This article was downloaded by:
On: *25 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Separation Science and Technology

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713708471

Application of Chemometrics to Modeling Produced Water Contamination

J. McFarlane[a]
[a] Separations and Materials Research Group, Nuclear Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

To cite this Article McFarlane, J.(2005) 'Application of Chemometrics to Modeling Produced Water Contamination', Separation Science and Technology, 40: 1, 593 — 609

To link to this Article: DOI: 10.1081/SS-200042524

URL: http://dx.doi.org/10.1081/SS-200042524

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Separation Science and Technology, 40: 593–609, 2005
ISSN 0149-6395 print/1520-5754 online
DOI: 10.1081/SS-200042524

Application of Chemometrics to Modeling Produced Water Contamination

J. McFarlane
Separations and Materials Research Group, Nuclear Science and
Technology Division, Oak Ridge National Laboratory,
Oak Ridge, TN, USA

Abstract: A partial least-squares statistical model was prepared to describe the contamination of produced water by soluble organic compounds. The model incorporated predictor variables used in earlier characterization studies: pH, temperature, pressure, salinity, water-to-oil ratio, and origin of the crude. Response variables included total extractable material and concentrations of aliphatic, aromatic, and polar organic molecules. The model was used to predict the uptake of water-soluble organics under a variety of conditions. The model will be applied to the prediction of hydrocarbon contamination of water produced in offshore drilling.

INTRODUCTION

Oil and gas production usually involves pumping large amounts of contaminated water along with the hydrocarbon phase. This high volume waste stream, on the order of a trillion gallons of water a year, is subject to National Pollution Discharge Elimination System permits. For Gulf of Mexico wells,

The author would like to thank the U.S. Department of Energy Natural Gas and Oil Technology Partnership Office for funding under contract AC10383OR41 and industrial partners through the Petroleum Energy Research Forum for their support.

Address correspondence to J. McFarlane, Separations and Materials Research Group, Nuclear Science and Technology Division, Oak Ridge National Laboratory, bldg 4500N, MS6181, Oak Ridge, TN 38731-6181, USA. Fax: 865 241-4829; E-mail: mcfarlanej@ornl.gov

the Environmental Protection Agency (EPA) limit on oil and grease content in produced water is a daily maximum of $42\,\mathrm{mg}\cdot\mathrm{L}^{-1}$ and a monthly average of no more than $29\,\mathrm{mg}\cdot\mathrm{L}^{-1}$ (1). After bulk separation from the hydrocarbon phase, the produced water contains organics, the concentrations of which have been measured on the order of a few ppm (2). The solubilities of individual components depend on the physical and chemical characteristics of the soluble organic molecules and insoluble colloidal particles (3). There is a variety of methods for removal of trace amounts of hydrocarbon; but the physical methods (i.e., adsorption and filtration) are subject to fouling, and the chemical methods can be compromised by the chemistry of the oil (4). Hence, it is important to know a priori the speciation and amounts of organic contaminants in the waste-water stream.

THEORY AND METHOD

The use of statistical analysis for model development has broad application throughout the sciences and social sciences. A statistical approach is particularly advantageous in the consideration of practical problems, when data come from a wide range of sources and comprise various forms, which do not lend themselves to a simple theoretical treatment. The chemometric approach has a number of advantages over phenomenological modeling as applied to environmental problems. Restrictions on the data set can be lax, allowing qualitative, as well as quantitative data to be used in the model. Depending on the adopted approach, predictor variables do not have to be strictly independent, as correlations will be eliminated in the transformation to factors represented by the score matrices. With some algorithms, even data sets with missing information can be accommodated, as long as sufficient independent data exist to determine all of the factors required to describe the system. In general, one factor (latent variable) is derived for each response variable in the model. Varmuzza gives an overview of chemometrics (5), which includes a useful description of the differences between methods such as multiple linear regression (MLR) and principal component analysis (PCA). Examples of the application of statistics to environmental problems are numerous. A couple of examples include Barbieri and coworkers who used PCA to investigate the pattern of freshwater contamination in Italian estuaries (6) and investigators at Oak Ridge National Laboratory who employed PCA to assess emissions from diesel fuels (7, 8).

Partial least-squares (PLS) analysis was used in this project; a method incorporating aspects of both MLR and PCA (9). In PLS, the covariance with respect to the response variable, y, is maximized, which allows for the calculation of correlation coefficients and also maximizes the variance between the predictor and response variables. PLS has been used widely in analytical chemistry applications, such as for calibration of spectra used in the determination of Pu(III) (10). A series of papers has been published by Brandvik and Daling on the application of a PLS algorithm to optimize the choice of which dispersants to use in the event of an oil spill at sea (11–13).

The goal of PLS analysis is to develop a linear model for the prediction of a response variable, y, based on a independent factors. During this discussion, it may be helpful to refer to Fig. 1. The factors are derived from the decomposition of an nxm matrix of predictor data, X, into an nxa score matrix (T) and an mxa loading matrix (P). The effective rank, a, of X, may be smaller than the actual number of variables, m, because of interdependencies between measured variables. The X scores, t_h, are made to be orthogonal by multiplying through by weight vectors, w_h, and the residuals, E, are small. Similarly, the nxp Y matrix of response variables is recast in the form of an nxa score matrix and a pxa loading matrix (Q), with residuals in F.

In practical terms, to construct the model, predictor and response data from a "training" data set are cast in matrices, X and Y, respectively, after transformation into a mean-centered, variance-scaled form. Following the NIPALS algorithm (14), model development involves a stepwise simultaneous breakdown of both matrices in terms of score vectors (t_h and u_h), weight vectors (w_h), and scalar regression parameters (b_h) to give an alternative description of the data in terms of a set of orthogonal eigenvectors for each factor, h.

Predictions are based on calculating a new set of latent variables (t_h) for a new set of predictor data, X, that is first loaded into the residual matrix, E. Each term of the score vectors, t_{hi}, and the loading vector, q_{hi}, from the

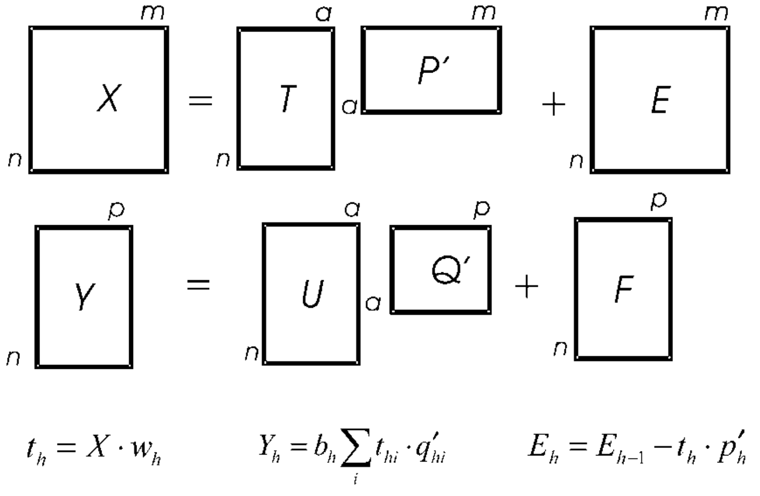

Figure 1. Schematic of partial least-squares model development and application (see text for explanation of symbols and methodology).

model, is used in combination with the regression parameter to determine a new response vector Y_h. Meanwhile, the residual matrix is recalculated using the score vector, t_h, and the loading vector, p_h. If more than one response variable is involved, the process will repeat with the "new" X matrix, E_h.

Evaluation of the model can be performed using standard statistical methods. Plots of scores from the input (t_h) and response (u_h) matrices allow assessment of the fit (or regression parameters, b_h). Correlations between input or response factors can be observed by plots of t_h vs. $t_{j \neq h}$, and u_h vs. $u_{j \neq h}$. Examples of the correlation plots are given in Fig. 2: t_1 vs. u_1, and t_1 vs. t_2. Numerical evaluations are also possible, through calculation of sums of squares (15), the F-test (16), and the PRESS algorithm (17), where each set of data in the training set is sequentially replaced followed by a comparison of the model parameters.

Programming of the PLS analysis was performed using the MAPLE 7 environment for symbolic mathematics (18) and was run on both a PC (Windows 2000, Windows XP) and an Apple G4 Powerbook (OS 9). Appendix A contains the MAPLE code used for model development.

RESULTS AND DISCUSSION

The PLS analysis was applied to the characterization data taken at ORNL (3). The X and Y matrices are given in (Table 1), showing which variables were chosen for the training data set before transformation to mean-centered and variance-scaled. Table 2 gives the loading and weight matrices derived from PLS analysis, as well as the regression coefficients for each of the resulting four factors. The model was then used to predict solubility as a function of the variables characterized for the produced water/crude oil contacts carried out earlier at ORNL. Results of the calculation are presented in the graphs shown in (Figs. 3–5) for pH, temperature, and salinity, respectively. The model predictions are shown as a dashed line on each graph. Also included are the experimental data for total extractable material according to EPA methodology (19).

In agreement with the experimental results, trends were observed with pH and temperature (Figs. 3 and 4), respectively. Salinity gave a poor correlation although a slight decrease in solubility was seen with both the total extractable material and the polar fraction (Fig. 5). As expected, a slight increase in solubility was observed in the organic fraction as a function of temperature. The trend in solubility with applied pressure was also explored using the PLS method. Because essentially no trends could be derived for the model, the results from this analysis are not presented in this paper. The solubility data used for the training set did not show a pressure dependence because the oil samples used in the experiments on which the PLS model was based were

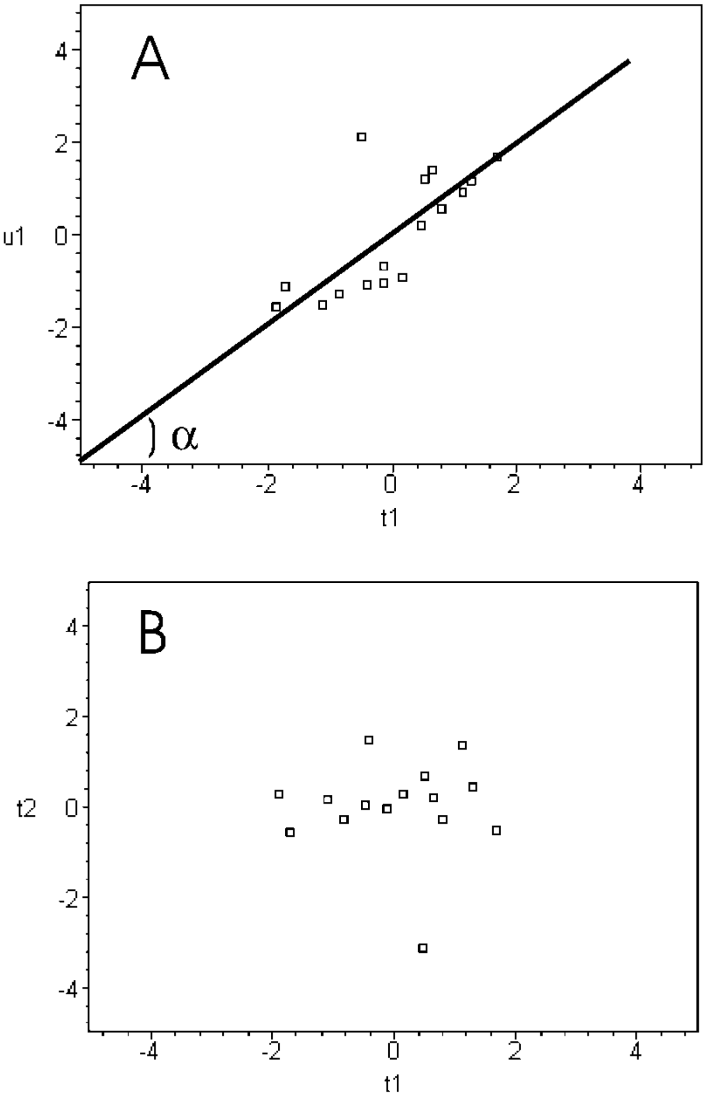

Figure 2. Examples of plots showing correlations between score eigenvectors. A: Regression between predictor and response variables. B: Correlation between first two orthogonal predictor factors.

degassed before they were shipped to Oak Ridge National Laboratory (ORNL). Only organics with very low ambient vapor pressures were sampled.

Predictions for solubility as a function of pH are presented in Fig. 3 as the dashed straight line. The solid line is a pseudo-titration curve arising from

Table 1a. Input variables for construction of X matrix

Temperature (°C)	pH	[Cl$^-$] (mg L^{-1})	Pressure (bar)	Volume ratio $H_2O : (H_2O + oil)$
25	7.00	40,600	1.0	0.80
25	9.03	65,000	1.0	0.80
25	5.98	65,000	1.0	0.80
25	7.00	65,000	1.0	0.67
25	4.73	65,000	1.0	0.80
75	7.00	65,000	1.0	0.80
25	7.00	114,500	1.0	0.80
25	7.00	65,000	1.0	0.80
50	7.00	65,000	1.0	0.80
25	8.09	65,000	1.0	0.80
25	7.00	65,000	1.0	0.50
25	7.00	65,000	1.0	0.20
25	7.00	65,000	6.9	0.80
25	7.00	65,000	17.0	0.80
25	7.00	65,000	29.3	0.80
25	7.00	65,000	1.0	0.80

Table 1b. Response variables (mg · L^{-1}) for construction of Y matrix

Total extractable material	Aliphatic compounds	Aromatic compounds	Polar compounds
12.0	0.066	0.030	4.60
28.6	0.066	0.280	6.94
11.7	0.005	0.040	5.30
11.6	0.060	0.030	6.94
11.9	0.039	0.070	3.10
24.8	0.066	0.330	12.6
14.5	0.253	0.050	6.10
21.0	0.016	0.110	2.35
21.9	0.216	0.080	6.70
23.5	0.010	0.110	27.9
30.6	0.050	0.305	5.80
30.0	0.051	0.240	6.94
47.5	0.103	0.313	3.43
6.1	0.015	0.024	6.25
4.1	0.020	0.090	2.08
3.9	0.016	0.166	4.07

Table 2. Calculated loading, and weight matrices and regression coefficients

p			
0.473	−0.152	0.096	0.684
0.567	0.200	0.533	0.047
0.145	−0.869	−0.225	−0.020
−0.575	0.085	0.130	0.728
−0.319	−0.416	0.800	0.023
q			
0.545	0.235	−0.316	−0.312
0.363	−0.919	−0.237	0.177
0.582	0.316	−0.298	0.933
0.483	0.025	0.869	−0.043
w			
0.496	−0.126	−0.036	0.646
0.615	0.206	0.534	0.087
0.170	−0.886	−0.240	0.006
−0.490	0.093	−0.034	0.762
−0.342	−0.387	0.837	0.005
b			
0.906	0.641	0.278	0.245

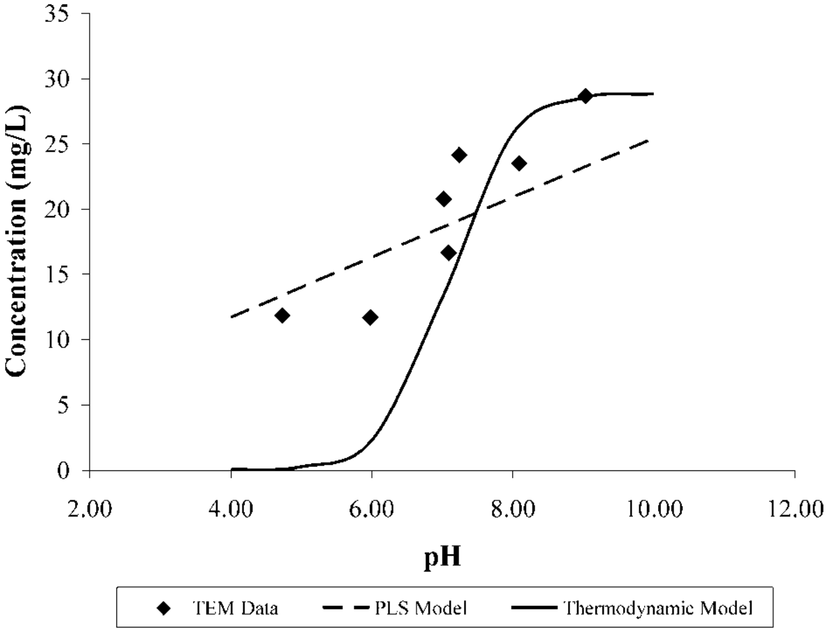

Figure 3. Predicted solubility as a function of pH.

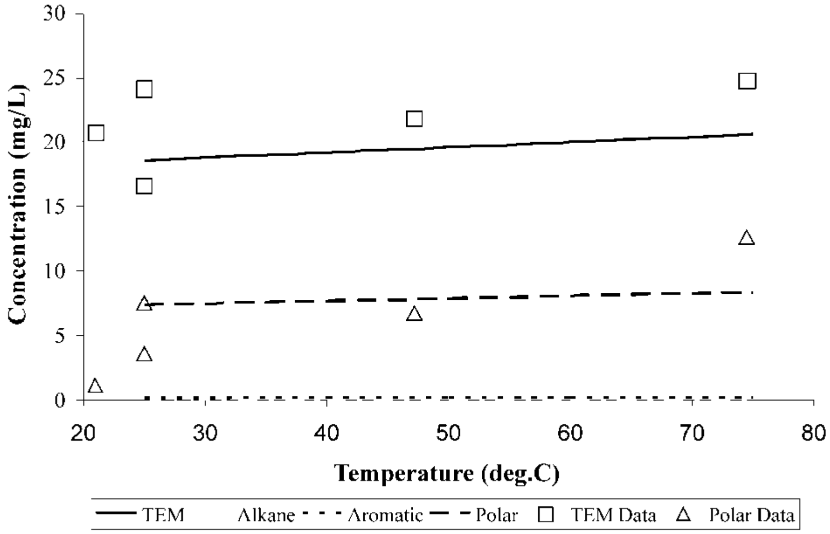

Figure 4. Solubility (mg/L) as a function of temperature (°C). Data for total extractable material (squares) and polar organics (triangles) are superimposed on PLS predictions of solubility. A slight increase in solubility was observed over the temperature range 25–75°C.

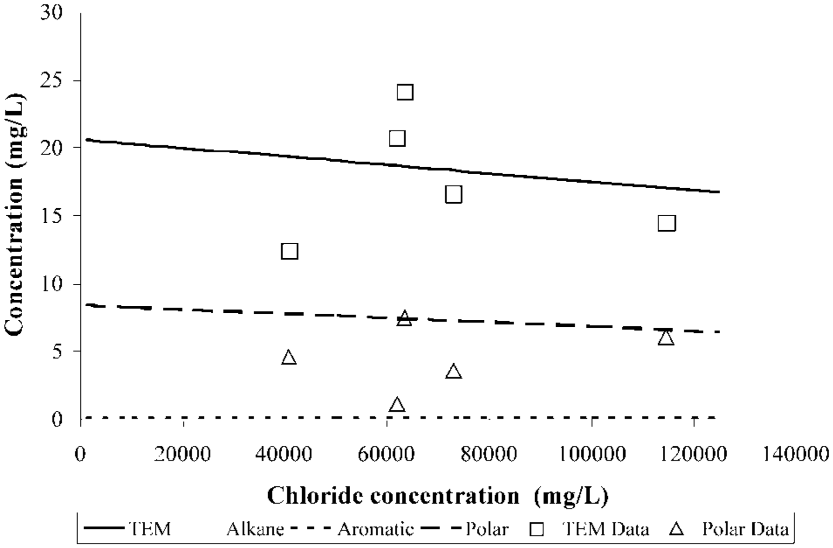

Figure 5. Solubility (mg/L) as a function of chloride ion concentration (mg/L). Data for total extractable material (squares) and polar organics (triangles) are superimposed on PLS predictions of solubility. In these experiments, the concentrations of aromatic and aliphatic compounds were at or below the detection limit.

thermodynamic calculations on a produced water brine–hydrocarbon system (20). Although the statistical model does not reproduce the "s"-shaped curve seen in the thermodynamic model, the analysis was performed without any assumptions concerning the chemistry of the water or the oil. The thermodynamic model, however, required the input of mole fractions of specific compounds into the system, and the phase behavior was calculated based on these predetermined variables. Use of a more sophisticated input variable in the statistical model, such as degree of dissociation rather than pH, should give results that more closely reproduce the pseudo-titration curve seen in the data. Unlike a thermodynamic formulation, however, a statistical model cannot be used to calculate physical properties a priori. For instance, differentiation of the Gibbs free energy with respect to solute mole fraction will give the upper and lower consulate temperatures of a mixture, but a statistical model gives no such insight into phase separation behavior.

CONCLUSIONS

Based on partial least squares analysis, a statistical model was constructed based on data for organic contamination of produced water. When used for predictions, the PLS model successfully reproduced trends in agreement with the data, showing an increase in solubility with pH and with temperature. The advantage of the PLS model is that field data can be easily incorporated into the model. For this reason, it will likely be the favored approach for implementation outside the laboratory. The usefulness of the model will be improved by eliminating apparent weaknesses in the model, such as the imposition of a linear pH dependence where a nonlinear trend was observed in the data. In PLS, this can be achieved by careful selection of predictor variables that incorporate nonlinear effects, such as fraction dissociated rather than pH. Refinements such as these will require incorporation of a larger and more representative data set in model development.

APPENDIX A

Maple 7 Program for Partial Least-Squares Applied to Produced Water Data
Joanna McFarlane August 29, 2003

Introduction

The object is to build a model that predicts an outcome, Y, based on a predictor, X. This program was developed following the NIPALS algorithm outlined in Geladi, P. and Kowalski, B.R., "Partial Least-Squares Regression: A Tutorial," Analytica Chimica Acta 185, 1–17 (1986).

Tested with data from Geladi, P. and Kowalski, B.R., “An Example of 2-Block Predictive Partial Least-Squares Regression with Simulated Data”, Analytica Chimica Acta 185, 19–32 (1986).

```
> restart:with(stats):with(linalg):with(plots):
```

Defining the Data

The training data—or that used to develop the model—is typed or read into the program. It is converted to “mean-centred” and “variance-scaled” form. In this case, 9 separate samples (or objects) were used for model development. The X data are the measured variables (predictors) and the Y data are the concentrations (outcomes or responses).

```
>
X: = [[25,7,40638,1,0.8],[25,9.03,65000,1,0.8],[25,5.98,65000,1,0.8],
[25,7,65000,1,0.67],[25,4.73,65000,1,0.8],[75,7,65000,1,0.8],[25,7,
114500,1,0.8],[25,7,65000,1,0.8],[50,7,65000,1,0.8],[25,8.09,65000,1,0.8],
[25,7,65000,1,0.5],[25,7,65000,1,0.2],[25,7,65000,6.9,0.8],[25,7,65000,17,
0.8],[25,7,65000,29.3,0.8],[25,7,65000,1,0.8]]:array(X);

Y: = [[12,0.066,0.03,4.6],[28.6,.066,0.28,6.94],[11.7,0.005,0.04,5.3],[11.6,
0.062,0.03,6.94],[11.9,0.039,0.07,3.1],[24.8,.066,0.33,12.6],[14.5,0.253,
0.05,6.1],[21,0.016,0.11,2.35],[21.9,0.216,0.08,6.7],[23.5,0.01,0.11,27.9],
[30.6,0.05,0.305,5.8],[30,0.051,0.24,6.94],[47.5,0.103,0.313,3.43],[6.1,
0.015,0.024,6.25],[4.1,0.02,0.09,2.08],[3.9,0.016,0.166,4.07]]:
array(Y);
num_samples: = rowdim(array(X));
num_ind_var: = coldim(array(X));
num_dep_var: = coldim(array(Y));
```

Procedures from J.A.Rafter, M.L.Abell and J.P.Braselton— “Statistics with Maple,” Academic Press, Amsterdam, 2003.

```
> Columns: = proc(data::list,cols::list)
        if nops(cols) = 1 then
                [seq(seq(data[i,j],j = cols),i = 1.nops(data))];

        else

                [seq([seq(data[i,j],j = cols)],i = 1.nops(data))];fi;
end:

> ColumnToList: = proc(data::list,cols::list)

        local a;
```

```
        if nops(cols) = 1 then
                if type(cols[],range) then
                a: = [seq(seq(data[i,j],i = 1.nops(data)),j = cols)];
                convert(linalg[transpose](a),listlist);
                else
                [seq(seq(data[i,j],i = 1.nops(data)),j = cols)];
                fi;

        else

                [seq([seq(data[i,j],i = 1.nops(data))],j = cols)];

        fi;

end:
```

Conversion of the data to be mean-centered and variance-scaled

```
 > s: = 's':X1: = []:
            for i from 1 to num_ind_var do s: = Columns(X,[i]);
            s1: = describe[mean](s);
            s2: = describe[standarddeviation[1]](s);
            X1: = [op(X1),(evalf(map(x - > (x - s1)/s2,s)))];
od:
X2: = ColumnToList(X1,[1.num_samples]):
E: = array(X2);
s: = 's':Y1: = []:
for i from 1 to num_dep_var do
            s: = Columns(Y,[i]);
            s1: = describe[mean](s);
            s2: = describe[standarddeviation[1]](s);
            Y1: = [op(Y1),(evalf(map(x - > (x - s1)/s2,s)))];
    od:
Y2: = ColumnToList(Y1,[1.num_samples]):
F: = array(Y2);
```

Model Building

Initialization

```
> check: = 1:tstart: = ColumnToList(X2,[1]);

ustart: = ColumnToList(Y2,[1]);Xm: = matrix(X2):Ym: = matrix(Y2):
pm: = matrix(5,1,[1,1,1,1,1]):qm: = matrix(4,1,[1,1,1,1]):
wm: = matrix(5,1,[1,1,1,1,1]):
tm: = matrix(16,1,[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]):
```

```
um: = matrix(16,1,[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]):
bm: = []:j: = ‘j’:
> for j from 1 to num_dep_var do
        while check > 0.1 do
```

X block

```
  u_ip: = 1/innerprod(ustart,ustart);
  w: = multiply(ustart,Xm);
  wold: = scalarmul(w,u_ip);
  worm: = 1/norm(wold,2);
  wnew: = scalarmul(wold,worm);
  w_ip: = 1/innerprod(wnew,wnew);
  t1: = multiply(Xm,wnew);
  t2: = scalarmul(t1,w_ip);
```

Y block

```
  t_ip: = 1/innerprod(t2,t2);
  q: = multiply(t2,Ym);
  qold: = scalarmul(q,t_ip);
  qorm: = 1/norm(qold,2);
  qnew: = scalarmul(qold,qorm);
  q_ip: = 1/innerprod(qnew,qnew);
  u1: = multiply(Ym,qnew);
  u2: = scalarmul(u1,q_ip);
```

Check convergence

```
  check: = norm(t2 − tstart);
  ustart: = u2;tstart: = convert(t2,list);
od;
```

Calculate X-loadings and rescale scores and weights

```
  print(“*****CONVERGENCE********CONVERGENCE******);
  t_ip: = 1/innerprod(t2,t2);
  p: = multiply(t2,Xm);
  pold: = scalarmul(p,t_ip):
  porm: = 1/norm(pold,2):
  pnew: = scalarmul(pold,porm);
  tnew: = scalarmul(t2,1/(porm));
  wnew: = scalarmul(wnew,1/(porm));
  pm: = concat(pm,pnew):tm: = concat(tm,tnew):
  wm: = concat(wm,wnew):qm: = concat(qm,qnew):
  um: = concat(um,u2):
```

Calculation of regression coefficient

```
   b: = multiply(u2,t2)*t_ip;bm: = [op(bm),b]:
```

Calculation of residual matrices

```
   Xdiff: = multiply(tnew,transpose(pnew)):
   Xm2: = matadd(Xm,Xdiff,1, − 1):Xm: = matrix(Xm2);
   Ydiff: = multiply(tnew,transpose(qnew)):
   Ym2: = matadd(Ym,Ydiff,1, − b):Ym: = matrix(Ym2);
```

Check for convergence. Recalculated u for computation, Need to multiply by −1 to reproduce results in paper.

```
   if (j < num_dep_var) then ustart: = col(Ym,j + 1);
   ustart: = scalarmul(ustart, − 1);
   check: = 1:fi;

od:
```

Output files (loadings p, q and weights w are used for prediction, bare regression coefficients, scores t and u are used for assessment).

```
>pm: = delcols(pm,1.1);qm: = delcols(qm,1.1);tm: = delcols(tm,1.1);
wm: = delcols(wm,1.1);um: = delcols(um,1.1);
```

Model Testing

Sum of Squares, Plots

Sum of squares are first calculated for samples, and then calculated for each column of the X matrix and Y matrix to show the degree of minimization of the residuals. If the sum of squares is large for a particular factor, this means that it did not play a great role in the model as it was not well minimized during the computation. The overall sum of squares for X and Y show how well the model converged for the X and Y data respectively.

```
> Xm3: = convert(Xm,listlist):SSx_rows: = 0:i: = ‘i’:
> for i from 1 to num_samples do
        sum_squares_sample[i]: = describe[sumdata[2]](Xm3[i]);
        SSx_rows: = SSx_rows + sum_squares_sample[i]:

od;

> i: = ‘i’:SSx_cols: = 0:
> for i from 1 to num_ind_var do
        sum_squares_list: = convert(col(Xm,i),‘list’):
```

```
        sum_squares_x[i]: = describe[sumdata[2]](sum_squares_list);
        SSx_cols: = SSx_cols + sum_squares_x[i]:
od;
```

```
> i: = ‘i’:SSy: = 0:
> for i from 1 to num_dep_var do
        sum_squares_list: = convert(col(Ym,i),‘list’):
        sum_squares_y[i]: = describe[sumdata[2]](sum_squares_list);
        SSy: = SSy + sum_squares_y[i];
od;
```

Plots are a valuable tool to visualize the goodness of fit. Regression plots of y versus x give an idea of how much the response variable depends onthe input variable. Mapping of t scores indicates correlations between predictor variables, u scores indicates correlations between response variables.

```
> i: = ‘i’:j: = ‘j’:
> for i from 1 to (num_dep_var − 1) do
        t1plot: = col(tm,i):ts1: = cat(“t”,i):
        t2plot: = col(tm,i + 1):ts2: = cat(“t”,i + 1):
        set1: = [seq([t1plot[j],t2plot[j]],j = 1.num_samples)]:
        Tplot: = plot(set1, − 5.5, − 5.5, style = point,symbol = box,
        axes = boxed, color = black, labels = [ts1,ts2]):Tplot;
        u1plot: = col(um,i):us1: = cat(“u”,i):
        u2plot: = col(um,i + 1):us2: = cat(“u”,i + 1):
        set2: = [seq([u1plot[i],u2plot[i]],i = 1.num_samples)]:
        Uplot: = plot(set1, − 5.5, − 5.5, style = point, symbol = box,
                axes = boxed, color = black,labels = [us1,us2]): Uplot;
od;
```

Plots of t versus u gives an idea of how good the regression is.

```
> i: = ‘i’:j: = ‘j’:
> for i from 1 to num_dep_var do
        t1plot: = col(tm,i):ts1: = cat(“t”,i):
        u1plot: = col(um,i):us1: = cat(“u”,i):
        set1: = [seq([t1plot[j],u1plot[j]],j = 1.num_samples)]:
        Tplot: = plot(set1, − 5.5, − 5.5, style = point, symbol = box,
                axes = boxed,color = black,labels = [ts1,us1]):Tplot;
od;
```

Prediction

In this section, one can calculate a new Y matrix, given an input X matrix. In this example, there is a single X vector. The input data is scaled according to the mean and variance calculated in model development.

```
> XP: = [25,7,65000,1,.6]:
```

Changing input values to mean-centred, variance-scaled values.

```
> s: = ‘s’:XP1: = []:
> for i from 1 to num_ind_var do
        s: = Columns(X,[i]);
        s1: = describe[mean](s);
        s2: = describe[standarddeviation[1]](s);
        XP1: = [op(XP1),(evalf(map(x −  > (x − s1)/s2,XP[i])))];

od:
> XP1;
```

Decomposing the X block into score matrix (vector) t, using weights matrix w.

```
> TP: = multiply(XP1,wm);TP1: = multiply(TP,transpose(pm));
> XP2: = matadd(XP1,TP1,1, − 1);
```

Building up Y block using u scores

```
> UP: = multiply(TP,transpose(qm));
> i: = ‘i’:YP: = [i]:
> for i from 1 to num_dep_var do
        YP: = [op(YP),scalarmul(UP,bm[i])[i]];

od:
> YP: = [op(2.num_dep_var + 1,YP)];
```

Regenerate “physical” concentrations.

```
> s: = ‘s’:YP1: = []:
> for i from 1 to num_dep_var do
        s: = Columns(Y,[i]);
        s1: = describe[mean](s);
        s2: = describe[standarddeviation[1]](s);
        YP1: = [op(YP1),(evalf(map(x −  > s1 + (s2*x),YP[i])))];

od:
YP1;
```

REFERENCES

1. US EPA. *Notice of Final NPDES General Permit; Final NPDES Permit for New and Existing Sources and New Dischargers in the Subcategory of the Oil and Gas Extraction Category for the Western Portion of the Outer Continental Shelf of the Gulf of Mexico* (GMG290000), FRL-7119-5, Federal Register (66–243), Tuesday December 18, 2001, notices, 65209.

2. Bostick, D.T., Luo, H., and Hindmarsh, B. (2002) Oak Ridge National Laboratory Technical Memorandum ORNL/TM-2001/78. In *Characterization of Soluble Organics in Produced Water*; Oak Ridge National Laboratory: Oak Ridge, TN.
3. McFarlane, J., Bostick, D.T., and Luo, H. Characterization and Modeling of Produced Water, Proceedings of the Groundwater Protection Council Produced Water Conference, Oct. 16, 2002, http://www.gwpc.org/meetings/PW2002/Papers-Abstracts.htm.
4. Wills, J.A. (2000) M. Inst. Pet. For Ekologicheskaya Vakhta Sakhalina (Sakhalin Environmental Watch). In *Survey of Oilfield Offshore Drilling Wastes and Disposal Technologies to Reduce the Biological Impact of Sea Dumping*; May.
5. Varmuzza, K. Chemometrics: Multivariate View on Chemical Problems. In *Encyclopedia of Computational Chemistry* 346–366.
6. Barbieri, P., Adami, G., Piselli, S., Gemiti, F., and Reisenhofer, E. (2002) A Three-way Factor Analysis for Assessing the Time Variability of Freshwaters Related to a Municipal Water Supply. *Chemom. Intell. Lab. Syst.*, 62: 89–100.
7. McAdams, H.T., Crawford, R.W., and Hadder, G.R. (2000) Oak Ridge National Laboratory Technical Memorandum ORNL/TM-2000/5. In *A Vector Approach to Regression Analysis and its Application to Heavy-Duty Diesel Emissions*; Oak Ridge National Laboratory: Oak Ridge, TN, November.
8. McAdams, H.T., Crawford, R.W., and Hadder, G.R. (2000) ORNL/TM-2002-16. In *PCR+ in Diesel Fuels and Emissions Research*; Oak Ridge National Laboratory: Oak Ridge, TN, March.
9. Wold, S., Sjostrom, M., and Eriksson, L. Partial Least Squares Projections to Latent Structures (PLS) in Chemistry. In *Encyclopedia of Computational Chemistry*, 2006–2021.
10. Wangen, L.E., Phillips, M.V., and Walker, L.F. (1988) Los Alamos National Laboratory Report LA-11297-MS. In *Use of Multivariate Calibration for Plutonium Quantitation by the Pu(III) Spectrophotometric Method*; Los Alamos National Laboratory: Los Alamos, NM, May; 1–8.
11. Brandvik, P.J. (1998) Statistical Simulation as an Effective Tool to Evaluate and Illustrate the Advantage of Experimental Designs and Response Surface Methods. *Chemom. Intell. Lab. Syst.*, 42: 51–61.
12. Brandvik, P.J. and Daling, P.S. (1998) Optimisation of Oil Spill Dispersant Composition by Mixture Design and Response Surface Methods. *Chemom. Intell. Lab. Syst.*, 42: 61–72.
13. Brandvik, P.J. and Daling, P.S. (1998) Optimising Oil Spill Dispersants as a Function of Oil Type and Weathering Degree: A Multivariate Approach using Partial Least Squares (PLS). *Chemom. Intell. Lab. Syst.*, 42: 73–91.
14. Geladi, P. and Kowalski, B.R. (1986) *Partial Least Squares Regression: A Tutorial.*, 185: 1–17.
15. Geladi, P. and Kowalski, B.R. (1986) An Example of 2-Block Predictive Partial Least Squares Regression with Simulated Data. *Anal. Chim. Acta*, 185: 19–32.
16. Draper, N. and Smith, H. (1981) *Applied Regression Analysis*; Wiley: New York; 33–40.
17. Wold, S., Ruhe, A., Wold, H., and Dunn, W.J., (III) (1984) The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *Siam. J. Sci. Stat. Comput.*, 5 (3): 735–743.
18. MAPLE 7, Waterloo Maple Inc., 57 Erb Street W., Waterloo, Ontario, Canada, N2L 6C2, 2001. http://www.maplesoft.com.

19. (a) US EPA. *Methods for the Determination of Organic Compounds in Drinking Water*, Method 525.2 Rev 1.0 (March) 1994; (b) Edwards, J., Fannin, S.T., and Conoley, M. (1998) *Analysis of Semi-volatile Organic Compounds under US EPA Method 525.2*, ThermoQuest, GCQ Technical Report 9215.
20. McFarlane, J., Bostick, D.T., and Luo, H. (2003) Analysis and Modeling of Hydrocarbon Contamination in Produced Watersubmitted to Chemosphere.