Multi-component Quantification Algorithms of the FTIR Method Engine Exhaust Gas Analyzer

Kaori INOUE      Masayuki ADACHI

Abstract
The effective functioning of FTIR multi-component gas analyzers depends, in part, on the selection and application of appropriate quantification algorithms. The quantification algorithms of MEXA-FT engine exhaust analyzers have been optimized for measurement of exhaust having complex composition. Customized and installed in each instrument is a quantification matrix, in the form of a kind of calibration curve, based on the reference spectral data. By collecting the reference spectra for each instrument and optimizing the algorithms we assure that the MEXA-FT series of instruments operate with high-performance and high reliability.
1. Introduction

The MEXA-FT series gas analyzers employ the Fourier transform infrared (FTIR) method to perform continuous analysis of the multiple components of engine exhaust. The FTIR method yields quantitative data by using the shapes of infrared absorption spectra and the relationship of absorption intensities and components’ concentrations. Spectral data from sample gas containing many components can be treated as multivariate data and multivariate calibration algorithms can be applied to predict the concentrations of multiple components. This paper describes representative algorithms used for the FTIR method and their relationship to instrument performance.

2. Algorithms of multivariate calibration methods

An absorption spectrum obtained by FTIR may be regarded as a vector or a matrix because it is a set of digital data. In the following sections, boldface lower case characters (e.g., \(a\)) are used for vectors, and boldface upper case letters (e.g., \(A\)) are used for matrices. The inverse of matrix \(A\) is presented as \(A^{-1}\), a transposed matrix is presented as \(A^T\) and a matrix predicted using a multivariate model is presented as \(\hat{A}\). In this paper we cover the concepts of algorithms and leave detailed explanations to other literature.

2.1 Classical Least Squares (CLS) method

2.1.1 Concept of the CLS model

The classical least squares (CLS) method is a multiple regression analysis technique used to solve a simple equation as expressed in Eq. 1.

\[
y = Xb + f \quad \cdots \cdots (1)
\]

Here, \(f\) stands for the residual error included in the model, and \(b\) shall be determined so as to minimize \(f\). Solving for \(b\) is expressed in Eq. 2.

\[
b = \hat{X}^T y \quad \cdots \cdots (2)
\]

\(\hat{X}\) is called the general inverse matrix of Moore-Penrose, and is equal to \((X^TX)^{-1}X^T\).

2.1.2 Application to analysis of absorption spectra

When spectra of sample gases containing several components are given, the absorbance \(a_{ij}\) at wavenumber \(j\) (\(j = 1, 2, ..., n\)) in spectrum \(i\) (\(i = 1, 2, ..., m\)) is regarded as the sum of absorbance at unit concentration of a component \(l\) (\(l = 1, ..., d\)), \(kl_{ij}\), multiplied by its concentration, \(c_{il}\). In this case, the CLS model would be expressed as Eq. 3. \(A\) is an \(mn\) matrix that consists of absorbance data of the spectra; \(C\) is an \(md\) matrix of concentrations of components; \(K\) is a \(dn\) matrix of pure-component spectra at unit concentration, and \(E\) is a matrix of errors at each wavenumber of matrix \(A\).

\[
A = CK + E \quad \cdots \cdots (3)
\]

When the spectral data in \(A\) and the concentration values in \(C\) are already known, \(\hat{K}\) can be presumed as shown in Eq. 4.

\[
\hat{K} = CA \quad \cdots \cdots (4)
\]

The process of determining \(\hat{K}\) corresponds to calibration for quantitative analyses, and the spectra in \(A\) are called calibration spectra. To predict the components’ concentration in an unknown spectrum, the inverse matrix of \(\hat{K}\) (i.e., \(\hat{K}^{-1}\)) is used as shown in Eq. 5. Here, \(a_{\text{unk}}\) represents a sample spectrum and \(c_{\text{unk}}\) represents a set of concentrations of components included in the spectrum.

\[
c_{\text{unk}}^T = a_{\text{unk}}^T \hat{K}^{-1} \quad \cdots \cdots (5)
\]

Fig. 1 shows an outline of the CLS model. In the CLS model, \(\hat{K}\) corresponds to spectra for pure components at unit concentrations, which makes the model very easy to understand. Since absorbances in \(A\) are expressed as sums of concentrations of components, building the model requires the concentrations of all of the existing components in the spectra. Thus, in some cases, the model might be incorrect due to unknown impurities in the calibration spectra. Furthermore, the CLS model cannot account for nonlinearities in relationships of the concentrations and the absorbances. Because of this, when the model is applied to components that produce such nonlinearities, the linearity of the predicted concentrations must be adjusted using additional computation.

2.2 Principal Component Regression (PCR) method

2.2.1 Concept of the PCR model

By using the principal component regression (PCR) method, multivariate data can be expressed as a simpler
model by extracting dominant factors that sufficiently explain data tendencies. Such factors in the PCR model are represented as loading vectors. Specifically, a loading vector \( p \) of an \( mn \) matrix \( X \) can be calculated as an eigen vector. Generally, the number of loading vectors that are necessary to build a adequate model is smaller than that of the included components in the sample gas. Using an \( nh \) matrix \( P \) that consists of loading vectors \( p_j \) (\( j=1, ..., h \)), \( X \) could be converted as shown in Eq. 6.

\[
XP = T \quad \cdots \cdots (6)
\]

The mn matrix \( T \) calculated here is called a score matrix. In solving Eq. 6, \( X \) is expressed as a product of two small matrices as Eq. 7.

\[
X=TP^T = TP^T \quad \cdots \cdots (7)
\]

The nonlinear iterative partial least squares (NIPALS) algorithm is recognized as the fastest and most accurate method for obtaining the matrix \( P \) on a computer. Figure 2 shows a schematic flow of the NIPALS algorithm.

- **2.2.2 Application to analysis of absorption spectra**
  
  A matrix of calibration spectra in the PCR model, \( A \), is represented with an \( mh \) matrix \( T \) and an \( nh \) matrix \( P \) as in Eq. 8.

\[
A=TP^T+E \quad \cdots \cdots (8)
\]

Ignoring the residual matrix \( E \) and introducing the model of \( A=CP^T \) that is used in the CLS algorithm, Eq. 8 can be modified to produce Eq. 9.

\[
C=TP^T \hat{K} \quad \cdots \cdots (9)
\]

By defining a new matrix \( G \) that is equal to matrix \( P^T \hat{K} \), Eq. 9 can be modified to produce Eq. 10.

\[
\hat{G}=TC \quad \cdots \cdots (10)
\]

In cases of the PCR model, calculation of \( P \) and \( G \) from calibration spectra \( A \) and concentration values \( C \) corresponds to the calibration process. Eq. 11 is used for the prediction process.

\[
c_{\text{unk}} = a_{\text{unk}}^T \hat{G} \quad \cdots \cdots (11)
\]

Figure 3 shows an outline of the PCR model. Different from the CLS, the PCR technique allows us to construct a model without determining concentrations of all of the components in the calibration spectra. Furthermore, because PCR allows modeling of nonlinearities, using PCR makes it possible to directly obtain linear outputs, even for components that show nonlinear absorption. Appropriate numbers of loading vectors shall be determined based on eigen values or by use of an algorithm called cross variation.

- **2.3 Partial Least Squares (PLS) method**
  
  **2.3.1 Concept of the PLS model**

  The partial least squares (PLS) method uses a similar algorithm to that used in PCR. One of the differences between the methods is that dominant factors, ie. loading vectors, are extracted using two matrices (e.g., \( X \) and \( Y \)). PCR, on the other hand, uses only one matrix (e.g., \( X \)). Matrices and coefficients required in the PLS model can be calculated using an algorithm similar to the NIPALS method used in the PCR model.
2.3.2 Application to analysis of absorption spectra

In the calibration process used in PLS, loading vectors $p_k$ and $q_k$, weighted vectors $w_k$, and coefficients $b_k$ (k=1, ..., r, for all) are calculated from two matrices, i.e., calibration spectra $A$ and concentration values $C$. In the prediction process, an unknown spectrum $a_{unk}$ is decomposed into scores $t_k$ (k=1, ..., r) so as to reconstruct a concentration vector $c_{unk}$ as a predicted result.

In both the PLS and PCR methods, concentrations of target components are only required for all existing components in the calibration spectra. The nonlinearity of absorption spectra can also be included into the model. Additionally, since both the information of absorption and concentration are used in the algorithm, a more appropriate model for prediction of concentrations can be established. The most suitable number of loading vectors can be confirmed using the same method as used in PCR.

2.4 Extended PCR (e-PCR) and extended PLS (e-PLS)

Extended PCR (e-PCR) and extended PLS (e-PLS) were originally developed by Horiba in order to predict more accurate concentration values from FTIR absorption spectra. Only the e-PCR algorithm is discussed in this paper; the e-PLS algorithm is analogous and is not described. In the calibration process for e-PCR, the total intensity of absorption $A_e$, i.e., the sum of absorbances at each wavenumber in an absorption band, is introduced in place of the concentration matrix $C$. The calibration spectra $A$ can be represented as shown in Eq. 12 with the pure-component spectra $K_e$ calculated as the per-unit absorbance intensity.

$$A = A_e K_e + E \quad \cdots \cdots (12)$$

For this model, a matrix $\hat{G}_e$ that corresponds to the matrix $G$ in Eq. 10 can be calculated as shown in Eq. 13.

$$\hat{G}_e = T + A_e \quad \cdots \cdots (13)$$

In the prediction process, a vector that consists of absorption intensities caused by target components in an unknown spectrum, $a_{unk}$, is calculated as Eq. 14.

$$a_{unk}^T = a_{unk}^T P G_e \quad \cdots \cdots (14)$$

As the next step, $a_{unk}$ shall be converted into a concentration vector $c_{unk}$ using approximate equations, $c_k = f(a_{unk})$ (k=1, ..., d), that are predetermined for the target components. Compared to the concentration values used in the PCR model, the total intensities of absorption newly introduced into the e-PCR show excellent linearities with the absorbances at individual wavenumbers in the calibration spectra. This is because the change of the total intensity of absorption corresponds to that of the averaged value of absorbances at each wavenumber, while the absorbances do not always linearly change with the components’ concentrations. By excluding, in advance, the nonlinearities from the model (as above), the factors which represent the characteristics of the spectra of each component can be effectively extracted by the e-PCR method. This is the most important advantage of e-PCR and e-PLS, and makes it possible to obtain more accurate concentrations than from the ordinary PCR and PLS. The e-PCR and e-PLS methods are very useful for continuous analyses of multiple components when they are calculated on high-speed and high-accuracy equipment. The techniques are especially effective for the analyses of nonlinear components such as carbon monoxide. Figure 4 shows the comparison of the PCR and the e-PCR, as zero noise of carbon monoxide output.

![Comparison of the PCR method and the expanded-PCR method](image)

3. The construction of spectral data base

One of common characteristics of the multivariate algorithm described above is that calibration matrices are predetermined using calibration spectra at which concentrations of target components are already known. Here, in order to effectively eliminate the effect of nonlinearity and errors of the spectra, it is essential to use sufficient numbers of calibration spectra for the components of interest. An automatic spectral data collecting system has been established in order to save both time and work in preparing such large numbers of spectra.
3.1 Automatic collection of spectral data

Figure 5 shows a configuration of the automatic spectral data collection system. Up to eight cylinders of standard gases can be connected to a gas divider through a solenoid valve unit. Both the gas lines and an FTIR analyzer are controlled by a computer, and spectra of components at various concentrations are continuously and automatically collected. This system has made it possible to collect actual spectra for all of our analyzer products.

3.2 Making and installing the calibration matrix

The collected spectra are sent to another computer and stored in a spectra database for each analyzer. Next, calibration matrices are calculated using the most suitable algorithm. By calibrating in this manner, using actual spectra, we reduce to a negligible level the effects of variations of the spectral shapes arising from instruments. As a result, high accuracy of the multivariate analysis can be achieved. For practical reasons the calibration matrices are installed only in the computer that controls the FTIR analyzer.

4. Conclusion

In this paper we described the multivariate algorithms used in the multi-component analysis performed by the MEXA-FT series. The most suitable algorithms differ depending upon the sample gases and target components to be analyzed. The reliability of concentration outputs may also be affected by the calibration spectra, and the wavenumbers used for measurement. Both the algorithms for quantitative analysis and the spectral database are extremely important elements the MEXA-FT series gas analyzers. Adequate consideration for them results in the high performance that is characteristic of the MEXA-FT series.

References
1) Adachi, M., et al., SAE Transaction No. 920723