

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231169146>

Nonlinear multivariate calibration using principal components regression and artificial neural networks

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 1991

Impact Factor: 5.64 · DOI: 10.1021/ac00020a022

CITATIONS

161

READS

37

3 AUTHORS, INCLUDING:



Paul Gemperline

East Carolina University

85 PUBLICATIONS 2,253 CITATIONS

SEE PROFILE



Vasilis Gregoriou

National Hellenic Research Foundation

71 PUBLICATIONS 1,345 CITATIONS

SEE PROFILE

- (10) Meinschen, W. G.; Rinaldi, G. G. L.; Hayes, J. M.; Scholler, D. A. *Biomed. Mass Spectrom.* **1974**, *1*, 172-174.
- (11) Schmid, E. R.; Grundmann, H.; Fogy, I.; Papesch, W.; Rank, D. *Biomed. Mass Spectrom.* **1981**, *8*, 486-499.
- (12) Deniro, M. J.; Epstein, S. *Science* **1977**, *197*, 261-263.
- (13) Monson, K. D.; Hayes, J. M. *J. Biol. Chem.* **1980**, *255*, 11435-11441; *Geochim. Cosmochim. Acta* **1982**, *46*, 139-149.
- (14) Rossmann, A.; Schmidt, H. L. *Z. Lebensm.-Unters.-Forsch.* **1989**, *188*, 434-438.
- (15) Bengsch, E.; Grivet, J. P.; Schulten, H. R. *Z. Naturforsch.* **1981**, *36B*, 1289-1298.
- (16) Freeman, K. H.; Hayes, J. M.; Trendl, J. M.; Albrecht, P. *Nature* **1990**, *343*, 254-256.
- (17) Craig, H. *Geochim. Cosmochim. Acta* **1953**, 53-92; *Science* **1961**, *133*, 1833-1834.
- (18) Hagemann, R.; Nief, G.; Roth, E. *Tellus* **1970**, *22*, 712-715.
- (19) Martin, M. L.; Delpuech, J. J.; Martin, G. J. *Practical NMR Spectroscopy*; Wiley-Heyden: London-Philadelphia-Rheine, **1980**; Chapter 9.
- (20) Bengsch, E. French Patent FR 2530026, July 9, 1982.
- (21) Smith, B. N. *Naturwissenschaften* **1976**, *62*, 390-390.
- (22) Bricout, J.; Fontes, J. C. *Ann. Falsif. Expert. Chim.* **1974**, 211-215.
- (23) Dunbar, J. *Fresenius' Z. Anal. Chem.* **1982**, *311*, 578-580.
- (24) Moussa, I.; Naulet, N.; Martin, M. L.; Martin, G. J. *J. Phys. Chem.* **1990**, *94*, 8303-8309.
- (25) Simon, H.; Rauschenbach, P.; Frey, A. *Z. Lebensm.-Unters.-Forsch.* **1988**, *136*, 279-284.
- (26) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. J. *Magn. Res.* **1983**, *52*, 335-338.
- (27) Guillou, C.; Trierweiler, M.; Martin, G. J. *Magn. Reson. Chem.* **1988**, *26*, 491-496.
- (28) Martin, G. J.; Naulet, N. *Fresenius' Z. Anal. Chem.* **1988**, *332*, 648-651.
- (29) Craig, H. *Geochim. Cosmochim. Acta* **1957**, *12*, 133-149.
- (30) Rauschenbach, P.; Simon, H.; Stichler, W.; Moser, H. *Z. Naturforsch.* **1979**, *34C*, 1-4.
- (31) Misselhorn, K.; Bruckner, H.; Mussig-Zufika, M.; Grafarend, W. *Branntweinwirtschaft* **1983**, 162-170.
- (32) Rinaldi, G.; Meinschen, W. G.; Hayes, J. M. *Biomed. Mass Spectrom.* **1974**, *1*, 412-414.
- (33) Krueger, D. A.; Krueger, H. W. *Biomed. Mass Spectrom.* **1984**, *11*, 472-474.
- (34) Hoffman, P. G.; Salb, M. J. *Agric. Food Chem.* **1979**, *27*, 352-355.
- (35) Bricout, J.; Koziet, J.; Derbesy, M.; Beccat, B. *Ann. Fals. Exp. Chim.* **1981**, *74*, 691-696.
- (36) Martin, G. E.; Alfonso, F. C.; Figert, D. M.; Burggraff, J. J. *Assoc. Off. Anal. Chem.* **1981**, *64*, 1149-1153.

RECEIVED for review March 7, 1991. Accepted July 16, 1991.
This work was supported in part by a grant from Sanofi Bioindustries and the CNRS.

Nonlinear Multivariate Calibration Using Principal Components Regression and Artificial Neural Networks

Paul J. Gemperline,* James R. Long,¹ and Vasilis G. Gregoriou²

Department of Chemistry, East Carolina University, Greenville, North Carolina 27857

This manuscript describes methods for detecting and modeling nonlinear regions of spectral response in multivariate, multi-component spectroscopic assays. Simulated data and experimental UV/visible data were used to study the capability of multivariate linear models to approximate nonlinear response. The sources of real and apparent nonlinearity simulated included nonlinear instrument response functions (e.g. stray light), concentration-dependent wavelength shifts, and concentration dependent absorption bandwidth changes. A weighting algorithm was devised to reduce the influence of nonlinear spectral regions in principal component regression (PCR) calibrations, thereby improving the performance of multivariate linear calibration models. Second-order calibration methods using quadratic principal component scores and nonlinear calibration methods using artificial neural networks were compared to unweighted and weighted linear calibration methods. Orthogonal transformation of the input variables was used to significantly improve neural network training speed and reduce calibration error. Some conditions where second-order and nonlinear calibration techniques outperform linear calibration techniques have been identified and are described.

INTRODUCTION

For the past three years, our laboratory has been working in collaboration with scientists at Burroughs Wellcome Co.

* Corresponding author.

¹ Current address: Air Force and Engineering Services Center, Tyndall AFB, FL 32403-6001.

² Current address: Department of Chemistry, Duke University, Durham, NC 27706.

to develop UV/visible spectroscopic assays for routine determination of the active ingredients and preservatives in pharmaceutical products. Our goal has been to develop simple, accurate ($\pm 0.2\%$), and reliable assays using digitally stored calibration models from UV/visible diode array spectra. Simple check standards are used to verify the accuracy of the stored calibration models so that lengthy recalibration procedures can be avoided. We quickly encountered problems with some of the assays, which we attributed to nonlinear spectral response (1). The symptoms included a lack of short-term and long-term reproducibility (large bias between days, ca. $\pm 1.0\%$) and curvature in the residuals for the estimated concentrations. Such errors are unacceptable in a heavily regulated industry like the pharmaceutical industry. In order to understand this problem more thoroughly, we undertook a study to examine the effects of nonlinear spectral response on PCR calibration models. During the course of the study we identified some circumstances where PCR and PLS satisfactorily model nonlinear response by inclusion of extra factors or latent variables in the calibration model. We have also identified some circumstances where extra factors do not give a satisfactory model of the nonlinear response. To handle the latter circumstances, we report a weighting method for reducing the effects of nonlinear spectral regions on PCR calibration models. We also report the capability of PCR models with quadratic terms and artificial neural networks (ANN) to model nonlinearity.

BACKGROUND

Principal component regression (PCR) and partial least squares (PLS) are two multivariate full-spectrum calibration techniques that have received considerable attention in the chemometrics literature (2-6). Many researchers have observed circumstances where inclusion of extra principal components or latent variables in the calibration model can model

nonlinear response (7–10). The theory for both methods is thoroughly described in the literature (11, 12).

We recently reported the use of artificial neural networks (ANN) with back-propagation of error for developing nonlinear calibration models for spectroscopic assays (13). A brief review and introduction to ANN's may be found in that paper, as well as complete descriptions on the use of artificial neural networks for spectroscopic calibration. Several other chemical applications of neural networks have recently appeared in the literature. Meyer and co-workers have used the pattern recognition capabilities of neural networks to identify alditols (14) and oligosaccharides (15) from NMR spectra. Robb and Munk used linear neural networks for automated functional group analysis of infrared spectra (16). Wythoff and co-workers developed a neural network application for automatic peak detection (17). Elrod et al. used neural networks to predict the products of electrophilic aromatic substitution (18). Jansson recently published an overview of neural networks (19).

Other algorithms for multivariate calibration of nonlinear data have been recently introduced into the chemical literature, including the data preprocessing technique called multiplicative scatter correction (MSC) (20), projection pursuit regression (21), and locally weighted least squares (22). Kowcientniak and Parczewski recently reported a method for performing nonlinear transformations of variables prior to least-squares modeling (23) and Frank and Lanteri reported a method for finding optimal nonlinear transformations of response and predictor variables prior to a univariate regression step (24); however, these latter two approaches were used for univariate calibrations, not multivariate calibrations.

Sources of Nonlinear Response. Real and apparent nonlinear response can be attributed to several different sources, including instrumental sources, physical sources, and chemical sources. Most of the nonlinearities from these sources give rise to one of the following three spectral artifacts: (a) curvature in the concentration-response function, (b) shifts in the position of absorption bands, and (c) changes in the width of absorption bands. The first effect may be classified as a *real* nonlinear effect. The latter two effects give rise to *apparent* nonlinear response in full-spectrum calibration techniques like PCR and PLS. Curvature or other systematic trends can be observed in plots of concentration residuals when these sources of apparent nonlinearity are present. Without prior knowledge, the effect of apparent nonlinearity is difficult to distinguish from real nonlinear response. While chemical equilibria are an important source of variability in spectroscopic measurements, it is well-known that linear multivariate calibration techniques can model these effects exactly, in at least-squares sense, under ideal conditions (21).

Sources of curvature in the instrument response function include nonlinear detector response, stray light at high optical densities, and baseline drift. Sample turbidity can introduce wavelength-dependent baseline offsets. Stray light will comprise a greater fraction of the measured signal at the peaks of strong absorption bands compared to other portions of the spectrum that exhibit low absorbance values; therefore, in complicated spectra, curvature in the concentration-response function caused by high levels of stray light can be expected to be wavelength dependent.

Shifts in the position and width of absorption bands can be attributed to changes in the sample temperature (25) and solvent composition, especially when a change in the solvent composition causes a large change in the dielectric strength of the solvent (26, 27). The magnitude of the shifts and bandwidth changes can be different for different absorption bands. In solids, small changes in the sample temperature can cause large shifts in the position and width of absorption

bands, especially for crystalline species with intermolecular hydrogen bonds (25). In crystalline materials, band broadening can be caused by a change in the amount of structural disorder induced by changes in the concentration of impurities.

We have observed that small wavelength shifts and bandwidth changes seem to cause significant problems in full-spectrum calibration techniques compared to single-wavelength univariate techniques and discrete wavelength techniques like multiple linear regression (MLR). These problems are especially troublesome when we attempt to transfer calibrations from one instrument to another. These kinds of errors can also be classified as a source of apparent nonlinearity in full-spectrum techniques. We have observed that these kinds of errors introduce additional small principal components into calibration models that have derivative-like features. The extra factors seem to be especially susceptible to perturbations by random noise.

There are compelling reasons why multivariate nonlinear calibration techniques and mathematical methods for reducing the effect of nonlinear response in linear multivariate calibration techniques should be developed. Traditional methods for reducing or avoiding nonlinear response will not always work. For example, a traditional method for reducing nonlinear response for samples that exhibit high optical density is to use a shorter cell path length or dilute the sample to give absorbance values that are within the instrument's linear working range. In on-line process monitoring applications, it may be impossible to dilute the process stream. Reducing the cell path length or diluting the sample may help linearize signals from strong absorbers; however, it may also render the signals of weak absorbers too weak for quantitation. In applications using reflectance measurements of solids, "dilution" is impractical and time consuming. Furthermore, some sources of nonlinearity cannot be reduced by either of the above techniques, for example changes in sample temperature and the dielectric strength of solutions. Precise temperature regulation may be difficult to maintain in process streams and solids, while changes in the dielectric strength of solutions are likely to depend on the composition of the solvent or concentration of the species to be quantitated. Mathematical approaches to solving these problems can be very fast, requiring only a few seconds of computer time, and may provide a viable alternative for solving the above problems.

F-Test for Finding Nonlinear Spectral Regions. In mixtures where the concentration of all of the absorbing components is known and the Beer-Lambert law strictly holds, the "multiple multivariate" regression model can be used to describe the absorption spectra.

$$\mathbf{A} = \mathbf{CB} \quad (1)$$

$$(\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{A} = \hat{\mathbf{B}} \quad (2)$$

In eq 1, \mathbf{A} is an $n \times m$ matrix (n samples by m wavelengths) of dependent variables (spectra), \mathbf{C} , is an $n \times k$ matrix (n samples by k components) of independent variables (concentration), and \mathbf{B} is the $k \times m$ matrix of least-squares regression coefficients that minimize $\|\mathbf{A} - \mathbf{CB}\|$. The notation, $\|\mathbf{X}\|$, represents the Euclidean norm of \mathbf{X} . The standard least-squares estimate of the regression coefficients, $\hat{\mathbf{B}}$ is given by eq 2. In this calibration model and for PCR and PLS calibration models, it is assumed that the spectra and concentrations are mean centered prior to the regression step to include nonzero intercepts in the models. For quadratic PCR and ANN calibration models, coefficients for treating nonzero intercepts are included in the models.

Nonlinear response can be detected by comparing the linear model described by eq 1 to a model having linear, first-order interaction, and quadratic independent variables. Here we assume that a quadratic model will give a better approxi-

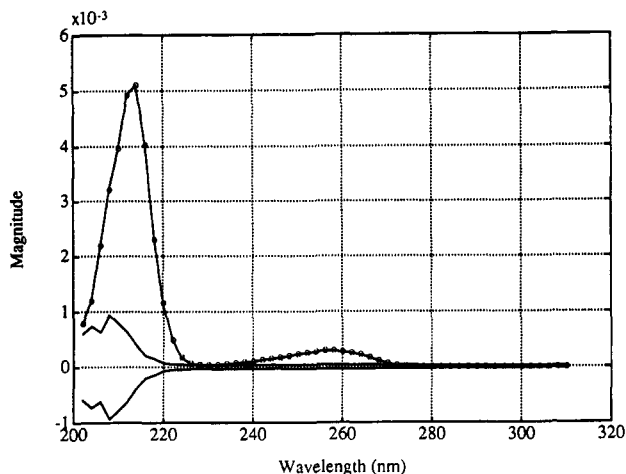


Figure 1. Linear regression coefficient (O) for the active ingredient, pseudoephedrine hydrochloride, in Actifed syrup and the 95% confidence region (—). Regression coefficients inside the bounded region are statistically no different from zero at the 95% probability level.

mation to the curved instrument response function compared to a linear model. To generate the quadratic model, we simply augment the columns of the centered concentration matrix, C , with the interaction and quadratic terms and substitute the new matrix into eqs 1 and 2:

$$C^* = [C; c; c_j(i \neq j)]C^2 \quad (3)$$

The interaction terms are formed by multiplying the concentration of the i th component by the j th component for all $i \neq j$. The quadratic terms are formed by squaring the concentration of all components. Extra regression coefficients for the interaction and quadratic terms are automatically estimated by eq 3. Plots of these regression coefficients as a function of wavelength can be used to identify regions where linear, interaction, and quadratic response is observed for each constituent. The confidence interval for the regression coefficient for the i th component at the j th wavelength can be determined from eqs 4–7, where $\text{Var}(b_{ij})$ is the i th diagonal

$$b_{ij} \pm \sqrt{\text{Var}(b_{ij})} \sqrt{(k+1)F_{k+1, n-k-1}(\alpha)} \quad (4)$$

$$\text{Var}(b_{ij}) = s_j^2 (C^T C)_{ii}^{-1} \quad (5)$$

$$s_j^2 = \sum_{jj} \quad (6)$$

$$\sum = \frac{(A - \hat{A})^T (A - \hat{A})}{n-1} \quad (7)$$

element of $s_j^2 (C^T C)_{ii}^{-1}$, k is the number of columns in C or C^* , and s_j^2 is the j th diagonal element of \sum . Equations 4–7 are correct for C or C^* even though elements of C are correlated provided C has full rank. The variance-covariance matrix of spectral residuals, \sum , is calculated according to eq 7 where \hat{A} is the matrix of response variables estimated from the regression model.

These confidence intervals can be plotted with the regression coefficients to aid scientists in judging the significance of the coefficients. For example, Figures 1 and 2 show plots of the linear and quadratic regression coefficients for one of the active ingredients, pseudoephedrine hydrochloride, in Actifed syrup. In addition to two active ingredients, the syrup also contains two preservatives, a sugar base, a flavoring agent, and a dye to add color. The confidence intervals have been plotted about the zero axis. Regression coefficients that lie inside the bounded regions are statistically no different from zero at the 95% confidence level. Although the linear regression coefficients in the region 208–220 nm are statistically significant, the significant higher order regression coefficients in this same range indicate that response in this region is

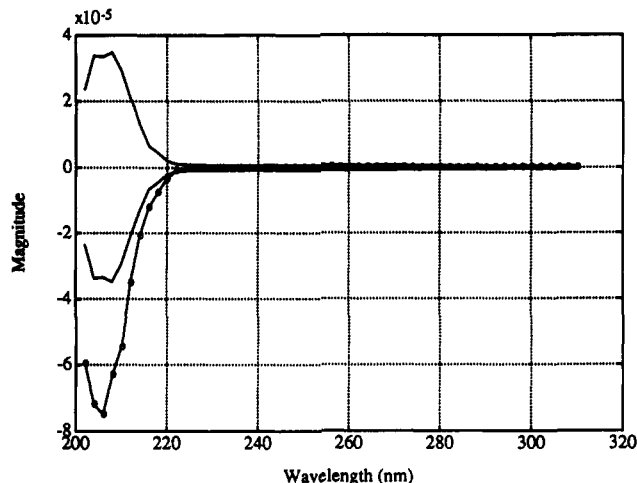


Figure 2. Quadratic regression coefficient (O) for the active ingredient, pseudoephedrine hydrochloride, in Actifed syrup and the 95% confidence region (—). Regression coefficients inside the bounded region are statistically no different from zero at the 95% probability level.

nonlinear. The region from 275 to 310 nm does not contain any significant linear or nonlinear information for pseudoephedrine hydrochloride. A significant improvement in the standard error of prediction (SEP) for PCR was observed for some of the components when the wavelength range from 202 to 220 nm was omitted.

Weighted PCR Regression. The confidence intervals and linear regression coefficients from PCR can be used to determine weighting coefficients that emphasize spectral regions where spectral variation makes an important contribution to the calibration model and deemphasize spectral regions where unmodeled variance due to nonlinear response or poor signal-to-noise ratios gives large spectral residuals. The PCR regression model is used to develop the weighting coefficients instead of the model shown in eqs 1–7 so that weights can be determined when some of the absorbing components are unknown. As in the previous section, it is assumed that the spectra and concentrations are mean centered prior to the regression step to accommodate nonzero intercepts in the model. The PCR regression model is defined by eqs 8 and 9, where c is the $n \times 1$ vector of concentrations for one con-

$$c = Ut \quad (8)$$

$$A = USV^T \quad (9)$$

stituent in the calibration standards, and t is a $k \times 1$ vector of regression coefficients that minimize $\|c - Ut\|$. In eq 8, U is the $n \times k$ matrix of the orthonormal scores for the first k principal components from the singular value decomposition of A according to eq 9, where A is an $n \times m$ matrix of n calibration spectra, S is the $k \times k$ matrix of singular values, and V^T is the $k \times m$ matrix of k eigenvectors. Equation 8 can be rewritten by substituting $U = AVS^{-1}$.

$$c = AVS^{-1}t \quad (10)$$

Now, an $m \times 1$ vector of regression coefficients, b , can be defined that minimizes $\|c - Ab\|$ according to eq 11. Equation 12 can be used to obtain the least-squares estimate of the regression vector. This vector of regression coefficients can

$$c = Ab \quad (11)$$

$$\hat{b} = VS^{-1}t \quad (12)$$

be plotted as though it were a "spectrum" to discover regions in the calibration spectra that have large residuals that do not contribute a significant amount of variance to the concentration of the constituent being quantitated. Because PCR is used to solve the "inverse calibration" problem there are

differences in how the confidence intervals are calculated compared to the way they were calculated for the "forward calibration" model in eqs 1-7. For the component being quantitated, the confidence interval at the j th wavelength is given by eqs 13 through 20. The term, $\text{Var}(b_j)$, is the variance

$$b_j \pm \sqrt{\text{Var}(b_j)} \sqrt{(k+1)F_{k+1, n-k-1}(\alpha)} \quad (13)$$

$$\text{Var}(b_j) = s^2(\mathbf{A}^T \mathbf{A})_{jj}^{-1} \quad (14)$$

$$s^2 = \frac{\mathbf{e}^T \mathbf{e}}{n-1} \quad (15)$$

$$\mathbf{e} = \mathbf{c} - \mathbf{A}\mathbf{b} \quad (16)$$

$$\mathbf{A}^T \mathbf{A} = (\mathbf{V}\mathbf{S}\mathbf{U}^T)(\mathbf{U}\mathbf{S}\mathbf{V}^T) \quad (17)$$

$$\mathbf{A}^T \mathbf{A} = \mathbf{V}\mathbf{S}^2\mathbf{V}^T \quad (18)$$

$$(\mathbf{A}^T \mathbf{A})^{-1} = \mathbf{V}\mathbf{S}^{-2}\mathbf{V}^T \quad (19)$$

$$\text{Var}(b_j) = s^2(\mathbf{V}\mathbf{S}^{-2}\mathbf{V}^T)_{jj} \quad (20)$$

associated with the regression coefficient at the j th diagonal (wavelength) of $(\mathbf{A}^T \mathbf{A})_{jj}^{-1}$, where s^2 is the residual variance for the estimated concentrations of the constituent being quantitated. In order to avoid inversion of the nearly singular matrix, $\mathbf{A}^T \mathbf{A}$, in eq 14, the decomposition of \mathbf{A} into k principal components is used (eq 17), which simplifies to eq 18. The inverse, $(\mathbf{A}^T \mathbf{A})^{-1}$, is easily calculated according to eq 19, allowing eq 14 to be rewritten in a form that is easy to compute (eq 20). These equations can be used to determine a matrix of regression coefficients and confidence intervals, $\mathbf{B} \pm \Delta = [\mathbf{b}_1 | \mathbf{b}_2 | \dots | \mathbf{b}_i] \pm [\delta_1 | \delta_2 | \dots | \delta_i]$, for any number of components that must be simultaneously quantitated.

Weighting coefficients for the i th component are determined from the unweighted PCR regression coefficients and confidence intervals from eq 13 for the i th component, $b_i \pm \delta_i$. We define the vector of weights for the i th component, \mathbf{w}_i , as the set of all w_{ij} greater than zero such that $w_{ij} = \text{abs}(b_{ij}) - \delta_{ij}$.

$$\{w_{ij} > 0 | w_{ij} = \text{abs}(b_{ij}) - \delta_{ij}\} \quad (21)$$

$$\mathbf{A}_i^* = \mathbf{A}\mathbf{w}_i \quad (22)$$

The i th weighted data matrix is then used to develop the PCR calibration model for the i th component. Wavelength variables that have weights less than or equal to zero are deleted from the data matrix, \mathbf{A}^* , before the regression step. This weighting function automatically deletes spectral regions that have large residuals and do not contribute a significant amount of variance to the concentration of the constituent being quantitated. As an example, the PCR regression vector and confidence interval for the determination of triprolidine hydrochloride in Actifed syrup is shown in Figure 3. The corresponding weight vector is shown in Figure 4. A significant improvement in SEP for triprolidine was observed when weighted PCR was used.

Quadratic PCR. In some cases it was observed that quadratic principal component scores could be used to facilitate modeling of real and apparent nonlinear response. The quadratic principal component model is determined by substituting an augmented matrix of principal component scores, \mathbf{U}^* , into eqs 8 through 12:

$$\mathbf{U}^* = [\mathbf{1} | \mathbf{U} | \mathbf{U}^2] \quad (23)$$

This approach can give improved results compared to linear calibration models like PCR and PLS if there is curvature in the scores versus concentration which can be more accurately approximated by a polynomial having the form $c_i = b_0 + b_1 u_i + b_2 u_i^2$.

Artificial Neural Networks. In a previous paper, we showed that artificial neural networks can be used for spectroscopic calibration (13). Figure 5 shows a schematic diagram

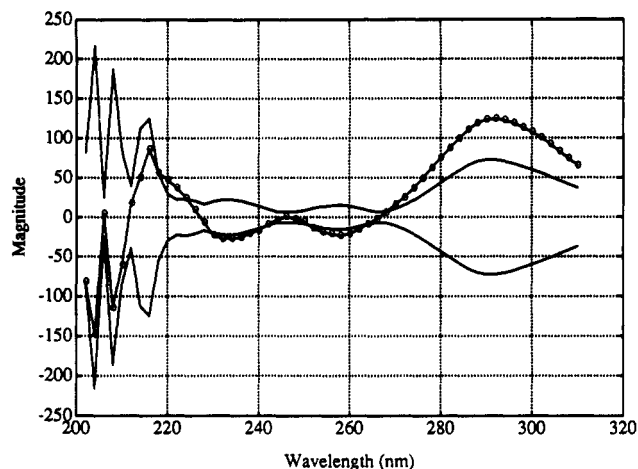


Figure 3. PCR regression vector (O) for the active ingredient, triprolidine hydrochloride, in Actifed syrup and the 95% confidence region (—). Regression coefficients inside the bounded region are statistically no different from zero at the 95% probability level.

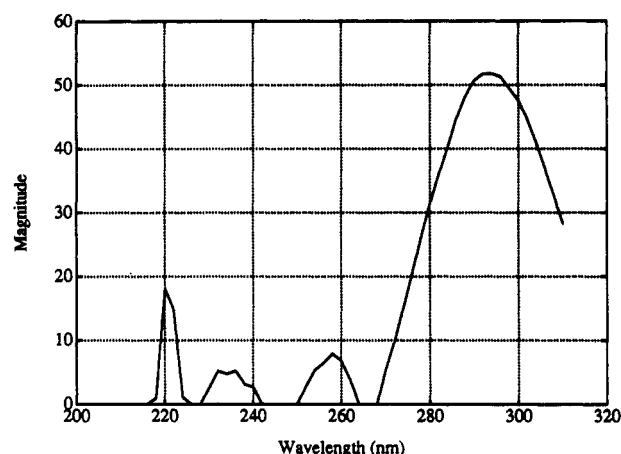


Figure 4. Weighting vector for the PCR calibration of triprolidine hydrochloride in Actifed syrup. The weight vector was derived from the regression vector and confidence intervals shown in Figure 3.

of a typical artificial neural network we used in these investigations. The artificial neural network in Figure 5 consists of three layers of nodes: an input layer, a hidden layer, and an output layer. In calibration applications, the input signals are absorbance values, a_i , measured at many different wavelengths. The input nodes have linear outputs ($o_i = a_i$) and serve to buffer the input signals to the hidden nodes. Their linear outputs are sent to the nodes in the hidden layer through weighted connections, as shown in Figure 5. The outputs of the hidden layer nodes are given by weighted sum, $o_j = f(\sum o_i w_{ij})$, after transformation by some function, $f(x)$. The nodes labeled "bias" output a constant value of 1.0. The weights connecting the bias nodes to the next layer of nodes are automatically adjusted during the training process to permit incorporation of nonzero intercepts into neural network calibration models. In our applications, only one linear output node was used in the output layer, having an output given by the sum of weighted inputs from the hidden layer nodes, $o_k = \sum o_j w_{jk}$. The weights are adjusted during the training process to give an output equal to the scaled concentration of the component of interest. Scaling and other important network training parameters are described in the experimental section.

In this study, we incorporated different linear and nonlinear transfer functions into the hidden layer nodes to facilitate nonlinear fitting. The transfer functions tested included a linear function, $y = x$, a quadratic function, $y = x^2$, a sigmoid function, $y = 1/(1 + e^{-x})$, and a hyperbolic tangent function,

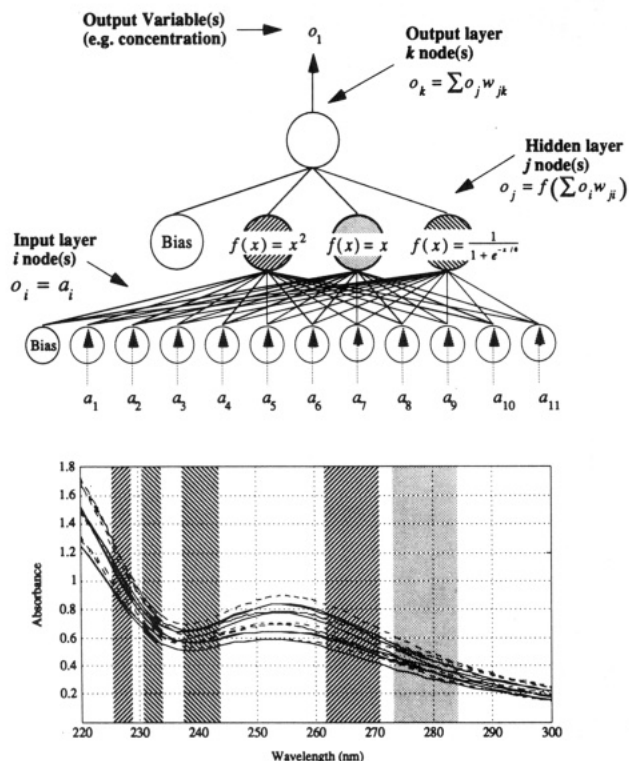


Figure 5. Schematic diagram showing the architecture of a hypothetical artificial neural network for spectroscopic calibration. Linear output functions are used for the input nodes and the output node. Different nonlinear nodes are used in the hidden layer. Weighted connections are adjusted during training. Shaded spectral regions indicate where large positive or negative weights connect the corresponding hidden node.

$y = \tanh(x)$. Currently, we use trial-and-error to select the number of nodes in the hidden layer and the transfer functions used in the hidden layer nodes. The network architecture that gives the minimum prediction error is retained. To guard against overfitting we reject neural network models where SEC is significantly lower than SEP or vice versa. We are currently devising techniques for automatically selecting the best network architecture.

During the training process, the weights are iteratively adjusted by back-propagating the error at the output node through the network. The training process is stopped when the desired level of precision is obtained in the estimated concentrations. The most exciting aspect of this approach is that the weights can be automatically adjusted by the training process to accommodate linear response and different types of nonlinear response in different spectral regions. One method for viewing this behavior takes advantage of how spectral variation propagates *forward* through the network. Figure 6 shows the square root of the spectral variation that propagates forward in the neural network used for the calibration of pseudoephedrine hydrochloride. The network had two hidden layer nodes, one sigmoid and one linear. Two variance "spectra" are shown, one for each node. The spectral variation at each hidden node was determined by multiplying the input variables (scaled spectra) by the vector of weights connecting the input nodes to the corresponding hidden layer node. The variance was then calculated at each wavelength from the resulting scaled and weighted spectra. From Figure 6, it can be seen that different portions of the input spectra are important for determining the response of the linear and sigmoid hidden nodes. We are currently exploring additional graphical techniques for understanding how neural networks model nonlinear spectral variation.

The back-propagation training rule for neural networks is equivalent to the steepest decent technique of function min-

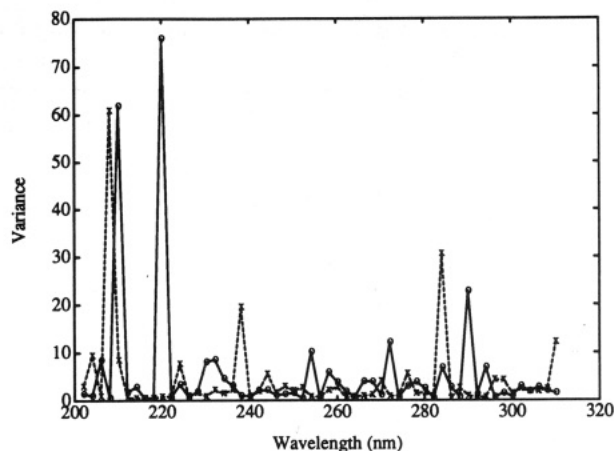


Figure 6. Plot of the square root of the spectral variance propagated forward from the input variables (scaled spectra) for the neural network calibration of pseudoephedrine hydrochloride in Actifed syrup: (x) spectral variance propagated forward at the linear hidden node; (o) variance propagated forward at the sigmoid hidden node.

imization. The steepest decent technique is used to minimize the network output error by adjusting the weights to decrease the output error in the direction of local downhill gradient. It is well-known that the steepest decent technique performs poorly (slow convergence), especially when the input variables are highly correlated (29), which is precisely the circumstance encountered in our spectroscopic calibration work. In order to improve training speed and decrease the overall calibration error in neural networks, orthogonal transformations of the measured response variables can be used instead of the measured response variables. The singular value decomposition is used to perform the orthogonal transformation, giving k principal components

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (24)$$

where \mathbf{A} is the $(n \times m)$ matrix of n calibration spectra measured at m correlated response variables and $k < m$, $k < n$ columns of \mathbf{U} are used as the new transformed orthogonal variables. For "unknown" samples, the orthogonal variables are computed according to eq 25, where \mathbf{a}_{unk} is the unknown

$$\mathbf{u}_{\text{unk}} = \mathbf{a}_{\text{unk}}\mathbf{V}\mathbf{S}^{-1} \quad (25)$$

sample's measured spectrum, \mathbf{u}_{unk} is the unknown sample's vector of new orthogonal variables, and \mathbf{S} and \mathbf{V} are determined from the singular value decomposition of the calibration spectra according to eq 24.

The number of orthogonal variables, k , used in the network is selected by "pruning" input nodes from the network having variables with small singular values. As long as nonsignificant nodes are pruned from the network only small insignificant increases are observed in the calibration error. A large increase in the calibration error is observed when important input variables are pruned from the network.

Without orthogonal input variables, network training required about 10 000–150 000 iterations per component (5–45 min on a 25-MHz 80386 personal computer) depending on the complexity of the network and the calibration problem. When orthogonal input variables were used, network training required 500–10 000 iterations per component (5 s to 5 min) depending on the complexity of the network and the calibration problem.

EXPERIMENTAL SECTION

Two-component and three-component simulated UV/visible data matrices were generated to examine three different nonlinear effects: (a) the effect of high stray light, (b) the effect of concentration-dependent peak shifts, and (c) the effect of concentration-dependent bandwidth changes. The simulated spectra

Table I. Parameters Used To Generate Simulated Spectra for Concentration-Dependent Peak Shifts and Bandwidth Changes

component	band 1		band 2	
	position (amt of shift)	width (amt of shift)	position (amt of shift)	width (amt of shift)
A	220.0 (± 0)	8.0 (± 0)	275.0 ($\pm 1.28C_B$)	25.0 ($\pm 1.28C_B$)
B	200.0 ($\pm 2.56C_A$)	15.0 ($\pm 1.28C_A$)	260.0 (± 0)	25.0 (± 0)
C			245.0 ($\pm 1.28C_C$)	15.0 ($\pm 1.28C_C$)

^a All simulated band positions, widths, and shifts are given in nanometers. C_A , C_B , and C_C represent the concentration of component A, B, and C respectively. See Figure 7 for plots of the simulated spectra.

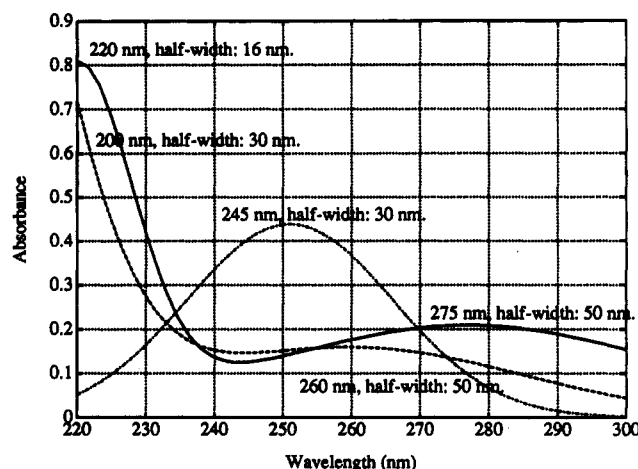


Figure 7. Simulated UV/visible spectra for two-component and three-component mixtures: (—) component A; (---) component B; (-·-) component C.

were generated from overlapping Gaussian bands. The pure spectra for all three simulated components are shown in Figure 7. The position (μ) and width at half-height (2σ) for each absorption band is shown on the plot.

A composite three-level factorial design was used to select the concentration of the calibration standards, expressed as percent of nominal assay concentration. These experimental designs are shown in Figure 8. A single standard at 100% of the nominal assay concentration for each component is the center point of the experimental design. Each component was varied $\pm 20.0\%$ to give the points at the vertices of the experimental design and $\pm 32.0\%$ to give the points on the faces of the experimental design. These designs were selected because they contain sufficient information to permit estimation of the linear, quadratic, and interaction coefficients for quadratic response surfaces.

Simulated "unknowns" (validation data sets) were also generated by using composite three-level factorial designs. A single standard at 100% of the nominal assay concentration is the center point of the design. Each component was varied $\pm 10.0\%$ to give the points at the vertices of the experimental design and $\pm 16.0\%$ to give the points on the faces of the experimental design. These different response levels were chosen in order to provide a more valid way of testing the calibration models, since it is possible for the calibration functions to pass through the calibration design points while providing very different behavior at other positions in the calibration domain or "assay working range". Proper estimation of these unknowns can be used to demonstrate that a calibration model is useful for interpolation between calibration points.

Simulated Stray Light. Stray light was simulated by transforming the simulated mixture spectra according to eq 24:

$$A_{ij} = -\log(10^{-A_{ij}} + 0.001) \quad (26)$$

A plot of the resulting instrument-response function is shown in Figure 9. Normally distributed random error ($\mu = 0.000$ AU, $\sigma = 0.001$ AU) was added to each simulated mixture spectrum after transformation. The transformation was performed after the pure component spectra were added together to form the simulated mixture spectra; therefore the nonlinearity is a complicated function of the sum of the concentration of all species in the

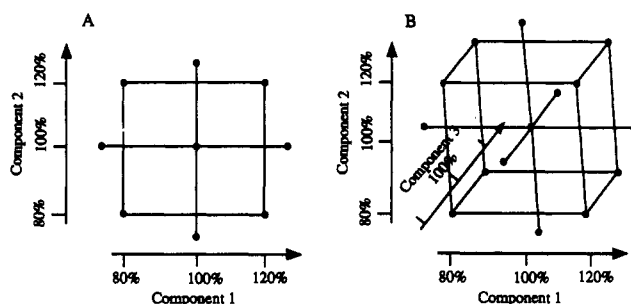


Figure 8. Experimental designs for simulated two-component and three-component mixtures. Each design point designates the composition of a calibration standard. Different factor levels were used for simulated "unknowns" (see text).

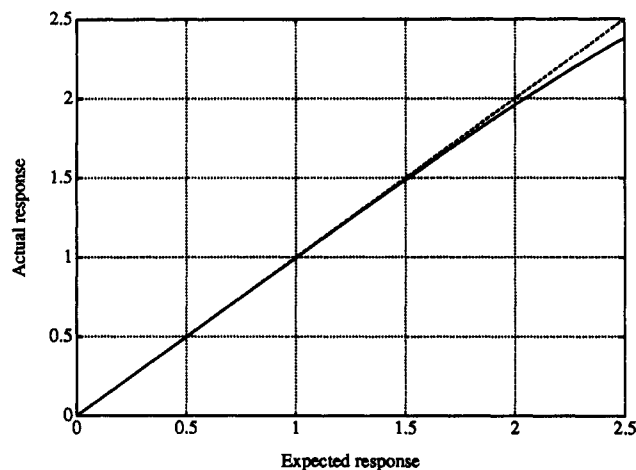


Figure 9. Linear instrument response function (---) and nonlinear instrument response function (—) with 0.1% stray light.

mixtures. The nominal assay concentration (100%) was adjusted so that the mixture spectra had a maximum absorbance level of approximately 2.3–2.5 in the range from 220 to 230 nm. Those portions of the mixture spectra that exhibit a total absorbance greater than 1.5 gave significant nonlinear response, while portions which exhibit a total absorbance less than 1.5 gave nearly linear response; therefore the nonlinearity is wavelength dependent.

Simulated Peak Shifts and Bandwidth Changes. Simulated spectra were generated by producing Gaussian bands at various peak positions and bandwidths. For the data sets with simulated peak shifts, the program used to generate the simulated spectra was modified to introduce concentration-dependent peak shifts according to the specifications shown in Table I. Normally, distributed random error ($\mu = 0.000$ AU, $\sigma = 0.001$ AU) was added to each simulated mixture spectrum after peak shifts were introduced. For the data sets with simulated bandwidth changes, the program was modified to introduce a concentration-dependent bandwidth changes according to the specifications shown in Table I. Normally, distributed random error ($\mu = 0.000$ AU, $\sigma = 0.001$ AU) was added to each simulated mixture spectrum after bandwidth changes were introduced. It is important to note that the simulated spectra for components A and B have two absorption bands each and that only one of the two absorption bands from each was altered to create peaks shifts and bandwidth changes.

Table II. Artificial Neural Network Specifications and Training Parameters

parameter	value
input layer nodes	variable (see text)
input layer transfer function	linear
hidden layer nodes	variable (see text)
output layer nodes	1
output layer transfer function	linear
learning rate	0.05 or 0.15 ^a
momentum	0.01 or 0.05 ^b
gain	1.0
training presentation order	random

^a A value of 0.05 was used for training neural networks with full spectra. A value of 0.15 was used for training neural networks principal component input variables. ^b A value of 0.01 was used for training neural networks with full spectra. A value of 0.05 was used for training neural networks principal component input variables.

UV/Visible Spectroscopic Data. Calibration models were developed for two pharmaceutical products using UV/visible spectroscopy. One product was an experimental injectable product containing a proprietary active ingredient and benzyl alcohol as a preservative. The nominal assay concentrations (100%) for the active ingredient and benzyl alcohol were 0.1 and 0.9 mg/mL, respectively. Mixed calibration standards and test standards were prepared by the appropriate dilution of stock solutions according to the experimental design shown in Figure 8a. Additional details may be found in ref 1. The UV/visible spectra of the mixed standards were obtained by using a Hewlett-Packard 8452A diode array spectrophotometer in the wavelength range from 270 to 290 nm in increments of 2 nm.

Calibration models were also developed for the active ingredients and preservatives in Actifed syrup. The nominal assay concentrations (100%) were 0.13, 0.025, 0.006, and 0.006 mg/mL of pseudoephedrine hydrochloride, triprolidine hydrochloride, sodium benzoate, and methyl paraben, respectively. Mixed calibration standards and test standards were prepared according to a five-component experimental design (two active ingredients, two preservatives, plus excipients) analogous to the design shown in Figure 8b. A complete description of the experimental design and additional experimental details may be found in ref 13. UV/visible spectra of the mixed standards were measured from 200 to 310 nm in 2-nm steps by using the Hewlett-Packard diode array spectrophotometer described above.

Network Training Parameters. The input and output variables for each data set were scaled prior to neural network training to lie in the range from 0.2 to 0.8 by adding an offset and multiplying by a constant. Scaling helps make the network training parameters (e.g. learning rate, momentum, and gain) less sensitive to the application at hand. With proper scaling the training parameters in Table II can be used for all of the applications described in this manuscript.

We used a random presentation order for training vectors with weight updates after presentation of each training vector. Presentation of a single training vector was counted as one iteration. Test data sets were analyzed at frequent intervals (ca. 20–50 iterations) to evaluate SEP. The network was automatically saved as new minimum values of SEP were obtained. Training was stopped when SEP converged to an average minimum value.

We found that injecting small amounts of random noise (± 0.01) with the scaled input variables early in the training process helped speed up convergence. This approach appears to reduce the network's tendency to temporarily converge to local minima. As the network converged to a minimum value of SEP the random "noise injectors" were turned off.

The number of input nodes used depends on the application. For scaled absorption spectra, one input node was used for each simulated or digitized wavelength. For neural networks with orthogonal principal components (PC-ANN) k input nodes were used, where k is the number of principal components used as shown in the column labeled "config^b" in Tables III through V. The transfer functions and number of nodes used in the hidden layer are also shown in Tables III and IV in the columns labeled

"config". For example, the notation "3LS" indicates that three principal components were used as input variables and two nodes were used in the hidden layer, one with a linear transfer function and the other with a sigmoid transfer function. Other important network training parameters are shown in Table II.

RESULTS

Six different calibration algorithms were tried on each data set: (1) principal component regression (PCR), (2) partial least squares (PLS), (3) weighted PCR (WPCR), (4) quadratic PCR (QPCR), (5) artificial neural networks (ANN), and (6) artificial neural networks with orthogonal principal components (PC-ANN). The performance of each method is shown in Tables III through V. The simulated two-component and three-component linear data sets serve as benchmarks for judging the performance of nonlinear calibration algorithms on the nonlinear data sets. For example, if a nonlinear calibration algorithm were able to completely compensate nonlinear response in a data set, then we would expect the calibration and prediction error to be equivalent to the error obtained by a linear calibration algorithm on a linear data set.

The columns labeled % SEC in Tables III–V show the relative standard error of calibration for the j th component where \bar{c}_j is the mean concentration of the j th component, \hat{c}_{ij} is the estimated concentration of the j th component for the i th standard, n is the number of calibration standards, and k is the number of columns in U (see eq 8) or U^* (see eq 23):

$$\% \text{ SEC}_j = \frac{100}{\bar{c}_j} \left[\frac{\sum_{i=1}^n (\hat{c}_{ij} - c_{ij})^2}{n - k - 1} \right]^{1/2} \quad (27)$$

In cases where mean centering is not performed, $n - k$ is used in the denominator of eq 27 instead of $n - k - 1$. For artificial neural networks, k , the number of degrees of freedom used in the model is unknown, therefore as an approximation, n , is used in the denominator of eq 27.

The columns labeled % SEP in Tables III–V show the relative standard error of prediction for the j th component in the "unknowns" or validation data sets. In eq 28, \bar{c}_j is the mean concentration of the j th component in the calibration

$$\% \text{ SEP}_j = \frac{100}{\bar{c}_j} \left[\frac{\sum_{i=1}^n (\hat{c}_{ij} - c_{ij})^2}{p} \right]^{1/2} \quad (28)$$

data set, \hat{c}_{ij} is the estimated concentration of the j th component for the i th "unknown," and p is the number of "unknowns". An F -test can be used to compare SEP for two calibration methods according to eq 29 where p_i and p_j are the number of "unknown" validation samples used to determine SEP. Using this approach, the PC-ANN result for

$$F(p_i, p_j) = (\text{SEP}_i / \text{SEP}_j)^2 \quad (29)$$

component B in the two-component stray light data set is significantly better than PCR ($F = 6.65$) and PLS ($F = 4.66$) at the 95% confidence level ($F_{\text{critical}} = 3.18$; $p_i, p_j = 9, 9$).

Overall, PCR and PLS gave useful results for all of the simulated data sets. In many cases PLS gave slightly lower values for SEP compared to PCR; however, the differences in SEP were not significant at the 95% confidence level. In general, the quadratic PCR results are slightly better than the results for PCR and PLS. The differences in SEP are not significant at the 95% confidence level except for components A and B in the three-component data set with bandwidth changes. Compared to PCR and PLS, the weighted PCR results are worse except for pseudoephedrine and triprolidine in Actifed syrup. The weighted PCR results for these two components are significantly better at the 95% confidence

Table III. Linear and Nonlinear Calibration Results for Simulated Two-Component Data Sets

two-component simulated data sets	PCR			PLS			quad PCR			wt PCR			ANN			PC-ANN		
	factors	% SEC	% SEP	factors	% SEC	% SEP	factors	% SEC	% SEP	factors	% SEC	% SEP	config ^a	% SEC	% SEP	config ^b	% SEC	% SEP
linear																		
comp A	2	0.033	0.039	3	0.008	0.023	2	0.033	0.034	2	0.035	0.043	LS	0.124	0.137	3LL	0.029	0.029
comp B	2	0.024	0.049	3	0.008	0.041	2	0.024	0.045	2	0.025	0.046	L	0.086	0.084	3LL	0.020	0.019
stray light																		
comp A	3	0.115	0.096	3	0.109	0.096	3	0.030	0.100	3	0.147	0.104	LS	0.186	0.185	3LS	0.098	0.094
comp B	3	0.031	0.061	3	0.030	0.061	3	0.038	0.062	3	0.046	0.061	LS	0.083	0.121	3LL	0.042	0.058
peak shifts																		
comp A	6	0.068	0.279	5	0.021	0.160	3	0.025	0.220	3	0.110	0.322	LS	0.116	0.358	5LLS	0.280	0.217
comp B	3	0.186	0.483	3	0.181	0.484	3	0.044	0.475	3	0.288	0.415	LS	0.174	0.601	3LLS	0.209	0.336
bandwidth changes																		
comp A	4	0.225	0.308	4	0.193	0.309	3	0.405	0.270	4	0.374	0.339	LS	0.186	0.429	4LSS	0.187	0.263
comp B	4	0.218	0.451	4	0.197	0.453	3	0.567	0.331	4	0.309	0.873	LS	0.203	0.303	5LSS	0.360	0.405

^a L and S represent nodes in the hidden layer with linear and sigmoid transfer functions, respectively. ^b N represents the number of orthogonal input variables used.

Table IV. Linear and Nonlinear Calibration Results for Simulated Three-Component Data Sets

three-component simulated data sets	PCR			PLS			quad PCR			wt PCR			ANN			PC-ANN		
	factors	% SEC	% SEP	factors	% SEC	% SEP	factors	% SEC	% SEP	factors	% SEC	% SEP	config ^a	% SEC	% SEP	config ^b	% SEC	% SEP
linear																		
comp A	3	0.171	0.195	4	0.087	0.183	4	0.170	0.141	3	0.196	0.243	LL	0.277	0.418	4LS	0.177	0.141
comp B	3	0.238	0.299	4	0.110	0.282	4	0.231	0.211	4	0.321	0.397	LLS	0.426	0.614	4LS	0.242	0.212
comp C	3	0.065	0.081	3	0.065	0.081	3	0.045	0.077	4	0.089	0.104	LL	0.152	0.166	3L	0.066	0.076
stray light																		
comp A	8	0.657	0.414	4	0.405	0.382	4	0.674	0.432	7	0.512	0.446	LS	0.277	0.418	9LS	0.474	0.331
comp B	6	1.160	0.788	4	0.783	0.727	4	1.245	0.716	7	0.928	0.805	LS	0.426	0.614	9LS	0.848	0.658
comp C	5	0.215	0.210	4	0.156	0.200	4	0.242	0.192	7	0.336	0.308	LLS	0.152	0.166	5LS	0.178	0.172
peak shifts																		
comp A	9	0.437	0.521	6	0.082	0.510	6	0.403	0.804	5	0.456	0.577	LLS	0.345	0.906	8LLS	0.276	0.361
comp B	8	0.879	0.680	5	0.621	0.715	6	0.439	1.326	10	0.576	1.200	LLS	1.244	1.099	9LS	0.580	0.687
comp C	4	0.486	0.548	5	0.160	0.508	6	0.170	1.063	10	0.251	0.487	LLS	0.264	0.685	6LLS	0.210	0.296
bandwidth changes																		
comp A	10	1.276	0.975	5	0.609	1.004	5	0.876	0.541	8	1.124	1.234	LLS	0.557	1.040	9LSS	0.652	0.669
comp B	8	1.660	1.581	6	0.348	1.655	6	1.089	0.754	8	1.759	1.766	LLS	0.688	1.949	9LLS	1.181	1.266
comp C	11	1.373	1.426	5	1.008	1.184	6	2.035	1.008	11	1.125	1.849	LLS	0.746	0.940	10LSS	1.035	1.048

^a L and S represent nodes in the hidden layer with linear and sigmoid transfer functions, respectively. ^b N represents the number of orthogonal input variables used.

Table V. Linear and Nonlinear Calibration Results for Experimental Data Sets

experimental data sets	PCR			PLS			quad PCR			wt PCR			ANN			PC-ANN		
	factors	% SEC	% SEP	factors	% SEC	% SEP	factors	% SEC	% SEP	factors	% SEC	% SEP	config ^a	% SEC	% SEP	config ^b	% SEC	% SEP
injectible product																		
benzyl alcohol	2	0.404	0.312	2	0.400	0.315	3	0.064	0.252	2	0.406	0.311	S	0.260	0.198	2SS	0.108	0.181
active ingredient	3	0.178	0.352	3	0.162	0.343	3	0.090	0.342	3	0.158	0.397	SS	0.126	0.234	3SS	0.128	0.247
Actifed syrup																		
pseudoephedrine	8	1.58	7.67	5	1.22	3.52	4	1.60	6.47	7	1.39	1.61	LS	1.26	2.37	6LS	1.21	1.23
triprolidine	4	11.66	12.05	12	2.15	11.93	1	19.46	18.63	5	2.18	2.86	S	1.23	3.14	14LS	1.57	2.14
methyl paraben	7	0.22	0.85	4	0.28	0.40	7	0.21	0.76	4	0.26	0.79	LSS	0.46	0.23	7LS	0.20	0.36
sodium benzoate	3	0.56	0.49	4	0.26	0.72	6	0.24	1.88	3	0.78	0.73	LSS	0.23	0.36	7LS	0.23	0.31

^a L and S represent nodes in the hidden layer with linear and sigmoid transfer functions, respectively. ^b N represents the number of orthogonal input variables used.

level. The ANN results are not significantly different from the PCR and PLS results except for pseudoephedrine and triprolidine in Actifed syrup. The ANN results for these two components are significantly better at the 95% confidence level. Overall the PC-ANN results are consistently better (e.g. lower values of SEP) than those of PCR and PLS; however, most of the improvements in SEP are not significant at the 90% confidence level. Notable exceptions to this observation are the results for component B in the two-component stray light data set and the experimental data sets. The results for benzyl alcohol, pseudoephedrine, triprolidine, and sodium benzoate are all significantly better (e.g. lower values of SEP) at the 95% confidence level.

From Tables III-V, it can be seen there is not much difference between the results produced by PCR and PLS for the linear simulated data sets. Quadratic PCR and weighted PCR gave approximately the same performance as PCR and PLS for the simulated linear data sets while ANN gave worse results. We found that ANN calibration performed poorly on linear data sets when correlated input variables are used. Switching to orthogonal input variables solved the problem. In the three-component data set, SEP for components A and B is larger than SEP for component C because the pure spectra of components A and B are highly similar (see Figure 7), while the pure spectrum of component C is dissimilar from A and B.

For the simulated data sets with stray light, linear calibration models using PCR and PLS gave useful results. In most cases PLS gave slightly lower values for SEP compared to PCR while PC-ANN gave results slightly better than PCR and PLS. Although the results are not shown here, PCR and PLS calibrations were also developed for the simulated stray light data without any noise. For the two-component data set, nearly perfect calibration models were obtained with five factors (SEP = 0.004%). When random error was added, the predictive performance of PCR and PLS was degraded since fewer principal components or latent variables could be used in the calibration model. The increased noise masked the useful information in the smallest principal components. We believe the PC-ANN calibration technique is able to provide slightly lower values for SEP for these data sets by approximating some of the curvature in the largest principal components. Plots of some of the regression vectors and confidence intervals for the two- and three-component data sets with stray light suggest the spectral region from 220 to 228 nm should be avoided; however, weighted PCR gave no improvement in SEP compared to PCR and PLS. The regression vector plots were difficult to interpret because the shape of the confidence intervals and regression vectors changed significantly when the number of principal components used in the regression was changed.

For the simulated data sets with concentration-dependent peak shifts and concentration-dependent bandwidth changes, higher order eigenvectors ($k > 2$ or $k > 3$) contained features that resembled derivative-like spectral features of the pure component spectra. These higher order factors allowed PCR and PLS to give useful calibration models, with PLS frequently giving slightly better performance than PCR. For the simulated data sets with bandwidth changes, quadratic PCR gave slightly better results than PCR and PLS. Weighted PCR results for the data sets with peak shifts and bandwidth changes were generally worse than the results for unweighted PCR and PLS. Examination of the weight vectors for these data sets was difficult because changing the number of factors in the PCR model resulted in significant changes in the regression vectors and confidence intervals. It appeared from plots of the weight vectors that nonlinear response could not be isolated to a specific region of the calibration spectra.

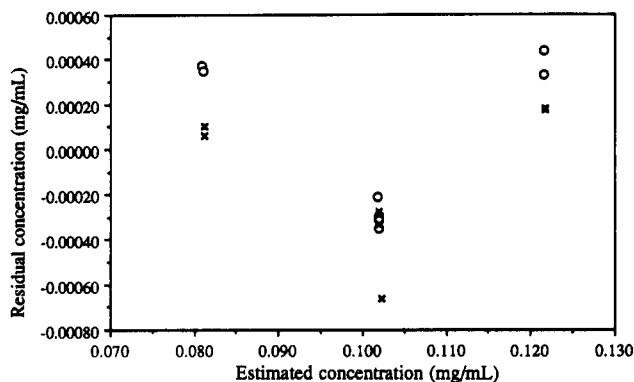


Figure 10. Plot of the residual concentration of benzyl alcohol versus the estimated concentration of benzyl alcohol for the two-component experimental injectable product data set using a two-factor PCR calibration model: calibration standards (O); test standards (x). Pronounced curvature due to nonlinear response can be observed in the residuals.

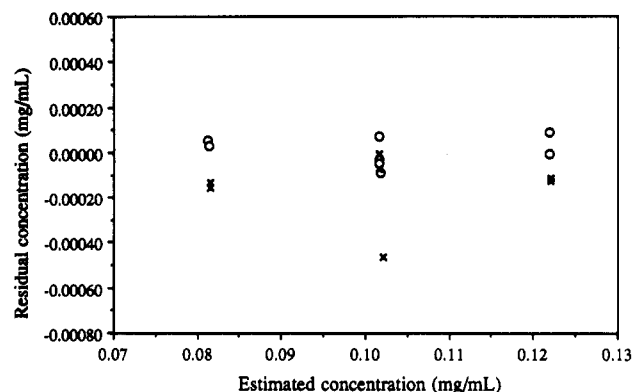


Figure 11. Plot of the residual concentration of benzyl alcohol versus the estimated concentration of benzyl alcohol for the two-component experimental injectable product data set using a PC-ANN calibration model with two principal components and two hidden nodes with linear and sigmoid transfer functions, respectively: calibration standards (O); test standards (x). Compared to Figure 10, the curvature in the residuals has been significantly decreased.

Even though useful calibrations could be obtained for nonlinear data sets with PCR and PLS, pronounced curvature was observed in many of the plots of the residual concentration (actual-estimated) versus estimated concentration. For example, Figure 10 shows a plot of the residual concentration of benzyl alcohol versus the estimated concentration of benzyl alcohol for the experimental injectable product using a two-factor PCR calibration model. In PC-ANN, calibration models were automatically determined that were able to compensate for curvature like that seen in Figure 10. For example, Figure 11 shows a plot of the residual concentration of component B for the same data set after calibration using PC-ANN with two nodes in the hidden layer, one linear and one sigmoid. The residuals in Figure 11 are much smaller, have a better random distribution, and have less curvature compared to Figure 10.

We previously reported the linear calibration results for the experimental two-component injectable product (1) and observed nonlinear response that we presumed to be due to interaction between the high molecular weight active ingredient and the preservative, benzyl alcohol, which helps solubilize the active ingredient. Quadratic PCR offers a slight improvement in SEP compared to the results from PCR and PLS. PC-ANN gives consistently lower calibration errors and prediction errors for this experimental data set.

Calibration results for the Actifed syrup data were also previously reported (13). The test set samples have large spectral residuals in the region between 202 and 216 nm where triprolidine and pseudoephedrine absorb most strongly. We

believe the large spectra residuals arose because of the low light throughput of the diode array spectrophotometer in this wavelength range. Propagation of the large spectral residuals at these wavelengths gives small weights in weighted PCR; consequently, weighted PCR gave good results for these two components. Overall, PC-ANN gave the best results for all four components with prediction errors substantially lower than the other methods.

CONCLUSIONS

Several general conclusions can be drawn from the results of this study: (1) When very low noise data is available, linear calibration techniques like PCR and PLS can give good approximations to all types of simulated nonlinearity by incorporating extra factors in the calibration model. These extra factors can help achieve local linearization of nonlinear response surfaces. We usually found PLS to give better performance than PCR, especially for nonlinear calibration problems. (2) Subtle nonlinear behavior cannot be modeled with extra factors by linear calibration methods like PCR and PLS when sufficient noise is present to perturb or mask the extra factors. (3) When nonlinear response or poor signal-to-noise ratios can be isolated in discrete spectral regions, weighting factors can improve the predictive performance of PCR; otherwise, weighted PCR gives slightly poorer performance than unweighted PCR. Plots of regressions vectors from PCR and their corresponding confidence intervals are difficult to interpret unless isolated spectral regions of nonlinear response or poor signal-to-noise ratios like those described in (3) are present. (4) Under the conditions described in (2) (above) nonlinearity can be accommodated by second-order calibration techniques like quadratic PCR or nonlinear calibration techniques like artificial neural networks provided the method can give a better approximation to the curved response surface than linear calibration methods. (5) Artificial neural networks having the appropriate architecture can be used to develop linear calibration models that perform as well as linear calibration models developed by PCR or PLS. (6) Artificial neural networks can be used to develop nonlinear calibration models that perform slightly better than PCR or PLS when various real and apparent sources of nonlinear response are present. (7) For spectroscopic calibration applications and other applications with highly correlated response variables, orthogonal transformation of the response variables can significantly improve the training speed and the overall precision obtained from artificial neural networks.

The promising results obtained in this study are sufficient to justify continued efforts to develop additional applications of spectroscopic calibration using artificial neural networks; however, before the technique gains widespread acceptance, methods for automatically determining the optimum network configuration must be developed. Refinements in training algorithms for artificial neural networks can be expected to improve its usefulness in calibration problems.

LITERATURE CITED

- Gemperline, P. J.; Salt, A. J. *Chemom.* 1989, 3, 343-357.
- Frank, I. E.; Kowalski, B. R. *Anal. Chem.* 1982, 54, 232R-243R.
- Delaney, M. F. *Anal. Chem.* 1984, 56, 261R-277R.
- Ramos, L. S.; Beebe, K. R.; Carey, W. P.; Sanchez, E.; Erickson, B. C.; Wilson, B. E.; Wangen, L. E.; Kowalski, B. R. *Anal. Chem.* 1986, 58, 294R-315R.
- Brown, S. D.; Barker, T. Q.; Larivee, R. J.; Monfre, S. L.; Wilk, H. R. *Anal. Chem.* 1988, 60, 252R-273R.
- Brown, S. D. *Anal. Chem.* 1990, 62, 84R-101R.
- Naes, T. *Technomet.* 1985, 27, 301-311.
- Phelan, M. K.; Barlow, C. H.; Kelly, J. J.; Jinguji, T. M.; Callis, J. B. *Anal. Chem.* 1989, 61, 1419-1424.
- Carey, W. P.; Wangen, L. E. *Anal. Chem.* 1989, 61, 1667-1669.
- Kelly, J. J.; Barlow, C. H.; Jinguji, T. M.; Callis, J. B. *Anal. Chem.* 1989, 61, 313-320.
- Gemperline, P. J. *J. Chemom.* 1989, 3, 549-568.
- Lorber, A.; Wangen, L. E.; Kowalski, B. R. *J. Chemom.* 1987, 1, 19-31.

- (13) Long, J. R.; Gregoriou, V. G.; Gemperline, P. J. *Anal. Chem.* **1990**, *62*, 1791-1797.
- (14) Thomsen, J. U.; Meyer, B. J. *Magn. Reson.* **1989**, *84*, 212.
- (15) Meyer, B.; Hansen, T.; Nute, D.; Albersheim, P.; Darbill, A.; York, W.; Sellers, J. *Science* **1991**, *251*, 542-544.
- (16) Robb, E. W.; Munk, M. E. *Mikrochim. Acta* **1990**, *1*, 131-155.
- (17) Wythoff, B. J.; Levine, S. P.; Tomellini, S. A. *Anal. Chem.* **1990**, *62*, 2702-2709.
- (18) Elrod, E. W.; Maggiora, G. M.; Trenary, R. G. *J. Chem. Inf. Comput. Sci.* **1990**, *20*, 477-484.
- (19) Jansson, P. A. *Anal. Chem.* **1991**, *63*, 357A-362A.
- (20) Isaksson, T.; Naes, T. *Appl. Spectrosc.* **1988**, *43*, 1273-1284.
- (21) Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley and Sons, Ltd.: Chichester, U.K., 1989.
- (22) Naes, T.; Isaksson, T.; Kowalski, B. R. *Anal. Chem.* **1990**, *62*, 664-673.
- (23) Koscielniak, P.; Parczewski, A. *Anal. Chim. Acta* **1985**, *177*, 197-202.
- (24) Frank, I. E.; Lanteri, S. *Chemom. Intell. Lab. Syst.* **1988**, *3*, 301-313.
- (25) Bratos, S.; Lascombe, L.; Novak, A. In *Molecular Interactions*; Ratajczak, H.; Orvill-Thomas, W. J., Eds.; John Wiley and Sons, Ltd.: Chichester, U.K., 1980.
- (26) Nyquist, R. A. *Appl. Spectrosc.* **1990**, *44*, 438-442.
- (27) Nyquist, R. A. *Appl. Spectrosc.* **1990**, *44*, 426-433.
- (28) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*, 2nd ed.; Prentice Hall: Englewood Cliffs, NJ, 1988.
- (29) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, MA, 1986.

RECEIVED for review March 20, 1991. Accepted July 16, 1991.
We thank the Burroughs Wellcome Co. for financial and in-kind support of this research.

Fourier Transform Infrared Spectroscopic Method for the Quantitative Trace Analysis of Transition-Metal Carbonyl-Labeled Bioligands

Michèle Salmain, Anne Vessières, and Gérard Jaouen*

Ecole Nationale Supérieure de Chimie de Paris, UA CNRS 403, 11 rue Pierre et Marie Curie, 75231 Paris, Cedex 05, France

Ian S. Butler*

Department of Chemistry, Otto Maass Building, McGill University, 801 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6

A quantitative FT-IR spectroscopic method has been developed for the trace analysis in chlorinated organic solvents of transition-metal carbonyl-labeled bioligands. In order to illustrate the widespread analytical potential of the method, three derivatives of the female hormonal steroid 17β -estradiol, containing $\text{Cr}(\text{CO})_3$, $\text{Cp}_2\text{Mo}_2(\text{CO})_4$ ($\text{Cp} = \eta^5\text{-C}_5\text{H}_5$), and $\text{Co}_2(\text{CO})_8$ as labels, and the anticonvulsant drug phenobarbital, labeled with $(\eta^5\text{-C}_5\text{H}_5)\text{Mn}(\text{CO})_3$, were examined. The cobalt carbonyl marker proved to be the best suited for quantitative analysis purposes, and the minimum tracer quantity detectable for this particular marker (64 scans, 4-cm^{-1} resolution, 3.5 min) was optimized in CCl_4 solution at about 300 fmol (or 0.3 pmol, 180 pg) by using an ultralow volume (23.0 μL), gold light-pipe IR solution cell and a liquid nitrogen cooled, InSb (indium antimonide) IR detector. The repeatability of this radically different analytical procedure over the concentration range 1.0×10^{-6} to 5.0×10^{-8} M was good (coefficient of variance $\leq 6\%$) and the method provides the basis for a new immunological test—carbonylmetal-immunoassay (CMIA).

INTRODUCTION

The application of Fourier transform methods to IR spectroscopy has led to this technique becoming an extremely sensitive and valuable quantitative microanalytical tool (1, 2). When a gas-phase chromatography/matrix-isolation apparatus is coupled to a FT-IR spectrometer, the minimum identifiable quantity for an organic molecule is about 250 pg (3), and recently, on-line GC/FT-IR spectra have been obtained for injected quantities of *n*-alkanes, even as low as 50 pg (4). The latter analyses were performed by trapping the *n*-alkane eluates as small spots directly onto a moving mirror

held at 77 K, which was then passed through a 100- μm square IR beam. In the biomedical field, FT-IR spectroscopy has provided some important insights into the structure and function of proteins (5), and FT-IR microscopy has been successfully employed to study bone calcification (6). An unusual new biomedical application of FT-IR spectroscopy has just been announced that involves measurement of the high-pressure (pressure-tuning) spectra of nanogram sections of intact human colonic tissue. The distinct spectral differences observed between normal and malignant tissue suggest that this qualitative technique may become diagnostically useful for colonic and other cancers (7). FT-IR spectroscopy has also proved quantitatively important in clinical chemistry, e.g., for determining the concentrations of various components in human blood plasma such as total protein, glucose, triglycerides, total cholesterol, urea, and uric acid (8, 9). Surprisingly, however, despite the growing interest in the application of FT-IR spectroscopy in the biomedical area, its quantitative use at the microanalytical level has yet to be demonstrated.

In immunoassay work radiochemical techniques are usually the quantitative method of choice (10). However, because of problems such as licensing and the general fear of the public of radioactivity and the nuclear industry, there is currently an intense effort to develop new, sensitive, nonradioisotopic methods for clinical analysis. Any alternative analytical procedure must be capable of functioning quantitatively in the picomole range in order to be competitive with existing radioisotopic technologies. Several techniques are being actively promoted at the present time, among them, chemiluminescence (11) and time-resolved fluorescence spectroscopy (12). We have chosen to direct our efforts in the clinical analysis field in a totally different direction, viz., toward the application of FT-IR spectroscopy in the detection of transition-metal carbonyl fragments as nonradioactive tracers for