N_b-1}\sum_{i=1}^{N_b}\left(\hat{y}^i_{m,n}-\bar{\hat{y}}_{m,n}\right)^2\right\}^{1/2} \quad (18)$$

$$m = 1,2,\ldots,J_y;\ n = 1,2,\ldots,N$$

The 95% confidence interval applicable to the prediction result with respect to each quality variable at each time can be calculated as

$$\left[\bar{\hat{y}}_{m,n} - 1.96\cdot std(\hat{y}_{m,n}), \bar{\hat{y}}_{m,n} + 1.96\cdot std(\hat{y}_{m,n})\right] \quad (19)$$

The confidence bound indicates how confident an estimation of quality is.[38] The narrower these confidence bounds, the higher the reliability of the estimation. The calculated confidence bounds, can, thus, help process operators decide whether to believe the model's quality predictions.

Simulations and Discussions

In this section, the performance of calibration modeling and quality prediction using the proposed method is illustrated through both numerical and real cases. Quantitative and qualitative analyses are performed with respect to the effects of the proposed optimization criterion, as well as different suboptimization weight values on recovery of real LV structure, and the improvement of the modeling performance resulting from the use of bootstrap technique.

Case study 1

The performance of M-ICR will first be compared with those of PLS and PCR when it comes to interpreting the extracted LVs. The example is similar to that used by Lee et al.[33]

Consider three source variables with the following distributions

$$\begin{aligned} s_1(k) &= 2\cos(0.08k)\sin(0.06k)\\ s_2(k) &= \text{sign}[\sin(0.3k) + 3\cos(0.1k)]\\ s_3(k) &= \text{uniformly distributed noise in the range}[-1,1] \end{aligned} \quad (20)$$

From the three source signals, process data are generated by linear mixing as $\mathbf{x}^T = \mathbf{s}^T\mathbf{A}$ with the following mixing matrix

$$\mathbf{A} = \begin{bmatrix} 0.86 & -0.55 & 0.17 & -0.33 & 0.89\\ 0.79 & 0.65 & 0.32 & 0.12 & -0.97\\ 0.67 & 0.46 & -0.28 & 0.27 & -0.74 \end{bmatrix} \quad (21)$$

A single-quality variable is defined by $\mathbf{y} = 2\mathbf{s}_1 + 7\mathbf{s}_2$, so quality attribute has no direct relationship with the third source signal.

A thousand mixes are generated for model training, and six different levels of normal-distributed random noises with the mean of zero, and different variance values (σ) as shown in Table 1 are added to the data. Taking the case with variance, $\sigma = 0.5$ for example, the real source signals and the mixed data are shown in panels a, b and c of Figure 1. The feature extraction results using PLS, the conventional FastICA, modified ICA,[33] and the proposed M-ICR algorithm (taking $\alpha_{imp} = 0.7$ and $\beta_{imp} = 0.3$, for example) are shown in panels d through g. Compared with the original sources, the first two LVs extracted by PLS (Figure 1d) do not recover the quality-related ones. This result may be because

Table 1. Quality Prediction Results with Different Algorithms Involving No Bootstrapping

Regression method		MSE				
		1 LV	2 LVs	3 LVs	4 LVs	5 LVs
Noise variance $\sigma = 0.1$	PLS	0.2145	0.1338	0.1330	0.1321	0.1317
	original ICA	0.9962	0.9770	0.1352	0.1351	0.1317
	Lee's M-ICA	0.2221	0.1325	0.1327	0.1318	0.1317
	proposed M-ICR	0.1521	0.1317	0.1317	0.1318	0.1317
Noise variance $\sigma = 0.3$	PLS	0.3321	0.2697	0.2675	0.2676	0.2670
	original ICA	0.9980	0.3514	0.2741	0.2672	0.2670
	Lee's M-ICA	0.3584	0.3564	0.2731	0.2670	0.2670
	proposed M-ICR	0.2712	0.2667	0.2668	0.2670	0.2670
Noise variance $\sigma = 0.5$	PLS	0.3993	0.3500	0.3473	0.3472	0.3469
	original ICA	0.9998	0.4442	0.4248	0.3502	0.3492
	Lee's M-ICA	0.4367	0.4223	0.3468	0.3469	0.3469
	proposed M-ICR	0.3481	0.3468	0.3467	0.3469	0.3469
Noise variance $\sigma = 0.7$	PLS	0.4474	0.4077	0.4049	0.4047	0.4046
	original ICA	0.4858	0.4741	0.4135	0.4047	0.4046
	Lee's M-ICA	0.4948	0.4772	0.4777	0.4047	0.4046
	proposed M-ICR	0.4047	0.4045	0.4044	0.4045	0.4046
Noise variance $\sigma = 0.9$	PLS	0.4855	0.4531	0.4503	0.4501	0.4501
	original ICA	1.0026	0.5412	0.5035	0.4501	0.4501
	Lee's M-ICA	0.5427	0.5244	0.5075	0.4531	0.4501
	proposed M-ICR	0.4500	0.4500	0.4499	0.4500	0.4501
Noise variance $\sigma = 1$	PLS	0.5020	0.4726	0.4698	0.4697	0.4697
	original ICA	0.5557	0.5517	0.5185	0.4705	0.4697
	Lee's M-ICA	0.5650	0.5484	0.5496	0.5161	0.4697
	proposed M-ICR	0.4695	0.4695	0.4696	0.4696	0.4697

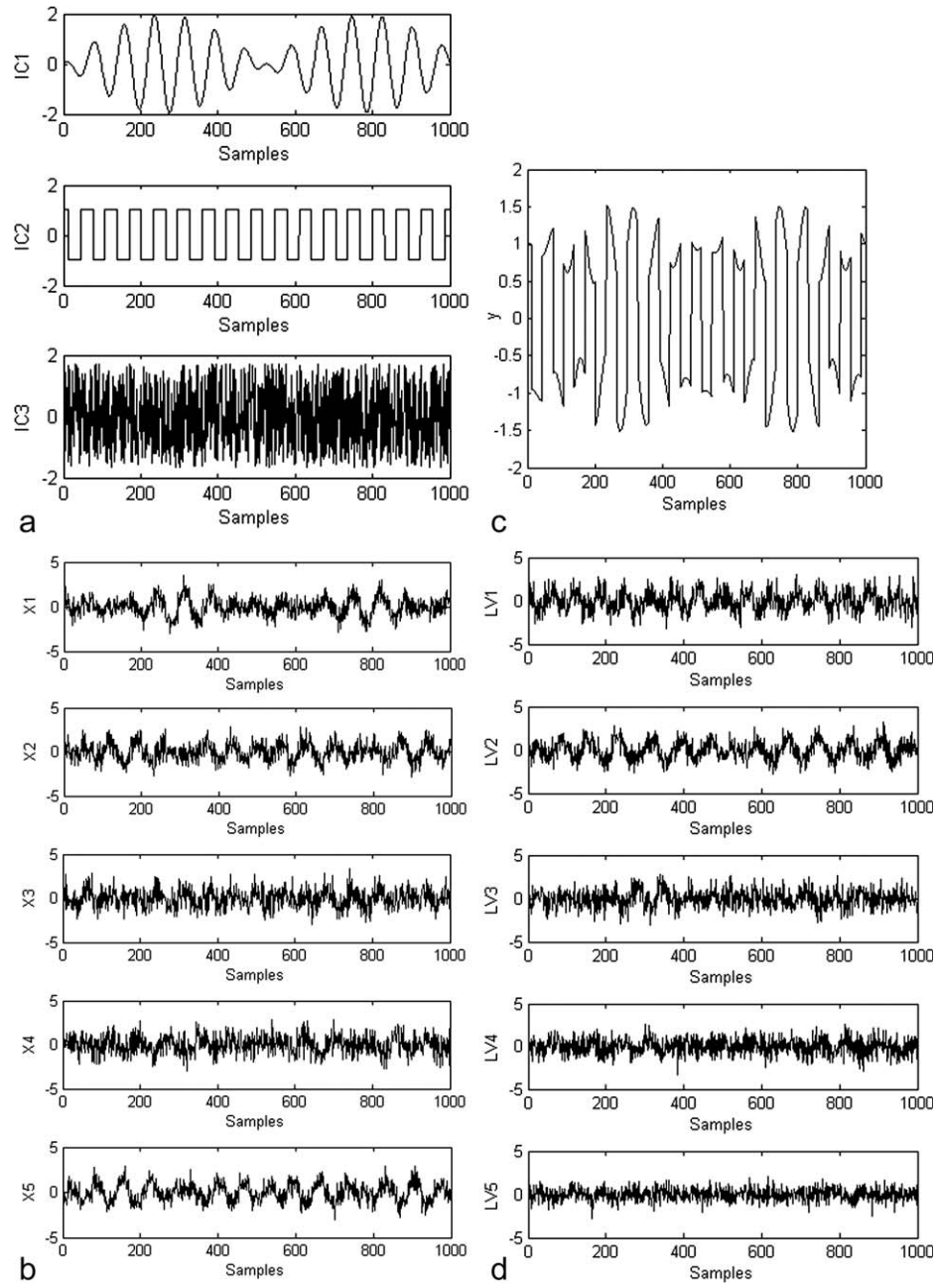

Figure 1. Real measurements and features estimated using the different algorithms.

(a) Real source signals, (b) normalized process measurements, (c) normalized quality variable, (d) LVs estimated using PLS, (e) ICA solution with the original FastICA algorithm, (f) ICA solution with Lee's modified ICA algorithm, and (g) proposed M-ICR solution.

the PLS algorithm focuses more on quality prediction than on recovery of true underlying sources, so the LVs tend to be linear combinations of source signals. In contrast, the conventional ICA and Lee's modified ICA algorithm pay more attention to the recovery of the true LV structure. The ICA solutions shown in Figure 1e and f reveal that both FastICA and modified ICA[33] recover the original sources, but this just provides an independent representation and reconstruction of the process data with no consideration of quality prediction or interpretation. In the FastICA solution shown in Figure 1e both original sources are recovered, but there is no guarantee that they will be chosen as significant regressors. In the M-ICA solution (Figure 1f), the first three ICs will be chosen as dominant ICs for regression modeling according to the work of Lee et al.[33] but the second extracted IC is more similar to the third source signal, which has no relationship with quality. Its participation in regression modeling will not improve the accuracy of quality predictions. Comparatively, using the proposed method, the two dominant ICs which really and significantly contribute to quality can be automatically identified and used for regression modeling, as shown in Figure 1g. The focus is entirely on sources which are really responsible for quality, and the influence of quality-irrelevant sources is excluded. The

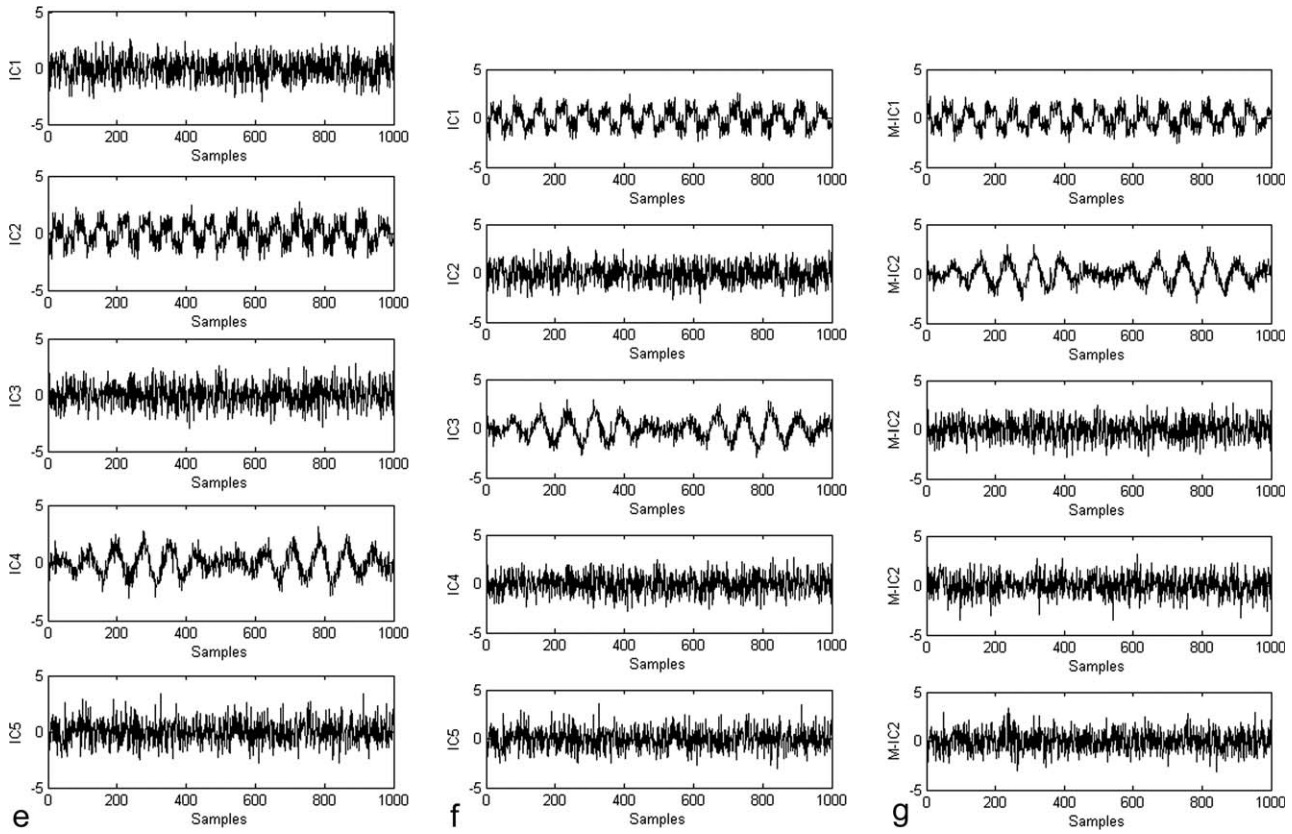

Figure 1. (Continued)

model structure involves fewer LVs, and interpretation is easier.

Under different levels of noise, the predictive performance of PLS, original ICA, Lee's modified ICA, and the proposed M-ICR algorithm are compared in Table 1. (The suboptimization parameters are set to be $\alpha_{\text{imp}} = 0.7$ and $\beta_{\text{imp}} = 0.3$ by trial and error. It should be noted, however, that $\alpha_{\text{imp}} > \beta_{\text{imp}}$ does not mean that the quality-related criterion is absolutely more important than statistical independence. This is actually determined by the joint influence of the three parts of the weight parameters as defined by Eq. 13.) The results are based on average over 1,000 testing samples and the different numbers of LVs recovered, where bootstrap-based model validation is not used. The popular performance index, mean-squared error (MSE) is used to quantitatively evaluate the prediction power of the regression models. It is calculated as $\text{MSE}_j = \frac{1}{N}\sum_{i=1}^{N_{te}} \left(y_{i,j} - \hat{y}_{i,j}\right)^2$ (where N_{te} is the number of testing samples; j denotes the quality variables, and here it is 1; $\hat{y}_{i,j}$ is the quality prediction corresponding to the real quality measurement $y_{i,j}$). The smaller the MSE value, the better the quality prediction result. Since a quality-related criterion is accommodated in the feature extraction, the ICs are automatically sifted out to be more suitable for calibration analysis, and, thus, provide possibility for improving quality predictions. Generally, the results confirm the superiority of the proposed M-ICR algorithm, which not only yields an interpretable LV model, but also shows prediction performance comparable to PLS in this numerical example. Specifically, the following findings can be derived.

Table 2. Optimization Parameters Using Correlation Analysis

		Correlation coefficient				
Parameters [α,β]	Objects	Original source 1	Original source 2	Original source 3	Quality variable	MSE (2 LVs)
[0.1,0.9]	M-IC1	−0.0554	0.9284	0.1358	0.8846	0.1325
	M-IC2	0.9641	0.0153	0.1010	0.2829	
	M-IC3	0.0550	0.0060	−0.8779	0.0211	
[0.7,0.3]	M-IC1	0.1100	0.9178	0.1404	0.8846	0.1317
	M-IC2	0.9638	−0.1408	0.0170	0.2827	
	M-IC3	−0.0081	0.0051	−0.8833	0.0224	
[0.9,0.1]	M-IC1	0.2463	0.8884	0.1386	0.9296	0.1336
	M-IC2	−0.2351	0.0674	−0.4173	0.0153	
	M-IC3	0.7876	−0.2259	−0.4786	0.0013	

Table 3. Bootstrap-Based Model Validation and Quality Prediction Results

No. of LVs	Contribution index (Ct)											
	Noise variance $\sigma = 0.1$		Noise variance $\sigma = 0.3$		Noise variance $\sigma = 0.5$		Noise variance $\sigma = 0.7$		Noise variance $\sigma = 0.9$		Noise variance $\sigma = 1$	
	M-ICR	PLS	M-ICR	PLS	M-ICR	PLS	M-ICR	PLS	M-ICR	PLS	M-ICR	PLS
1	1589.93	446.5832	1136.43	377.83	879.32	349.58	593.60	327.95	426.70	309.39	372.32	299.91
2	173.52	109.2810	89.46	67.39	44.20	54.19	25.14	49.82	17.73	48.84	15.45	48.97
3	3.51	189.4121	4.02	150.87	6.52	119.03	3.19	97.23	0.87	83.19	1.04	77.99
4	7.20	54.9743	0.59	66.61	2.76	45.11	1.16	34.62	0.33	28.51	0.01	26.32
5	2.05	166.6806	12.99	158.55	3.08	137.65	0.01	121.89	0.46	110.58	0.76	106.09
MSE	0.1279	0.1298	0.1995	0.2617	0.2710	0.3394	0.2716	0.3953	0.3367	0.4392	0.3655	0.4582

When the number of LVs is small, PLS yields better quality predictions than conventional ICA or Lee's M-ICA algorithm, although ICA recovers the true LVs. This may result from the fact that the first LVs extracted by PLS might be linear combinations of the true independent sources, which may have to be described by more ICs. With more numerous LVs, the MSE values with respect to all algorithms do not necessarily decrease continuously or significantly. This may be because the extracted LVs are not ranked according to their effects on quality prediction, so sometimes ICs bearing little quality significance may be extracted prior to quality-related ones. Such earlier LVs are not excluded from regression modeling although they are not very efficient for quality description, so their presence will affect the subsequent improvements in prediction performance even as quality-related LVs are added. This phenomenon highlights the necessity of sorting out the real quality-related LVs for regression modeling through, for example, a bootstrap technique.

With the proposed M-ICR algorithm, generally only two LVs are needed to achieve smaller MSE values, which agrees with the real case in which the quality variable is determined by two source signals. Increasing the number of LVs extracted for regression analysis would not further improve the quality prediction performance, but only complicate the model structure. With the proposed M-ICR algorithm in particular, the results may even deteriorate to some extent with the addition of extra LVs, since they may introduce quality-irrelevant variations.

Imposing noise generally degrades the quality prediction performance of all the algorithms. The exception of conventional ICA may be due to the randomness in conventional ICA. That is, when the noise level is low, the first ICs recovered may not be very informative about quality, whereas at higher noise levels, the first ICs recovered may by chance be more related with quality than those obtained when the noise level is lower. It is interesting that sometimes conventional ICA generates better quality predictions than Lee's M-ICA algorithm, generally at higher noise levels. This may have resulted from the fact that the noise begins to account for an increasing proportion of the real underlying information as the noise variance increases. Lee's M-ICA algorithm, which uses PCs for ICA initialization based on variance information, may be falsely led to identify noise sources as PCs.

To inspect the effects of the suboptimization significance factors α_{imp} and β_{imp} on feature extraction, they can be manually adjusted and the derived features evaluated in terms of their relationships with both the source signals and quality by correlation analysis. The results are summarized in Table 2 taking the first three LVs and three kinds of parameters settings for example. The quality prediction results are also compared based on a MSE criterion with respect to different parameters. Increasing α_{imp} and decreasing β_{imp} puts more weight on quality-related characteristics. Therefore, the extracted LVs should be more correlated with the quality variable, and less similar to the true underlying sources, which is generally the case in Table 2.

To improve the robustness of calibration modeling and also obtain a quantitative evaluation of modeling performance, the bootstrap technique is implemented on the basis of the original 1,000 training samples. This generates multiple ($N_b = 100$) bootstrap datasets, each with 800 modeling samples and 200 samples for testing, from which N_b regression parameters are derived. Then for each sample the same number of bootstrap-based quality estimations are obtained and used for model validation and performance evaluation. The analysis results are shown in Table 3 for the proposed M-ICR method in comparison with the PLS algorithm. The use of the bootstrap technique can, on the one hand, weaken the adverse effects of noise in the process and help to rank those LVs really related with quality, and, thus, determine the

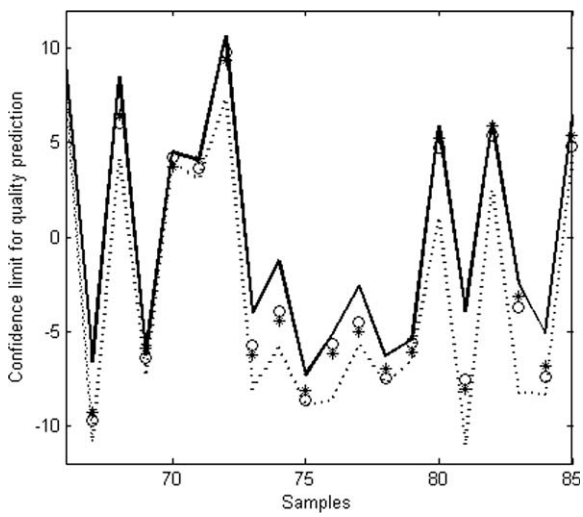

Figure 2. Quality prediction result and the corresponding confidence region (" * ", real measurement;"o", predicted quality value; solid line, the upper 95% confidence limit; dashed line, the lower 95% confidence limit).

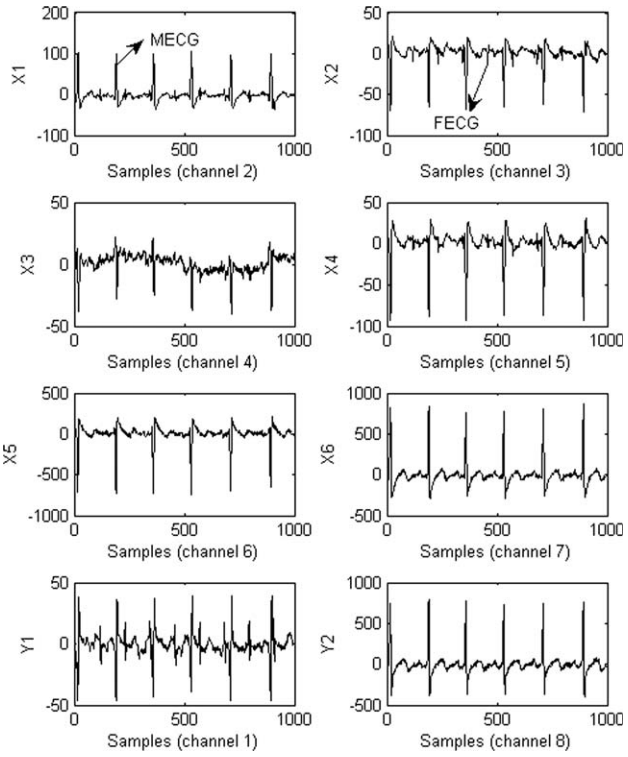

Figure 3. An eight-channel set of cutaneous ECG data.

proper order of the model. At the same time it can quantify the reliability of the quality prediction. Using the bootstrap technique, the retained LVs are sorted out and in this case only two LVs prove necessary for M-ICR, whereas three LVs are required for PLS, both less than the analysis result without bootstrapping shown in Table 1. Moreover, using bootstrap-based contribution index defined in Eq. 17, the originally extracted PLS LVs are shown to be not in accordance with the size of their contribution indices, and the LVs finally chosen are different from the first three LVs extracted in the original PLS. This illustrates the utility of bootstrap-based model validation. With the M-ICR method, only two LVs are needed for regression modeling in this case, yielding a simpler model structure but prediction results comparable to those from PLS. Moreover, from the bootstrap-based quality prediction results it is easy to derive a simple confidence region as defined by Eq. 18. For the sake of convenience, taking 25 samples for example, the quality prediction results with the first level of noise (variance $\sigma = 0.1$) are illustratively shown in Figure 2 along with the corresponding bootstrap-based confidence region. The figure gives a visual impression of the quality variations which can be captured.

Case study 2: Fetal electrocardiogram (FECG)

In this example, the eight-channel cutaneous potential recordings measured on a pregnant woman's skin are used to test the performance of the proposed method. The data is available from the following website: http://homes.esat.kuleuven.be/~smc/daisy/daisydata.html. Analysis of the fetal heart rate (FHR) has become a routine procedure for the evaluation of the well-being of the fetus, which is realized by examining the electrical activity of fetal heart. Due to various factors, the electrocardiogram (ECG) recordings are often measured on the mother's skin instead of directly placing an electrode on the fetal scalp. In this example used here, potential measurements are obtained from eight electrodes located on a pregnant woman's skin (cutaneous recordings), which contain contributions from several bioelectric phenomena (maternal and fetal activity, potential distributions generated by respiration and stomach activity, etc.), and are affected by various kinds of noise (thermal noise, noise from electrode-skin contact, etc.). Here, to highlight the effects of sampling, only 200 samples are used for model training, each covering one maternal heart beat and one fetal heart beat. Another 1,000 samples are used as test data to evaluate the modeling performance, and these include multiple beat-to-beat heart signals. Figure 3 shows the potential

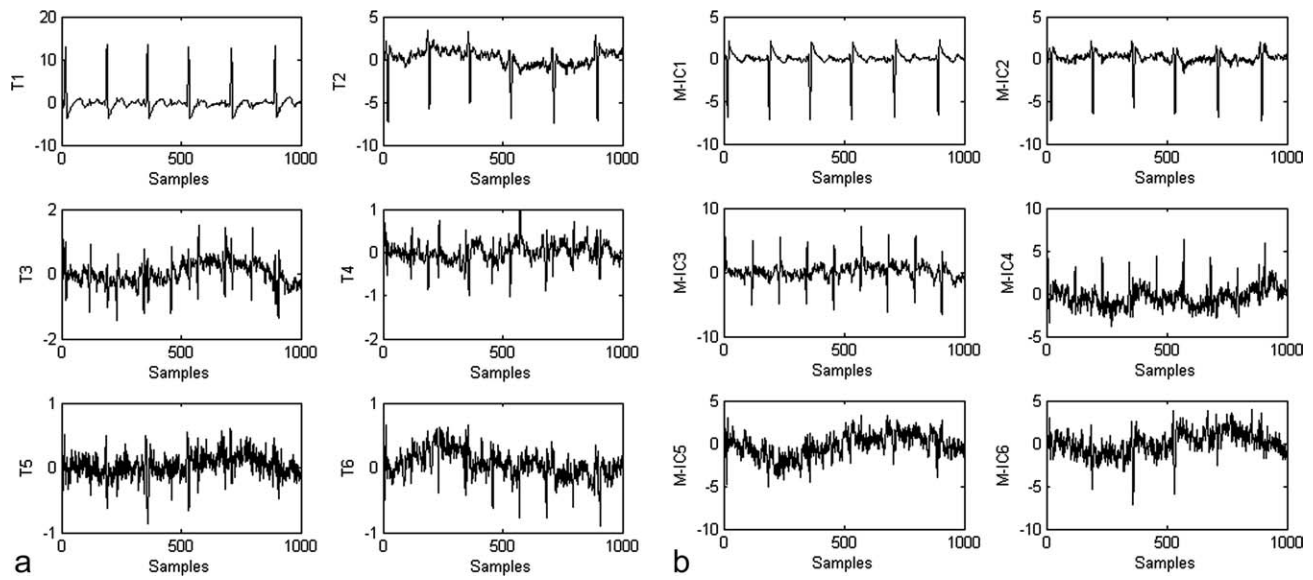

Figure 4. (a) LV estimates obtained through PLS, and (b) Source estimates obtained through M-ICR.

Table 4. Quality Prediction Results for Multichannel Cutaneous Potential Recordings

Method	Number of retained LVs					
	1	2	3	4	5	6
(a) PLS	0.5612	0.4307	0.1624	0.1294	0.1371	0.1365
	0.2389	0.1391	0.0661	0.0627	0.0309	0.0302
Original ICA	0.5003	0.3454	0.3614	0.2087	0.1329	0.1365
	0.2285	0.2153	0.2298	0.1609	0.0287	0.0302
Lee's M-ICA	0.4531	0.2933	0.1592	0.1658	0.1375	0.1365
	0.1734	0.0960	0.0923	0.0954	0.0290	0.0302
M-ICR ($\alpha = 0.8$, $\beta = 0.2$)	0.3864	0.2830	0.1532	0.1607	0.1325	0.1365
	0.1248	0.0887	0.0855	0.0894	0.0278	0.0302

Method	LVs	MSE
(b) PLS	4	0.1046
		0.0200
Lee's M-ICA	5	0.1075
		0.0179
M-ICR ($\alpha = 0.8$, $\beta = 0.2$)	4	0.0983
		0.0162

(a) Quality prediction MSEs with no bootstrap-based model validation.
(b) Quality prediction with bootstrap-based model validation.

recordings of test data. Channels 1–5 are abdominal signals; for channels 6–8 the electrodes have been placed further away from the fetus, on the thorax. Here, for the specific purpose of calibration and regression analysis, the measured potential values from channels 1 and 8 are employed as the response variables ($\mathbf{Y}_1$ and $\mathbf{Y}_2$), and the measurements from channels 2–7 served as descriptors ($\mathbf{X}_1 - \mathbf{X}_6$). Actually, the cutaneous recordings contain instantaneous linear mixtures of the aforementioned bioelectric signals, particularly the maternal electrocardiogram (MECG), and the fetal electrocardiogram (FECG). The instantaneous beat-to-beat FHR signal shows that channels 1 and 3 clearly contain weak fetal contributions. Due to the large amplitude of the MECG in the thoracic signals (channels 6–8), the FECG is less visible there.

Using PLS and the proposed method (with the initial significance factors α_{imp} and β_{imp}, set to be 0.8 and 0.2, respectively), LV model structures are derived from the 200 training samples. Using the resulting decomposition relationship, the underlying LVs of the 1,000 test samples are then calculated. The PLS-based source estimates are displayed in Figure 4a. Only two clear MECG-related components are obtained, as shown by the first two LVs. FECG signals may appear in LVs 3, 4, 5 and 6, but if so they are seriously contaminated by noise. In contrast, the LVs recovered through M-ICR are shown in Figure 4b. The two MECG signals and one FECG signal are accurately reconstructed, as shown by the first three ICs. Moreover, generally speaking, the ICs show better signal-to-noise ratios than the PLS LVs. The test results support Lathauwer's assertion[46] that ICA is a promising tool for estimating the underlying features of evoked potential recordings.

The direct quality prediction results using the different algorithms are first compared without bootstrap-based model validation in Table 4a. They are calculated using the 1,000 test samples and different numbers of retained LVs. With only two LVs, the proposed method yields lower MSE values than PLS. With more LVs, the PLS-based prediction results improve significantly. As the number of LVs increases beyond 5, the prediction results from all the algorithms generally are not further improved, but deteriorated. Bootstrapping is then employed for model validation and performance evaluation. Contribution indices of the ICs derived using the proposed method are calculated to the two quality variables, and are shown in Figure 5a, which gives a primary impression about the importance ranking of all the extracted ICs. The MECG and FECG contribute differently to the abdominal and thoracic quality signals as a result of their different locations placed on the pregnant woman's skin. The confidence interval applicable to the quality predictions based on bootstrapping is shown in Figure 5b taking example for 800 samples. Combined with the quality plots shown in Figure 3, the confidence profile apparently captures the evolving trends in the instantaneous beat-to-beat heart rates. After model validation, the final quality prediction results are summarized in Table 4b with different algorithms. For PLS and the proposed method, fewer LVs are required. In addition, the proposed M-ICR algorithm shows better quality prediction accuracy.

These experimental results show the promise of the proposed method for FECG analysis, although data limitations prevent assessing to what extent the extracted ICs might be useful in medical practice. Any practical application needs to be substantiated with further patient data.

Summary and discussion

This report has illustrated how the new regression strategy performs on both simulated and real data sets, and how it compares with PLS and conventional ICR. As a regularized regression estimation method, it is a more consistent and controllable option between ICR and PLS. It is hoped that this report will provide the basis for future work to determine under which circumstances each method or their combination is most appropriate. Future work might profitably take the following directions.

Use of the LV Structure Extracted from Quality Variables. As shown in the objective function of Eq. 7, the correlation between regressors and response is calculated and evaluated

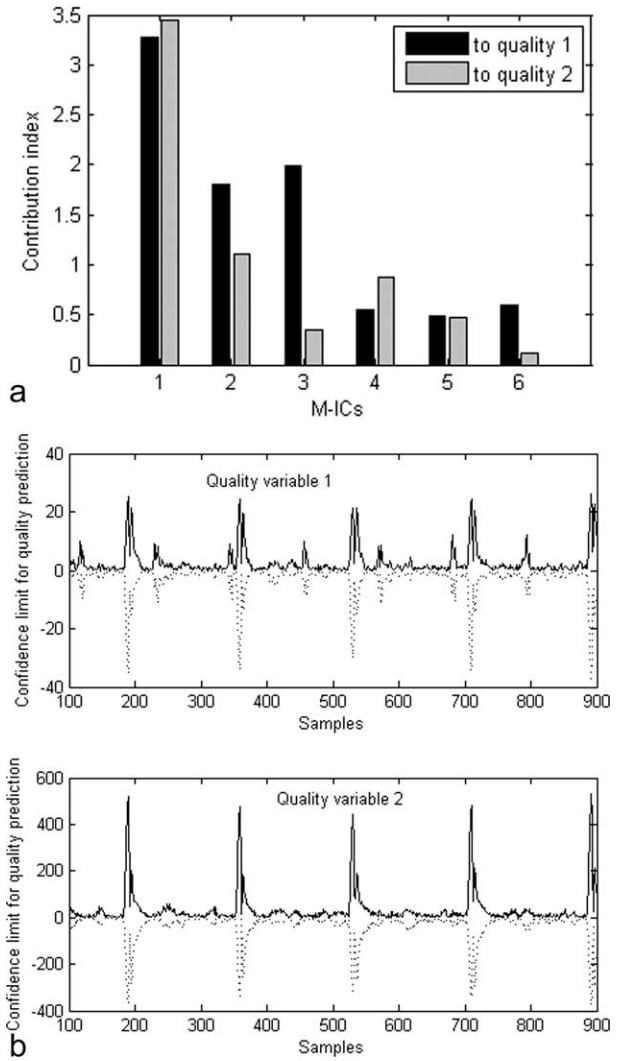

Figure 5. (a) Contribution indices of the ICs for model validation, and (b) confidence limits for the prediction of the two quality variables (solid line, the upper 95% confidence limit; dashed line, the lower 95% confidence limit).

directly using quality variables. This is rather different from PLS where the LV structure is recovered from multivariate response variables and employed in developing the correlation criterion. A natural question is whether or not efficiency might be gained in the M-ICR algorithm by using LVs instead of quality variables. It is clear that the use of quality LVs may result in more complex mathematical deduction and computation, but it may be necessary if the quality variables are contaminated by noise. On the other hand, when lacks dominant eigenvalues, which means the quality variables are not interrelated, then a correlation analysis using quality variables directly will come close to that obtained from quality LVs.

Setting of Suboptimization Weight Parameters. The effects of parameters on feature extraction and regression modeling have not yet been reduced to an explicit mathematical formula. The proper setting of parameter values has to be determined by cross-validation. Considering that the parameter values actually determine the balance between PLS and ICR involved in the regressor design, is it possible to work out some simple rule to follow? This work may be especially fruitful.

Pursuit of Global Optimization. The modified ICR algorithm is derived from the conventional FastICA method, which is a constrained optimization which makes some assumptions to simplify the computation. How do they influence the optimization result? Moreover, FastICA is likely to be trapped at local optima using the gradient-based algorithm. How to better solve multiobjective mathematical optimizations and get global optima remains an open question. Answering it may help to further improve LV modeling and benefit subsequent regression analysis.

Solving Nonlinear Problems. For reasons of computational and conceptual simplicity, the LV representation is a linear transformation of the original data. In complex reality, chemical measurements, for example, may be a nonlinear function of some real LVs. A nonlinear version of LV modeling is probably required to take into account nonlinear regression relationships between regressors and responses.

There are still many issues to be investigated in future, but the results of this study constitute a step forward toward improving LV extraction and regression modeling for practical applications.

Conclusions

In this study, an improved ICR modeling algorithm is formulated. It differs from conventional ICR methods in the way it extracts underlying explanatory sources for quality description. In the general formulation, 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 statistical independence of ICA and the quality-related measures of PLS are combined in the projection pursuit ''indices''. The major results of the study are as follows:

1. The proposed method yields more meaningful LVs better correlated with quality and with higher-order independence. This may indicate a parsimonious model structure compared with conventional ICA, achieved by automatically sorting out quality-relevant regressors.

2. Bootstrap-based quantitative analysis can verify the modeling results and evaluate modeling performance more convincingly.

3. Both numerical and realistic simulations have shown the effectiveness of the proposed method. M-ICR is shown to yield LVs that are chemically interpretable and quality predictions comparable with those from PLS.

Practitioners have used various LV estimation methods arising from their understanding of the LV structure of measured data, but there has been little agreement about which LV structure is better. The proposed method, as a regularized regression estimation tool, provides some way to combine PLS and ICR to derive a suitable LV structure for specific practical applications. It is a more consistent and controllable extension of ICR and PLS by changing certain parameters in the general objective function, and the balance is purely data driven. Although the proposed method might not yet be able to derive absolutely optimal features for

calibration modeling and quality prediction in every case, it shows desirable improvement by encapsulating both quality-related and higher-order independence characteristics into the feature extraction. Considerable further research, is, thus, recommended.

Acknowledgments

The authors would like to acknowledge the anonymous reviewers for their helpful and instructive comments, which are really of great benefit. This work is supported by China's National 973 program (2009CB320603) and the National Natural Science Foundation of China (No. 60774068).

Literature Cited

1. Wang XZ. *Data Mining and Knowledge Discovery for Process Monitoring and Control*. London: Springer; 1999.
2. Wang XZ, Li RF. Combining conceptual clustering and principal component analysis for state space based process monitoring. *Ind Eng Chem Res*. 1999;38:4345–4358.
3. Martin EB, Morris AJ. Enhanced bio-manufacturing through advanced multivariate statistical technologies. *J Biotechnol*. 2002; 99:223–235.
4. Chen FZ, Wang XZ. Discovery of operational spaces from process data for production of multiple grades of products. *Ind Eng Chem Res*. 2000;39:2378–2383.
5. Qin SJ. Statistical process monitoring: Basics and beyond. *J Chemomet*. 2003;17:480–502.
6. Cheng C, Chiu MS. A new data-based methodology for nonlinear process modeling. *Chem Eng Sci*. 2004;59:2801–2810.
7. Wong CWL, Escott R, Martin EB, Morris AJ. The integration of spectroscopic and process data for enhanced process performance monitoring. *Can J Chem Eng*. 2008;86:905–923.
8. Yu J, Qin SJ. Statistical MIMO controller performance monitoring. Part I: Data-driven covariance benchmark. *J Process Control*. 2008;18:277–296.
9. Martens H, Naes T. *Multivariate Calibration. 2nd ed*. Chichester: Wiley; 1994.
10. Brereton RG. Introduction to multivariate calibration in analytical chemistry. *Analyst*. 2000;125:2125–2154.
11. Li B, Hassel PA, Martin EB, Morris AJ. A non-linear nested PLS algorithm. *Comput Stat Data Anal*. 2003;48:87–101.
12. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied Regression Analysis and Other Multivariable Methods. 3rd ed*. Beijing: China Machine Press; 2003.
13. Trygg J. Prediction and spectral profile estimation in multivariate calibration. *J Chemomet*. 2004;18:166–172.
14. Li B, Morris AJ, Martin EB. Generalized partial least squares regression based on the penalized minimum norm projection. *Chemomet Intell Lab Syst*. 2004;72:21–26.
15. Ergon R. Reduced PCR/PLSR models by subspace projections. *Chemomet Intell Lab Syst*. 2006;81:68–73.
16. Jönsson S, Eriksson LA, van Bavel B. Multivariate characterisation and quantitative structure-property relationship modeling of nitroaromatic compounds. *Anal Chim Acta*. 2008;621:155–162.
17. Gustafsson MG. Independent component analysis yields chemically interpretable latent variables in multivariate regression. *J Chem Inf Model*. 2005;45:1244–1255.
18. Hyvärinen A, Oja E. Fast fixed-point algorithm for independent component analysis. *Neural Comput*. 1997;9:1483–1492.
19. Hyvärinen A, Oja E. Independent component analysis: Algorithms and applications. *Neural Netw*. 2000;13:411–430.
20. Girolami M. *Self-Organising Neural Networks: Independent Component Analysis and Blind Source Separation*. London: Springer; 1999.
21. Li RF, Wang XZ. Dimension reduction of process dynamic trends using independent component analysis. *Comput Chem Eng*. 2002; 26:467–473.
22. Albazzaz H, Wang XZ. Statistical process control charts for batch operations based on independent component analysis. *Ind Eng Chem Res*. 2004;43:6731–6741.
23. Kano M, Tanaka S, Hasebe S, Hashimoto I, Ohno H. Monitoring independent components for fault detection. *AIChE J*. 2003;49:969–976.
24. Lee JM, Yoo CK, Lee IB. Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chem Eng Sci*. 2004;59:2995–3006.
25. Zhao CH, Wang FL, Mao ZZ, Lu NY, Jia MX. Adaptive monitoring based on independent component analysis for multiphase batch processes with limited modeling data. *Ind Eng Chem Res*. 2008; 47:3104–3113.
26. Liu XQ, Lei X, Kruger U, Littler T, Wang SQ. Statistical-based monitoring of multivariate non-Gaussian systems. *AIChE J*. 2008; 54:2379–2391.
27. Kermit M, Tomic O. Independent component analysis applied on gas sensor array measurement data. *IEEE Sensor J*. 2003;3:218–228.
28. Ahn H, Choi E, Han I. Extracting underlying meaningful features and canceling noise using independent component analysis for direct marketing. *Expert Syst Appl*. 2007;33:181–191.
29. Chen J, Wang XZ. A new approach to near-infrared spectral data analysis using independent component analysis. *J Chem Inf Comput Sci*. 2001;41:992–1001.
30. Westad F. Independent component analysis and regression applied on sensory data. *J. Chemomet*. 2005;19:171–179.
31. Shao XG, Wang W, Hou ZY, Cai WS. A new regression method based on independent component analysis. *Talanta*. 2006;69:676–680.
32. Kaneko H, Arakawa M, Funatsu K. Development of a new regression analysis method using independent component analysis. *J Chem Inf Model*. 2008;48:534–541.
33. Lee J, Qin SJ, Lee I. Fault detection and diagnosis based on modified independent component analysis. *AIChE J*. 2006;52:3501–3514.
34. Luenberger DG. *Optimization by Vector Space Methods*. New York: Wiley; 1969.
35. Lu XF. *Application Basis of an Optimization Method*. Shanghai: Tongji University Press; 2003.
36. Westad F, Kermit M. Cross validation and uncertainty estimates in independent component analysis. *Analytica Chimica Acta*. 2003;490: 341–354.
37. Jackson JE. *A User's Guide to Principal Components*. New York: Wiley; 1991.
38. Zhang J. Inferential estimation of polymer quality using bootstrap aggregated neural networks. *Neural Networks*. 1999;12:927–928.
39. Zhang J. Developing robust non-linear models through bootstrap aggregated neural networks. *Neurocomputing*. 1999;25:93–113.
40. Zhang J, Morris AJ, Martin EB, Kiparissies C. Prediction of polymer quality in batch polymerization reactors using robust neural networks. *Chem Eng J*. 1998;69:135–143.
41. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.
42. Wehrens R, Putter H, Buydens LMC. The bootstrap: A tutorial, *Chemomet Intell Lab Syst*. 2000;54:35–52.
43. Centner V, Massart D, Noord OED, Jong SD, Vandeginste BM, Sterna C. Elimination of uninformative variables for multivariate calibration. *Anal Chem*. 1996;68:3851–3858.
44. Galvão RKH, Araújo MCU, Fragoso WD, Silva EC, José GE, Soares SFC, Paiva HM. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. *Chemomet Intell Lab Syst*. 2008;92:83–91.
45. Ye SF, Wang D, Min SG. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection. *Chemomet Intell Lab Syst*. 2008;91:194–199.
46. Lathauwer LD, Moor BD, Vandewalle J. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Trans Biomed Eng*. 2000;47:567–572.

Appendix A

This appendix will show how to mathematically formulate the basic search optimization solution of the proposed M-ICR algorithm.

Using a Lagrange operator, the initial objective function can be expressed as an unconstrained extremum problem

$$F(\mathbf{w},\lambda)=\alpha\cdot\frac{1}{J_y}\left(E\left(\mathbf{w}^T\mathbf{zy}\right)^T E\left(\mathbf{w}^T\mathbf{zy}\right)\right)+\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)^2+\lambda(\mathbf{w}^T\mathbf{w}-1) \quad (A1)$$

where λ is a constant scalar.

Calculating the derivatives of with respect to and and setting both equal to zero, the following equations can be obtained

$$\nabla F_{\mathbf{w}}=\frac{\partial F}{\partial \mathbf{w}}=2\alpha\cdot\frac{1}{J_y}E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}+2\lambda\mathbf{w}=0 \quad (A2)$$

$$\nabla F_{\lambda}=\mathbf{w}^T\mathbf{w}-1=0 \quad (A3)$$

where g is the first-order derivative of G.

Multiply both sides of Eq. A2 by $\mathbf{w}^T$

$$2\alpha\times\frac{1}{J_y}\mathbf{w}^T E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)\times\mathbf{w}^T E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}+2\lambda\mathbf{w}^T\mathbf{w}=0 \quad (A4)$$

According to Eq. A4, can be specifically calculated as

$$\lambda=-\alpha\cdot\frac{1}{J_y}\mathbf{w}^T E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)-\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)\mathbf{w}^T E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} \quad (A5)$$

Based on Kahn-Tucker conditions,[34] it is known that the optima of the objective function shown in Eq. 7 under the constraint $\mathbf{w}^T\mathbf{w}=1$ will be at the points shown in Eq. A2. Here, to improve the convergence speed, Newton's method is adopted to solve this problem, which makes use of the second-order derivative of the objective function.

Denoting the function on the lefthand side of Eq. 10 by Ψ, its Jacobian matrix $J\Psi(\mathbf{w})$ is

$$J\Psi(\mathbf{w})=2\alpha\cdot\frac{1}{J_y}E\left(\mathbf{zy}^T\right)E\left(\mathbf{yz}^T\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)E\{\mathbf{zz}^T g'(\mathbf{w}^T\mathbf{z})\}+2\beta\cdot E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\left(E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\right)^T+2\lambda\mathbf{I} \quad (A6)$$

Substituting Eq. A5 into Eq. A6, the following expression for the Jacobian matrix results

$$J\Psi(\mathbf{w})=2\alpha\cdot\frac{1}{J_y}E\left(\mathbf{zy}^T\right)E\left(\mathbf{yz}^T\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)E\{\mathbf{zz}^T g'(\mathbf{w}^T\mathbf{z})\}+2\beta\cdot E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\left(E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\right)^T-\left(2\alpha\cdot\frac{1}{J_y}\mathbf{w}^T E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)\mathbf{w}^T E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\right)\mathbf{I} \quad (A7)$$

To simplify the inversion of this matrix, the second term can be approximated as in the original FastICA algorithm, since the data have been whitened

$$E\{\mathbf{zz}^T g'(\mathbf{w}^T\mathbf{z})\}\approx E\{\mathbf{zz}^T\}E\{g'(\mathbf{w}^T\mathbf{z})\}=E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{I} \quad (A8)$$

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

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

$$J\Psi(\mathbf{w})=2\alpha\times\frac{1}{J_y}E\left(\mathbf{zy}^T\right)E\left(\mathbf{yz}^T\right)+2\beta E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\left(E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\right)^T-c\cdot\mathbf{I} \quad (A9)$$

where c is a constant, which is defined as follows

$$c=2\alpha\cdot\frac{1}{J_y}\mathbf{w}^T E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)\mathbf{w}^T E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}-2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)E\{g'(\mathbf{w}^T\mathbf{z})\} \quad (A10)$$

Substituting Eq. A5 into Eq. A2, the derivative of with respect to $F(\mathrm{w},\lambda)$ can then be expressed as

$$\nabla F_{\mathbf{w}}=2\alpha\cdot\frac{1}{J_y}E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)+2\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}-2\left(\alpha\cdot\frac{1}{J_y}\mathbf{w}^T E\left(\mathbf{zy}^T\right)E\left(\mathbf{w}^T\mathbf{zy}\right)+\beta\cdot\left(E\left\{G(\mathbf{w}^T\mathbf{z})\right\}-E\{G(\upsilon)\}\right)\mathbf{w}^T E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\}\right)\mathbf{w} \quad (A11)$$

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

$$J\Psi(\mathbf{w})\cdot\mathbf{d}_n=-\nabla F_{\mathbf{w}} \quad (A12)$$

The solution, therefore, depends heavily on whether or not the Jacobian matrix is invertible. When it is not, some strategy should be adopted to modify it and make it invertible. An approximate Newton iteration direction can, thus, be calculated. The detailed analysis of the invertibility of the Jacobian matrix is shown in Appendix B, as well as the corresponding regulation when invertibility cannot be achieved.

The Newton method loses its convergence ability when the search region is far from the optima of the objective function, so a one-dimensional (1-D) 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 the function $F(\cdot)$ is shown in Eq. A1. Adopting a 1-D searching strategy will give the method global convergence ability. If $\gamma=1$, it converges to the conventional Newton iteration method.

Finally, $\mathbf{w}$ is derived as follows

$$\mathbf{w}\leftarrow\mathbf{w}+\gamma\cdot\mathbf{d}_n \quad \text{i.e., } \mathbf{w}\leftarrow\mathbf{w}-\gamma\cdot\left(J\Psi(\mathbf{w})_m\right)^{-1}\nabla F_{\mathbf{w}} \quad (A13)$$

where $J\Psi(\mathbf{w})_m$ is the modified invertible Jacobian matrix as shown in Appendix B, and $\nabla F_{\mathbf{w}}$ is the first-order derivative of $F(\mathbf{w},\gamma)$, with respect to $\mathbf{w}$ as shown in Eq. A11.

Appendix B

Here the conditions under which the Jacobian matrix shown in Eq. A9 can be generally inverted are discussed and analyzed. An adjustment strategy is also discussed for when the conditions cannot be satisfied.

First, for a $(d \times J_y)$-dimensional symmetric matrix $2\alpha \cdot \frac{1}{J_y}\mathrm{E}(\mathbf{z}\mathbf{y}^{\mathrm{T}})\mathrm{E}(\mathbf{y}\mathbf{z}^{\mathrm{T}}) + 2\beta \cdot \mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}(\mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\})^{\mathrm{T}}$ (here for simplicity, assume that $J_y \ll d$), assume its rank is q. This means there are only q positive eigenvalues computable by direct singular value decomposition, which is delivered as follows

$$2\alpha \cdot \frac{1}{J_y}\mathrm{E}(\mathbf{z}\mathbf{y}^{\mathrm{T}})\mathrm{E}(\mathbf{y}\mathbf{z}^{\mathrm{T}}) + 2\beta \cdot \mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}(\mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\})^{\mathrm{T}} = \mathbf{L}\begin{bmatrix}\Delta & \mathbf{0}_{q\times(d-q)} \\ \mathbf{0}_{(d-q)\times q} & \mathbf{0}_{d-q}\end{bmatrix}\mathbf{L}^{\mathrm{T}} \tag{B1}$$

where the orthogonal matrix $\mathbf{L}(d \times d)$ satisfies $\mathbf{L}\mathbf{L}^{\mathrm{T}} = \mathbf{I}(d \times d)$ and $\mathbf{L}^{-1} = \mathbf{L}^{-\mathrm{T}}$. The diagonal matrix $\Delta(q \times q)$ employs the positive eigenvalues of $2\alpha \times \frac{1}{J_y}\mathrm{E}(\mathbf{z}\mathbf{y}^{\mathrm{T}})\mathrm{E}(\mathbf{y}\mathbf{z}^{\mathrm{T}}) + 2\beta \times \mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}(\mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\})^{\mathrm{T}}$ as its elements. Then the following relationship can be derived

$$\begin{aligned}
&2\alpha \times \frac{1}{J_y}\mathrm{E}(\mathbf{z}\mathbf{y}^{\mathrm{T}})\mathrm{E}(\mathbf{y}\mathbf{z}^{\mathrm{T}}) + 2\beta \times \mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\}(\mathrm{E}\{\mathbf{z}\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{z})\})^{\mathrm{T}} - c \times \mathbf{I} \\
&= \mathbf{L}\begin{bmatrix}\Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{L}^{\mathrm{T}} - c \cdot \mathbf{L}\mathbf{L}^{\mathrm{T}} \\
&= \mathbf{L}\left(\begin{bmatrix}\Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} - c \cdot \mathbf{I}\right)\mathbf{L}^{\mathrm{T}} \\
&= \mathbf{L}\begin{bmatrix}\Delta - c \cdot \mathbf{I}_q & \mathbf{0}_{q\times(d-q)} \\ \mathbf{0}_{(d-q)\times q} & -c \cdot \mathbf{I}_{d-q}\end{bmatrix}\mathbf{L}^{\mathrm{T}} \\
&= \mathbf{L}\mathbf{U}\mathbf{L}^{\mathrm{T}}
\end{aligned} \tag{B2}$$

where $\mathbf{U} = \begin{bmatrix}\Delta - c \cdot \mathbf{I}_q & \mathbf{0}_{q\times(d-q)} \\ \mathbf{0}_{(d-q)\times q} & -c \cdot \mathbf{I}_{d-q}\end{bmatrix}$ is diagonal.

From this the inverse of the Jacobian matrix can be calculated as

$$(\mathrm{J}\Psi(\mathbf{w}))^{-1} = \mathbf{L}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{L} \tag{B3}$$

However, when the first matrix block in U has a zero diagonal element, the Jacobian matrix is not invertible. Some strategy should thus be adopted to modify it so as to get a feasible Newton iteration direction. Here, the invertibility of the Jacobian matrix is judged continuously, and if it cannot be satisfied, an extra positive constant term is added to the diagonal to make it nonzero. The modified Jacobian matrix then becomes

$$\mathrm{J}\Psi(\mathbf{w})_m = \mathrm{J}\Psi(\mathbf{w}) + \mu \cdot \mathbf{I} \tag{B4}$$

where μ is the selected compensatory constant.

Appendix C

This appendix will discuss how to reveal the LVs on the basis of the modified ICR algorithm.

The initialized demixing matrix is set

$$\mathbf{W}(n \times d) = [\mathbf{I}_n \vdots \mathbf{0}] \tag{C4}$$

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

The detailed procedure is as follows:

(a) Determine the number of independent components n, and set the counter $p = 1$.

(b) Set, $\mathbf{w}_p$ the pth row of $\mathbf{W}$, as the initial demixing vector.

(c) Let $\mathbf{w} \leftarrow \mathbf{w} - \gamma \cdot (\mathrm{J}\Psi(\mathbf{w})_m)^{-1} \bigtriangledown \mathrm{F}_{\mathbf{w}}$. This step is an approximative Newton iteration procedure in pursuit of the maximization of the dual-objective optimization problem shown in Eq. 7.

(d) Orthogonalization and normalization: $\mathbf{w}_p = \mathbf{w}_p - \sum_{j=1}^{p-1} \mathbf{w}_p^T\mathbf{w}_j\mathbf{w}_j$, $\mathbf{w}_p = \mathbf{w}_p / \|\mathbf{w}_p\|$.

(e) If $\mathbf{w}_p$ does not converge (where 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).

(f) Set $p = p + 1$. If $p \leq n$, return to step (b) until all the desired ICs have been retained. The updated demixing matrix is then $\mathbf{W}(n \times d)$.

Manuscript received Feb. 7, 2009, and revision received Aug. 13, 2009.