

# Prediction of Molar Fractions in Two-Component Gas Mixtures Using Pulse-Echo Ultrasound and PLS Regression

Johan E. Carlson and Rolf Carlson

**Abstract**—The composition, and thus the energy content and monetary value of natural gas and biogas, vary considerably depending on the source. Present energy measurement techniques are not suitable for online use.

We show with experiments on mixtures of ethane ( $C_2H_6$ ) and oxygen ( $O_2$ ) how partial least squares regression (PLSR) can be used to predict the molar fraction of ethane in the mixtures, given spectral data from ultrasonic pulse-echo measurements. The PLSR technique is compared with the standard principal component regression (PCR), and we show that PLSR yields better predictive performance.

## I. INTRODUCTION

NATURAL gas contains a mixture of several gases, such as methane, ethane, hydrogen, etc. Sometimes also small, but highly undesired, fractions of oxygen. With the use of different sources of gas (e.g., different gas fields) the energy content of the gas delivered to customers may vary considerably. Variations up to 20% are common [1]. It is of interest to both provider and customer to know the composition of such gas mixtures as this determines the energy content (calorimetric value) and, thus, the combustion properties and the monetary value of the gas.

Today, the energy content of gases is measured using either gas chromatography or calorimetry. Both methods are accurate, but they require samples of the gas to be removed and analyzed separately, which makes the measurement slow. They also are relatively expensive. Because of this, the existing techniques are not suitable for on-line measurement at the customer side of the distribution line.

Ongoing research focuses on both physical modeling of wave propagation in gases [2]–[4] and on multivariate statistical analysis of measured ultrasonic data [5], [6]. The physical models suggest that phase velocity and frequency-dependent attenuation will change with the composition of the gas mixtures. In this paper we focus on exploring these changes using multivariate tools.

In this paper we show how partial least squares regression (PLSR) can be used to estimate molar fractions of

ethane in mixtures of ethane and oxygen, for a wide pressure range. PLSR has been applied in many experimental sciences, from chemometrics [7] to social sciences [8], where the goal is to explore relationships between a set of descriptors (variables) and a set of measured responses. The results of the PLSR are compared with those obtained by principal component regression (PCR).

The paper is organized as follows: Section II describes the theory behind PLSR and PCR and highlights some of the fundamental differences. Section III describes the measurement configuration and how data sets are prepared in order to evaluate the prediction models. Experimental results on mixtures of ethane and oxygen are given in Section IV.

## II. THEORY

Assume that we have a series of gas mixtures for which the relative amounts of the gases vary. Assume also that some measurement procedure has been used to characterize properties of these gas mixtures, and that each measurement yields a series of measured property variables. Assume also that there are linear relationships between the measured properties and the composition of the gas mixture. The goal of any regression tool is to find and explore these relationships.

### A. Physical Background

Before going into details about the PLSR, we first need to describe some of the physics of the gases and which properties can be expected to affect the wave propagation.

When sound propagates through gases, basically two properties can be directly observed: speed of sound and attenuation. Both of these are frequency dependent. They depend on viscous, molecular, and thermal relaxation phenomena that occur when molecules collide. Describing the details of underlying physics leading to changes in speed of sound and attenuation is beyond the scope of this paper, but they can be found in previous work [2]–[4], [9]. Good physical models are crucial for the understanding of the problem, so that proper instrumentation can be designed. Note that, even if a good model of the underlying physics is available, the parameters of such model are not necessarily identifiable from the observed ultrasound data. Furthermore, the end goal is not to describe the model, but

Manuscript received November 10, 2004; accepted September 11, 2005.

J. E. Carlson is with Embedded Internet Systems Laboratory (EIS-LAB), Department of Computer Science and Electrical Engineering, Luleå University of Technology, SE-971 87 Luleå, Sweden (e-mail: Johan.Carlson@ltu.se).

R. Carlson is with the Department of Chemistry, Tromsø University, N-9037 Tromsø, Norway.

to measure some implicit property (e.g., energy content or molar fractions). The choice of model thus depends on the goal of the study.

In this paper we demonstrate how variations in ultrasound spectra can be connected to molar fractions in gas mixtures. As will be shown, the PLSR model captures most of the variation from the measurements that contain information about the molar fraction.

*B. Motivation*

In this section we give a brief overview of common regression tools used to quantify relationships between physical variables and measured responses. There are some common problems with these techniques, and these are highlighted as a motivation for the use of the PLSR technique.

Let each row in the matrix  $\mathbf{X}$  be the measured ultrasound properties for one gas mixture, and let the corresponding row in  $\mathbf{Y}$  be the composition of the mixture. Section III-B describes which properties are used and how they are extracted from the measured ultrasound pulses.

For a multiple linear regression (MLR) model, the relationship between the descriptors  $\mathbf{X}$  and the responses  $\mathbf{Y}$  can be written as:

$$\mathbf{X}\mathbf{A} = \mathbf{Y}, \tag{1}$$

where  $\mathbf{A}$  is a matrix with the model coefficients so that  $\mathbf{X}\mathbf{a}_i = \mathbf{y}_i$ . If  $\mathbf{X}$  has full column rank (i.e., the columns are linearly independent), a least-squares estimate,  $\hat{\mathbf{A}}$ , of  $\mathbf{A}$  is obtained by:

$$\hat{\mathbf{A}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{2}$$

If there is any linear dependence between the columns of  $\mathbf{X}$ , the inverse  $(\mathbf{X}^T\mathbf{X})^{-1}$  will not exist. The most common remedy for this problem is called PCR. First a principal component analysis (PCA) [10] is made on  $\mathbf{X}$ . The significant principal components then are used to form the pseudo-inverse of  $\mathbf{X}$ . The resulting solution is identical to the least-squares solution obtained by the Moore-Penrose pseudo-inverse [11]. This is sometimes a powerful approach, but there is an underlying assumption that is not always fulfilled. We cannot be sure that large variations within  $\mathbf{X}$  necessarily is of greater importance when it comes to describing  $\mathbf{Y}$ . That is, the principal components that are discarded might contain valuable information about the cross-covariance between  $\mathbf{X}$  and  $\mathbf{Y}$ .

Another approach is to use regularization (often referred to as ridge regression), that is to add a diagonal matrix to  $\mathbf{X}$  (i.e.,  $\tilde{\mathbf{X}} = \mathbf{X} + \mu\mathbf{I}$ ) then form  $(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\mathbf{Y}$  to obtain a regularized least-squares solution for the linear relationship between the X block and the Y block. In this way, no columns of  $\mathbf{X}$  are discarded, but at the cost of introducing a bias. For good performance, finding the regularization parameter  $\mu$  is crucial [12].

The following sections describe the calculations used in PLSR and PCR. In Section IV the two methods are compared for measurements on two-component gas mixtures.

*C. PLSR*

The previous section highlighted some of the problems with the most common regression methods. One way to overcome these is by using a method called PLS (partial least squares or projections onto latent structures).

The PLS method was developed by Wold [13] and has been applied to many areas in experimental sciences. In this section we will give a brief description of the PLS method. An extensive overview can be found in [7].

The central idea of PLS, as opposed to PCR and ordinary least-squares (OLS) estimation, is that the PLS determines a set of basis functions (PLS components) for both the X block and the Y block, in such a way that they best describe the cross covariance of the blocks.

*D. Calculating PLS Components*

Determining the PLS components is an iterative process. The first step is to calculate the cross-covariance matrix  $\mathbf{C}_{YX}$ , as:

$$\mathbf{C}_{YX} = \mathbf{Y}^T\mathbf{X}, \tag{3}$$

where  $\mathbf{X}$  is an  $N \times K$  matrix and  $\mathbf{Y}$  is an  $N \times L$  matrix. Find the first principal component of  $\mathbf{C}_{YX}$  (i.e., the eigenvector of  $\mathbf{C}_{YX}^T\mathbf{C}_{YX}$ ) corresponding to the largest singular value of  $\mathbf{C}_{YX}$  [10]. This vector, also called weight vector, is denoted  $\mathbf{w}_1$ . The first score vector,  $\mathbf{t}_1$  ( $N \times 1$ ) is then:

$$\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1, \tag{4}$$

and the corresponding loading vector for the  $\mathbf{X}$  block,  $\mathbf{p}_1$  ( $K \times 1$ ), is given by:

$$\mathbf{p}_1 = \mathbf{X}^T\mathbf{t}_1 / (\mathbf{t}_1^T\mathbf{t}_1). \tag{5}$$

For the Y block, the first loading vector,  $\mathbf{q}_1$  ( $L \times 1$ ), is given by:

$$\mathbf{q}_1 = \mathbf{Y}^T\mathbf{t}_1 / (\mathbf{t}_1^T\mathbf{t}_1). \tag{6}$$

The residuals of the first iteration  $\mathbf{E}_1$  and  $\mathbf{F}_1$  for the X block and Y block, respectively, are then:

$$\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^T, \tag{7}$$

$$\mathbf{F}_1 = \mathbf{Y} - \mathbf{t}_1\mathbf{q}_1^T. \tag{8}$$

To find the next PLS component, the steps above are repeated with  $\mathbf{E}$  and  $\mathbf{F}$  as the new starting matrices. This procedure then is repeated until the desired number of components,  $A$ , has been determined. The scores and loadings of the X and Y blocks then are stored as columns in the matrices  $\mathbf{T}$ ,  $\mathbf{P}$ , and  $\mathbf{Q}$ , as:

$$\mathbf{T} = [\mathbf{t}_1 \ \mathbf{t}_2 \ \cdots \ \mathbf{t}_A], \tag{9}$$

$$\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_A], \tag{10}$$

$$\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_A]. \tag{11}$$

The weight vectors,  $\mathbf{w}$ , from the different iterations are stored in the matrix  $\mathbf{W}$  ( $K \times A$ ).

### E. Prediction of the Y Block

The original matrices  $\mathbf{X}$  and  $\mathbf{Y}$  can now be expressed in terms of the PLS components as:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}, \quad (12)$$

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{F}, \quad (13)$$

where  $\mathbf{E}$  and  $\mathbf{F}$  are residual matrices of the X and Y blocks, respectively.

Given an existing set of PLS components, a prediction of  $\mathbf{Y}$  is obtained by [7]:

$$\hat{\mathbf{Y}} = \mathbf{X}\tilde{\mathbf{W}}\mathbf{Q}^T, \quad (14)$$

where:

$$\tilde{\mathbf{W}} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}. \quad (15)$$

### F. Geometrical Interpretation of the PLS Components

As mentioned earlier, the PLS algorithm determines a set of components for the X and Y block that best describes the covariance between the blocks, in a least-square sense. The first step of the algorithm, determining the sample cross-covariance matrix  $\mathbf{C}_{YX} = \mathbf{Y}^T\mathbf{X}$  projects  $\mathbf{X}$  onto what is common between the X and Y blocks. The first weight vector,  $\mathbf{w}_1$ , is a principal component (i.e., a singular vector) of  $\mathbf{C}_{YX}$ , which means it represents the direction that describes most of the cross covariance. The first score and loading vectors of the X block,  $\mathbf{t}_1$  and  $\mathbf{p}_1$ , are projections of  $\mathbf{X}$  onto that weight vector. This means that they represent variation in  $\mathbf{X}$  that best describes the cross covariance between the blocks. The scores and loadings of the Y block ( $\mathbf{u}_1$  and  $\mathbf{q}_1$ ) are projection of  $\mathbf{Y}$  onto what is already described by the first PLS component of  $\mathbf{X}$ . The X and Y blocks then are deflated, and the residuals are used as input to the consecutive steps of the algorithm.

### G. Determining PLS Model Order

As in all linear modeling, finding an appropriate model order is an important, and difficult, problem. Numerous criteria are available for model order selection, e.g., the Akaike information criterion [14] and cross validation [15]. Some of the available methods tend to overestimate the model order, and others are overly conservative [16].

For PLS modeling problems, cross validation is the most commonly used method for model order estimation.

In this paper we have one training data set and one validation set. We estimate a PLS model from the training set, and we evaluate by looking at the prediction error sum of squares (PRESS) for the validation set as a function of the number of components in the PLS model. For the experiments presented in this paper, the PRESS is shown in Fig. 1.

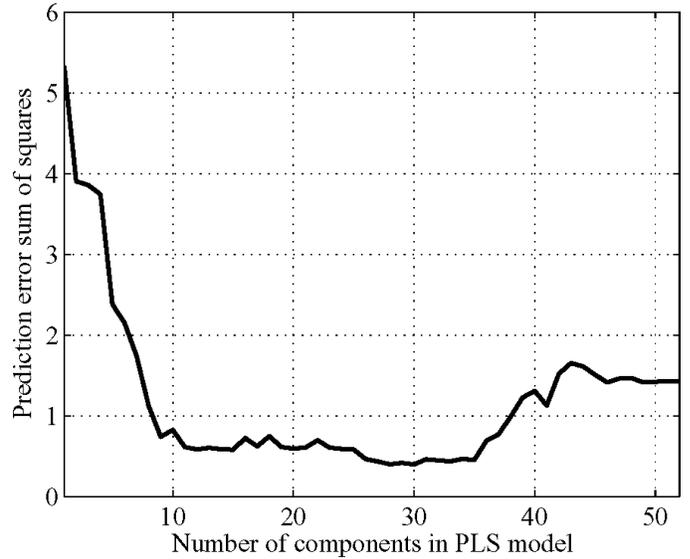


Fig. 1. Prediction error sum of squares as a function of the number of PLS components used in the model. Using more than 12 components does not improve the model significantly.

### H. Principal Component Regression

When  $\mathbf{X}$  does not have full rank, the ordinary least-squares solution in (2) does not exist because  $\mathbf{X}^T\mathbf{X}$  is not invertible. A common solution is to use PCR. This is the same as finding the Moore-Penrose pseudo-inverse of  $\mathbf{X}$ , denoted  $\mathbf{X}^\dagger$  and estimating the coefficients as:

$$\hat{\mathbf{A}}_{PCR} = \mathbf{X}^\dagger\mathbf{Y}. \quad (16)$$

For details on how to calculate  $\mathbf{X}^\dagger$ , see [11].

To determine the model order (i.e., the number of principal components to use), we examine how much of the total experimental variation (in percent) each component explains. This is done by studying the scaled singular values of  $\mathbf{X}$  because the singular value  $\sigma_i$  is proportional to the experimental variation in the direction of the  $i$ :th principal component. Let  $\bar{\sigma}_i$  be scaled versions of the  $r$  nonzero singular values  $\sigma_i$ , as:

$$\bar{\sigma}_i^2 = 100 \frac{\sigma_i^2}{\sum_{k=1}^r \sigma_k^2}. \quad (17)$$

Fig. 2 shows the cumulative sum of (17) for the gas mixtures evaluated in this paper.

### I. Prediction of Molar Fractions

This section describes the preprocessing of experimental data and how they are arranged prior to performing the PLS modeling. The actual measurement configuration and data acquisition is described in Section III.

Ultrasound pulses transmitted through a gas mixture are affected primarily by frequency-dependent attenuation and frequency-dependent sound velocity (i.e., dispersion). This affects the shape of the pulse in several ways. The

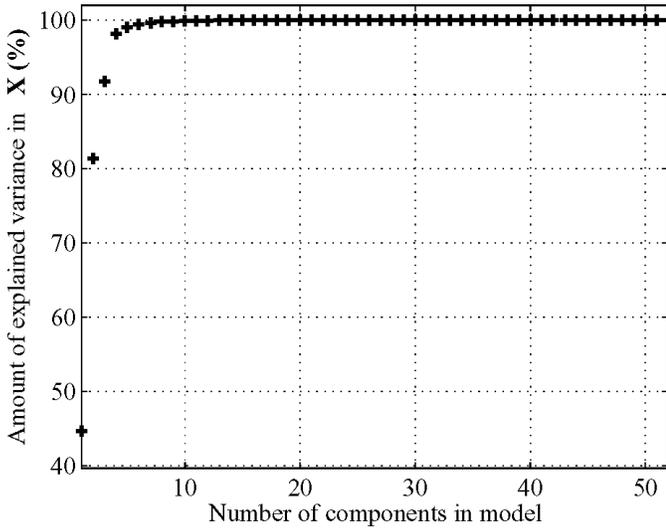


Fig. 2. Amount of explained variance of the  $\mathbf{X}$  block as a function of the number of components used for PCR. The first five components account for approximately 99% of the total variation of the  $\mathbf{X}$  block.

aim of this study is to explore how these variations in pulse shape can be used to predict the composition of the gas mixture. Because of the varying propagation delay through the gas for different compositions simply stacking the measured time signals as rows in the matrix  $\mathbf{X}$  does not work. An alternative is to study how the spectral properties of the pulses are affected. That is, let the variables of the  $\mathbf{X}$  block be the amplitude and phase of the pulse for different wavelengths. The  $\mathbf{Y}$  block then is made up of the variables we wish to predict, in this case the molar fractions of the constituent gases. In the case of a two-component mixture, the  $\mathbf{Y}$  block reduces to one vector consisting of values between zero and one, corresponding to the molar fraction of ethane in the mixture. It should be noted here that it is possible to extend the  $\mathbf{Y}$  block to contain also pressure, temperature, etc.

Two sets of data were prepared (see Section III-B), one for model estimation (referred to as the training set) and one for model validation (referred to as the validation set).

### III. EXPERIMENTS

#### A. Setup

For all the experiments in this paper, the ultrasound echoes were measured in a pulse-echo configuration (see Fig. 3). An ultrasound transducer was mounted in a measurement cell, transmitting pulses through the gas toward a stainless steel reflector. The acoustic properties of interest vary with frequency,  $f$ , and pressure,  $P$ . The temperature and frequency dependence on sound velocity and acoustic attenuation was previously investigated by, for example, Martinsson and Delsing [4], Dain and Lueptow [2], [3], and Bhatia [9]. Changes in frequency and pressure normally are studied as the ratio  $f/P$  [2]. There are two ways to vary this in a pulse-echo system. Transducers of

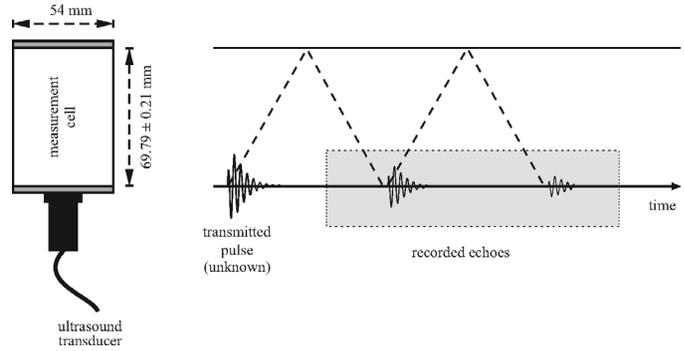


Fig. 3. The measurement cell and the pulse-echo principle. The transducer emits an unknown sound wave. The reflection from the bottom of the measurement cell then is recorded. Consecutive echoes can be recorded, as indicated in the figure. In this paper, however, only the first echo is used.

different center frequency can be used for a fixed pressure, or one transducer is used while the pressure is changed. The latter of the two principles was chosen for the work in this paper. A 1 MHz air transducer was used while the static pressure was varied. The effective diameter of the transducer was 15 mm. For this transducer and the propagation distance (c.f. 2), diffraction losses were assumed to be negligible [17].

A custom-built pressure chamber (see Fig. 4) was used to achieve different static pressures. The pressure was varied between 1.54 to 7.4 bar in 12–18 steps. Because the attenuation in ethane is extremely high at low pressures and high frequencies, we were limited to make measurements at higher pressure for that particular gas (above 1.86 bar).

The pressure in the chamber was measured with an Anderson TPP Pressure Transmitter (Anderson Instrument Company Inc., Fultonville, NY) with a range of up to 13.6 bar above atmospheric zero. The transmitter has an accuracy of approximately 30 mbar. This includes the combined effects of linearity, hysteresis, and repeatability.

The transducer was mounted on a stainless steel measurement cell (see Fig. 3), as seen in the lower left corner of Fig. 4. The measurement cell then was immersed into the pressure chamber. The whole setup then was placed in a temperature-controlled chamber (HT4010, Heraeus Vötsch, Balingen, Germany), keeping the temperature constant at the desired temperature.

To excite and receive acoustic pulses from the transducer, a Panametrics Pulser/Receiver Model 5072 (Panametrics, Waltham, MA) was used. For the transmitting mode, the pulser/receiver was set to deliver maximum energy to the transducer, which corresponds to a short voltage peak with 380 V amplitude with an energy of 104  $\mu\text{J}$ . In receive mode, the pulser/receiver has a built-in amplifier that can amplify the input as much as 59 dB. In all measurements presented here, the gain was set to 20 dB.

All pulses were sampled at 100 MHz with an 8-bit Tektronix TDS 724, 1 GHz digitizing oscilloscope (Tektronix UK Ltd., Bracknell, Berkshire, UK). For each measure-

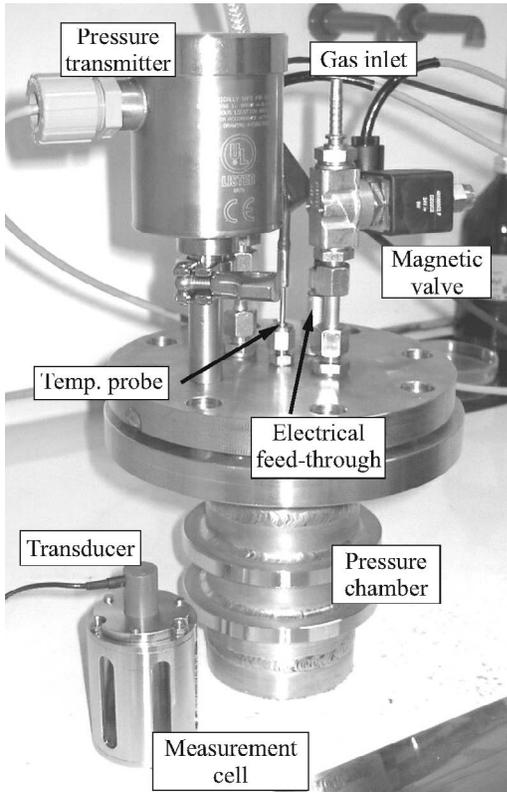


Fig. 4. The measurement equipment. All measurements were performed in a pressure chamber. The pressure was varied between 1.54 bar and 7.4 bar. The air transducer has a center frequency of 1 MHz.

ment, the temperature was recorded using an encapsulated PT100 sensor (Pentronic AB, Gunnebo, Sweden) mounted through the wall of the pressure chamber.

The transducer used was originally designed for operation in air with an acoustic impedance  $Z_{\text{air}} = 4154 \text{ Pa}\cdot\text{s}/\text{m}$  (at  $T = 20^\circ\text{C}$  and  $P = 1 \text{ bar}$ ). Because the acoustic impedance of the gases used are different from air, the transducer will not operate at its optimal performance; consequently, it would not transmit as much power as desirable.

### B. Measurements and Preprocessing

Measurements were made for pure oxygen ( $\text{O}_2$ ) and pure ethane ( $\text{C}_2\text{H}_6$ ), as well as mixtures of the two for molar fractions of ethane of 20%, 40%, 60%, and 80%. The temperature was kept constant at  $20^\circ\text{C}$  during all measurements. The static pressure was varied from 1.54 bar and 7.4 bar, in 12 steps.

For each experimental setting, 50 pulses were recorded. The first 25 were averaged and used as training data. An example of such a pulse is shown in Fig. 5 (top). The second 25 pulses also were averaged and used as validation data. That is, all models were calculated based on the training set, and their predictive performance was evaluated by predicting the withheld validation data set. For each of the averaged pulses,  $\mathbf{p}_i^T$ , the corresponding row in the matrix  $\mathbf{X}$  was obtained as follows:

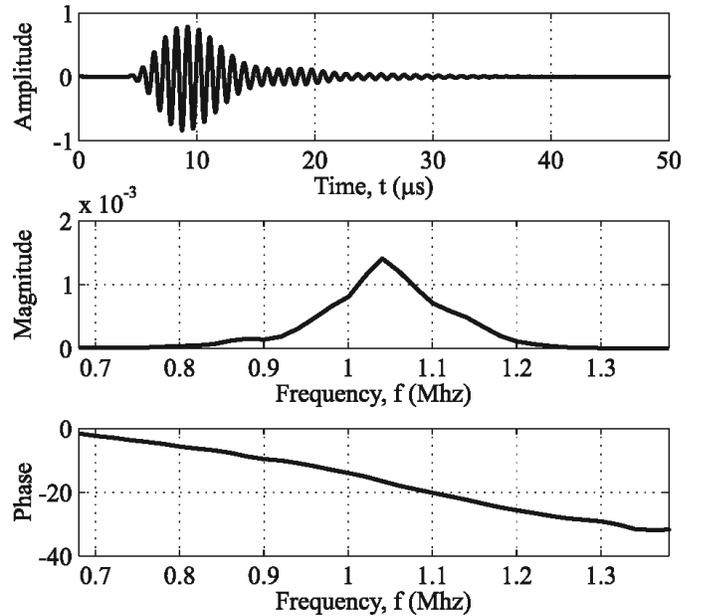


Fig. 5. One averaged pulse and the corresponding spectrum.

- Calculate spectral representation of the pulse, using the discrete Fourier transform (DFT).
- Calculate magnitude and unwrapped phase of the spectrum.
- Extract the frequency components around the center frequency that has sufficient signal-to-noise ratio. The extracted spectrum is shown in Fig. 5 (middle and bottom).
- Store as a row in the matrix  $\mathbf{X}$ , i.e.:

$$\mathbf{x}_i^T = \left[ \text{abs}(P_i(\omega)) : \text{arg}(P_i(\omega)) \right]. \quad (18)$$

Before the PCR and PLS models were calculated, the columns of both the  $\mathbf{X}$  and the  $\mathbf{Y}$  block were scaled to unit variance and centered to have zero mean. The reason for this is that we do not want to a priori assess any greater importance to any of the frequency components. This is a standard operation in most multivariate techniques. We also tested the technique without scaling the amplitude spectra (i.e., only centering), but this decreased the predictive performance of the PLS model.

## IV. RESULTS

Once the training and validation data sets were prepared, as described in the previous section, we need to determine the order of the linear models. Fig. 2 shows the cumulative sum of (17). The figure shows that five components are enough to describe approximately 99% of the experimental variation within the  $\mathbf{X}$  block. Figs. 6 and 7 show the true versus the predicted molar fraction of ethane in the mixtures. Although 99% of the total variation in the  $\mathbf{X}$  block is retained in the pseudo inverse, the predictive performance of both the PCR and the PLSR is poor.

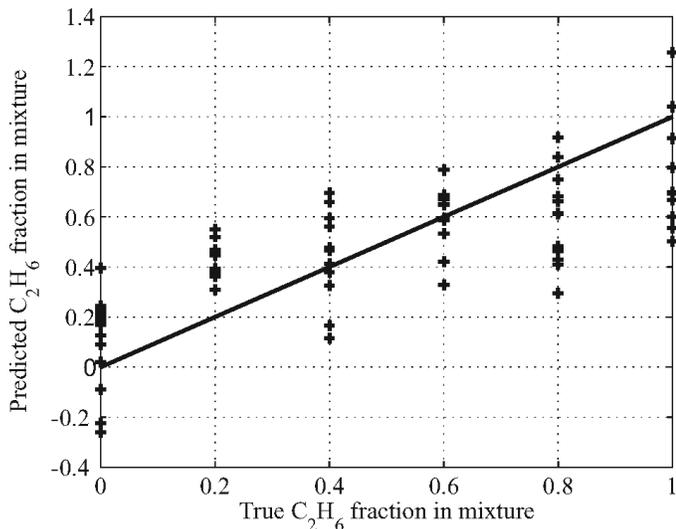


Fig. 6. Principal component regression using five components. The figure shows observed versus predicted molar fraction of ethane. The solid line represents the optimal predictor.

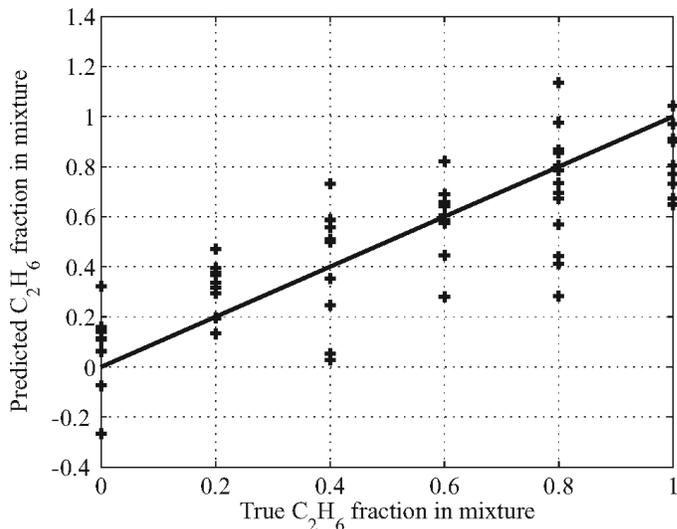


Fig. 8. Principal component regression using 12 components. The figure shows observed versus predicted molar fraction of ethane. The solid line represents the optimal predictor.

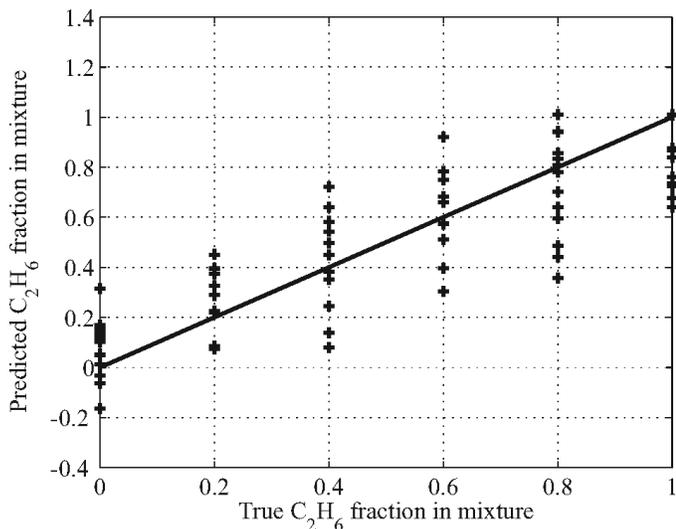


Fig. 7. PLS regression using five components. The figure shows observed versus predicted molar fraction of ethane. The solid line represents the optimal predictor.

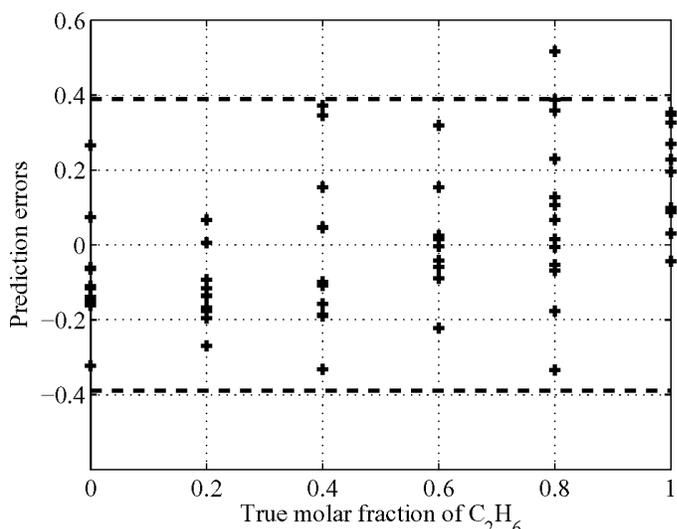


Fig. 9. Principal component regression using 12 components. The figure shows the prediction errors and the corresponding 95% confidence interval.

Fig. 1 shows the PRESS as a function of the number of PLS components. Guided by this, we decided that 12 components should be enough to describe most of the cross covariance between the X and the Y block.

Figs. 8 and 9 show the predictions and the prediction errors, respectively, for PCR, using 12 components. The dashed lines in Fig. 9 represent the 95% confidence interval for the prediction error. The prediction errors are plotted against molar fraction. Fig. 9 shows that the prediction error variance increases with increasing molar fraction of ethane. The main reason for this is that the signal-to-noise ratio decreases when the molar fraction of ethane increases.

Figs. 10 and 11 show the predictions and the prediction errors, respectively, for the PLS regression, using 12 com-

ponents. The dashed lines in Fig. 11 represent the 95% confidence interval for the prediction error. Comparing Fig. 9 and Fig. 11, we see that there are no apparent systematic variations in the prediction errors for the PLS model. This indicates that the PLS model is better at capturing the variations in the X block that correlate to variations in molar fractions than the PCR model. Fig. 12 shows the same prediction errors plotted against pressure. No systematic variation is visible here either, which indicates that the prediction errors are not correlated with either molar fraction or pressure.

Note that for the PCR, increasing the model from 5 to 12 components does not significantly improve its predictive performance. Using the complete X block, we arrive at the multiple linear regression estimate, which performance

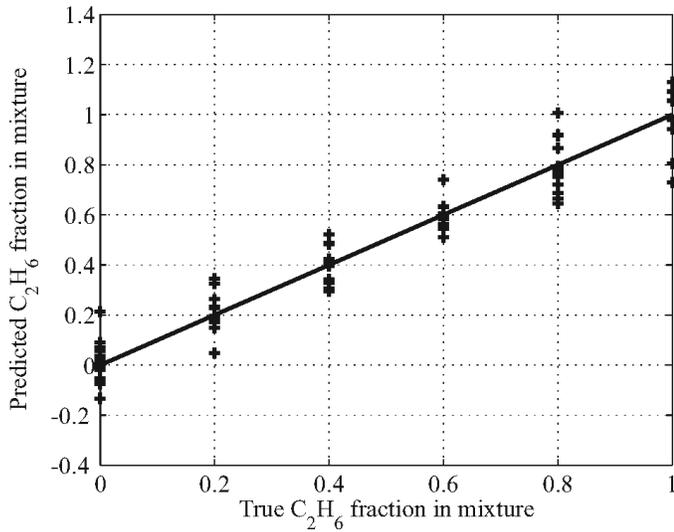


Fig. 10. PLS regression using 12 components. The figure shows observed versus predicted molar fraction of ethane. The solid line represents the optimal predictor.

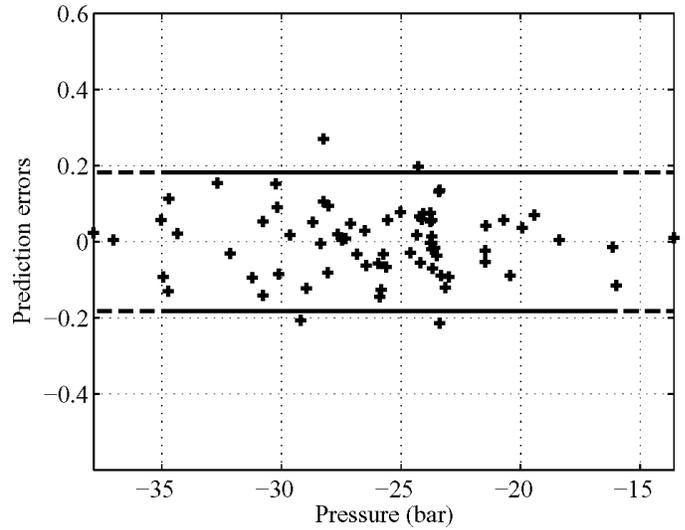


Fig. 12. Prediction errors of the 12 component PLS model versus pressure, and the corresponding 95% confidence interval.

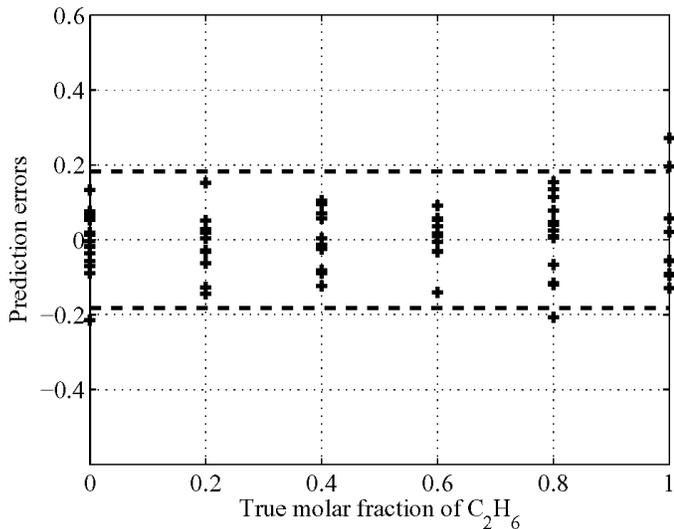


Fig. 11. PLS regression using 12 components. The figure shows the prediction errors and the corresponding 95% confidence interval.

is identical to the PCR results. This is expected because the PCR focuses on capturing as much as possible of the variation within the X block. The five component model already accounted for approximately 99% of this variation. However, the predictive performance of the PLS model is drastically improved when the additional components are added. This suggests that the cross covariance between the X and Y blocks is not connected to large variations in X. In this particular case, PLS regression is recommended over PCR.

It should be noted here that with prediction we mean a withheld replicate of a calibration experiment and not a true prediction in the sense that we predict the outcome of an experiment not present in the calibration run.

## V. DISCUSSION

The results presented in the previous section were obtained using only the first pulse-echo. Because the pulse bounces back and forth through the gas sample several times, it also is possible to estimate attenuation and phase velocity if the second echo also is recorded. For the ultrasound frequencies used in this study, however, the amplitude of the second echo was so low that reliable estimation of attenuation and phase velocity was not possible. It should be noted that even the first echo contains information about these effects, when different compositions are compared. These relative effects, however, are more difficult to relate to the underlying physics. A change of instrumentation in order to better exploit the potential of the ultrasound is being developed for future measurements.

The frequency-dependent attenuation,  $\alpha(\omega)$ , and phase velocity,  $c(\omega)$ , vary with the actual gases in the mixture. For any given mixture, the ultrasound frequencies should be adjusted so that variations in gas compositions which also are detectable as variation in the measured pulses. This is a trade-off between sensitivity and robustness, which will be investigated further in future work.

Although the PLS predictions are more accurate than those obtained using PCR, there is still a significant uncertainty in the results. The results demonstrate the potential of the technique, using only spectral data of one measured pulse-echo. By including more observed properties (e.g., phase velocity, frequency-dependent attenuation), it is likely that the performance is improved.

As pointed out by one of the reviewers, the known pressure could have been included in the X block. This was tested, but the predictive performance did not improve. One explanation to this is that a pressure change is analogous to a frequency shift of the ultrasound, and thus already represented by the magnitude spectra in the X block.

## VI. CONCLUSIONS

In this paper we describe how the principles of PLSR can be used to predict molar fractions of ethane in a mixture of ethane and oxygen from ultrasonic pulse-echo measurements. The results obtained with PLSR are compared to those obtained using a conventional pseudo-inverse or PCR technique. It is clear that the predictive performance of the PLS model exceeds that of the PCR method. The reason is that, although PCR captures as much as possible of the experimental variation of the X block, PLS aims to describe cross covariance between the X and the Y block.

Figs. 8–11 show that the predictions using PLSR have less variance and are less biased than the PCR predictions for high fractions of ethane.

## ACKNOWLEDGMENTS

The authors wish to express their gratitude toward Prof. Veli-Matti Taavitsainen for valuable discussions and to Dr. Pär-Erik Martinsson and Mr. Jesper Martinsson for their help with collecting and preparation of the experimental data. Generous grants from the Kempe Foundation and The Swedish National Energy Administration are also gratefully acknowledged. Financial support to R.C. from The Norwegian Research Council is gratefully acknowledged. The authors also would like to thank the reviewers for their rapid and thorough review, leading to significant improvements of the paper.

## REFERENCES

- [1] J. Delsing and I. Blom, "On-line measurements of energy content of bio gas and natural gas mixtures," Lund Institute of Technology, Tech. Rep. ISSN: 0282-3772, 1995.
- [2] Y. Dain and R. M. Lueptow, "Acoustic attenuation in three-component gas mixtures—Theory," *J. Acoust. Soc. Amer.*, vol. 109, pp. 1955–1964, 2001.
- [3] Y. Dain and R. M. Lueptow, "Acoustic attenuation in three-component gas mixtures—Results," *J. Acoust. Soc. Amer.*, vol. 110, pp. 2974–2979, 2001.
- [4] P.-E. Martinsson and J. Delsing, "Ultrasonic measurements of molecular relaxation in ethane and carbon monoxide," in *Proc. IEEE Ultrason. Symp.*, Oct. 8–11, 2002, pp. 494–499.
- [5] P.-E. Martinsson and J. E. Carlson, "Investigating the feasibility of using principal component analysis for ultrasonic classification of gas mixtures," in *Proc. IEEE Ultrason. Symp.*, Oct. 5–8, 2003, pp. 1396–1399.
- [6] J. E. Carlson and P.-E. Martinsson, "Exploring interaction effects in two-component gas mixtures using orthogonal signal correction of ultrasound pulses," *J. Acoust. Soc. Amer.*, vol. 117, no. 5, pp. 2961–2968, 2005.

- [7] S. Wold, M. Sjöström, and L. Eriksson, "PLS regression: A basic tool of chemometrics," *Chemometrics Intell. Lab. Syst.*, vol. 58, pp. 109–130, 2001.
- [8] M. Henningsson, E. Sundbom, B.-Å. Armelius, and P. Erdberg, "PLS model building: A multivariate approach to personality test data," *Scand. J. Psychol.*, vol. 422, pp. 399–409, 2001.
- [9] A. B. Bhatia, *Ultrasonic Absorption—An Introduction to the Theory of Sound Absorption and Dispersion in Gases, Liquids and Solids*. New York: Dover, 1985.
- [10] I. T. Jolliffe, *Principal Component Analysis*. 2nd ed. New York: Springer-Verlag, 2002.
- [11] G. Strang, *Linear Algebra and Its Applications*. 3rd ed. San Diego, CA: Harcourt Brace Jovanovich, 1986.
- [12] G. Golub and C. V. Loan, *Matrix Computations*. Oxford: Johns Hopkins Univ. Press, 1983.
- [13] H. Wold, *Systems Under Indirect Observation*. vol. I, II, Amsterdam, The Netherlands: North Holland, 1982.
- [14] H. Akaike, "A new look at statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, pp. 716–723, 1974.
- [15] I. N. Wakeling and J. J. Morris, "A test of significance for partial least squares regression," *J. Chemom.*, no. 7, pp. 291–304, 1993.
- [16] J. Shao, "Linear model order selection by cross-validation," *J. Amer. Stat. Assoc.*, vol. 88, pp. 486–494, June 1993.
- [17] G. S. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1987.



**Johan E. Carlson** was born 1973 in Umeå, Sweden. He received his M.Sc. degree in Computer Science and Engineering (1998) and his Ph.D. degree in Electrical Engineering (2002), at Luleå University of Technology, Sweden. He currently holds a position as Associate Professor at the Department of Computer Science and Electrical Engineering at Luleå University of Technology, Sweden. His research interests include: Ultrasonic measurement technology, signal processing, and multivariate data analysis.



**Rolf Carlson** was born in 1946 in Motala, Sweden. He earned the Fil.kand. (B.Sc.) in 1969 from the University of Lund, Lund, Sweden and the Fil.dr (Ph.D.) in organic chemistry in 1978, from Umeå University, Umeå Sweden. He was an Assoc. Professor and Research fellow in organic chemistry at Umeå University from 1979–1994. Since 1994 he has been a full professor in organic chemistry at Tromsø University, Tromsø, Norway. His research interests are organic synthetic chemistry, process organic chemistry, and development of strategies for design, analysis, and optimization of experimental procedures. He is the author of approximately 100 scientific papers and two books on synthesis optimization. He was awarded the Chevalier dans l'Ordre National du Mérite (French order) in 1994 and The Herman Wold Gold Medal in 2003.