

Modeling of Water-Soluble Organic Content in Produced Water

May 2006

**Prepared by
Joanna McFarlane**

DOCUMENT AVAILABILITY

Reports produced after January 1, 1996, are generally available free via the U.S. Department of Energy (DOE) Information Bridge:

Web site: <http://www.osti.gov/bridge>

Reports produced before January 1, 1996, may be purchased by members of the public from the following source:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-605-6000 (1-800-553-6847)
TDD: 703-487-4639
Fax: 703-605-6900
E-mail: info@ntis.fedworld.gov
Web site: <http://www.ntis.gov/support/ordernowabout.htm>

Reports are available to DOE employees, DOE contractors, Energy Technology Data Exchange (ETDE) representatives, and International Nuclear Information System (INIS) representatives from the following source:

Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831
Telephone: 865-576-8401
Fax: 865-576-5728
E-mail: reports@adonis.osti.gov
Web site: <http://www.osti.gov/contact.html>

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Nuclear Science and Technology Division

**MODELING OF WATER-SOLUBLE ORGANIC CONTENT
IN PRODUCED WATER**

Joanna McFarlane

Date Published: May 2006

Prepared by
OAK RIDGE NATIONAL LABORATORY
P.O. Box 2008
Oak Ridge, Tennessee 37831-6181
managed by
UT-Battelle, LLC
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	viii
ACRONYMS	ix
ACKNOWLEDGMENTS	xii
EXECUTIVE SUMMARY	xiii
1. INTRODUCTION	1
1.1 INTERNATIONAL DATA SETS ON PRODUCED-WATER COMPOSITION	1
1.2 WATER CHARACTERIZATION AT ORNL	3
1.3 APPROACH TO MODELING	4
2. THEORY	8
2.1 Liquid–Liquid Equilibrium Model	8
2.2 Partial Least-Squares Analysis	9
3. APPLICATION TO PRODUCED-WATER MODELING	11
3.1 CONCENTRATIONS OF WATER-SOLUBLE ORGANICS IN CRUDE OIL	12
3.2 RANGES OF PHYSICAL AND CHEMICAL VARIABLES	26
3.3 MODEL BASED ON THERMODYNAMIC INPUTS INTO PLS	28
3.4 COMPARISON WITH PRODUCED WATER DATA	32
4. CONCLUSIONS	37
5. REFERENCES	39
Appendix A MAPLE 7 Code for Thermodynamic Equilibrium Calculations	A-1
Appendix B MAPLE 7 Code for Determination of UNIFAC Activity Coefficients	B-1
Appendix C MAPLE 7 Code for Partial Least-Squares Analysis	C-1

LIST OF FIGURES

Figure		Page
1	Dependence of Henry's Law constants of polar organics on pH.....	6
2	Comparison of model results with total extractable material data from contact experiments of simulated produced-water brine with actual crude oil.	8
3	Results of partial least-squares analysis, showing the leading independent factor (best aligned with concentration) versus the leading dependent factor (best aligned with pH)	11
4	Results of principal component analysis of crude oil properties.....	14
5	Correlation of water-hydrocarbon partitioning with density for paraffinic and naphthenic compounds.	19
6	Correlation of water-hydrocarbon partitioning with gram molecular mass for paraffinic and naphthenic compounds.....	20
7	Correlation of water-hydrocarbon partitioning with gram molecular mass for all aromatic compounds considered.	21
8	Correlation of water-hydrocarbon partitioning with normal boiling point for all soluble organics considered.	23
9	Comparison of calculated vs literature values for octanol-water partition coefficients.....	25
10	PLS diagnostics.....	31
11	Calculated vs experimental aqueous concentrations.....	33
12	Logarithmic plot of predicted vs measured concentrations in milligrams per liter.....	34

LIST OF TABLES

Table	Page
1 Results of uncertainty analysis in water characterization experiments.....	5
2 Properties of selected crude oils.....	13
3 Properties of selected aliphatic, olefinic, and naphthenic compounds in crude oil.....	16
4 Properties of selected aromatic compounds in crude oil.....	17
5 Properties of selected polar compounds in crude oil.....	18
6 Calculated distribution coefficients ($K_{W/HC}$) of selected organics.....	27
7 Uncertainties in computed PLS results.....	30
8 Crude oil composition.....	32
9 Experimental vs calculated distribution coefficients and aqueous concentrations.....	35

ACRONYMS

API	American Petroleum Institute
BTEX	benzene, toluene, ethylbenzene and xylene
DOE	U.S. Department of Energy
EPA	U.S. Environmental Protection Agency
NAP	naphthalenes
NIPLS	Nonlinear Iterative Least Squares (model)
NPD	naphthalene, phenanthrene, and dibenzothiophene
NPDES	National Pollution Discharge Elimination System
NRTL	Non-Random Two-Liquid (method/model)
ORNL	Oak Ridge National Laboratory
PAH	polyaromatic hydrocarbons
PERF	Petroleum Energy Research Forum
PLS	partial least-squares analysis
PRESS	predictive residual sum of squares (algorithm/analysis)
TEM	total extractable material
TPH	total petroleum hydrocarbons
UNIFAC	UNIQUAC Functional-group Activity Coefficient
UNIQUAC	Universal Quasichemical
VOC	volatile organic carbons

ACKNOWLEDGMENTS

This work was supported through the Natural Gas and Oil Technology Partnership, administered by the U.S. Department of Energy's National Petroleum Technology Office under contract AC103830OR41. Produced-water and oil samples were supplied by industrial partners through the Petroleum Energy Research Forum Program.

EXECUTIVE SUMMARY

A model for the solubility of organic compounds in produced water was developed and compared with the data from the Petroleum Energy Research Forum (PERF) Program 98-04 characterization experiments.

Large amounts of brine are often associated with oil and gas production. Because these produced waters are in contact with oil in the formation at high pressures, they can become contaminated with water-soluble organic compounds. The discharge of produced water in the Gulf of Mexico is regulated by National Pollution Discharge Elimination System (NPDES) permits, which specify that total oil and grease in the water be below a daily maximum of $42 \text{ mg}\cdot\text{L}^{-1}$. However, analysis of the produced water for total petroleum hydrocarbons as specified by U.S. Environmental Protection Agency (EPA) methods 413.1 or 1664 do not distinguish between carboxylic acids and other polar compounds and oil and grease (typically comprised of paraffinic and aromatic hydrocarbons). Hence, remediation of the billion barrels of produced water per annum is based on aqueous organic concentrations that exceed the actual content of oil and grease.

The goal of the project was to provide a model, based on analysis and modeling of oil/brine samples, to be used to predict the nonvolatile water-soluble organic content in brines associated with deep-well oil production. The model predicted the production of water-soluble organics as a function of measurable parameters such as crude composition, physical conditions at the wellhead, and produced water composition. Because such information identifies and quantifies the production of water-soluble contaminants, it may be used in the design of efficient and cost-effective water treatment options for the next generation of offshore platforms for deep-water wells. Industry may use this information to develop treatment guidelines prior to construction of facilities, thus assisting in the development of a more selective and focused approach to produced-water cleanup and leading to cost savings and reduced environmental impact.

The measurement of the solubility of organic fractions in brines was conducted as part of PERF project 98-04. Distribution coefficients for various categories of water-soluble organic components (aliphatic, aromatic, and polar) were derived from these data. The results indicated that a significant fraction of the water-soluble organic component in the brine comprised C1 to C3 organic acids. A sensitivity analysis was conducted on the solubilities of the various components with respect to pressure (1 to 60 atm), temperature (25 to 75°C), pH (4.7 to 9.5), brine-to-oil ratio (20 to 80%), and salinity (35,000 to 150,000 ppm Cl^- by weight). The factors that had the most effect on solubility were pH, followed by temperature and pressure. The influence of pH is indicative of the presence of organic acids, because their solubility depends on their dissociation in the aqueous phase, based the negative logarithm of the acid-dissociation constant, or pKa.

An empirical analysis was conducted to determine if the approach could be used to quantify the effect of parameters such as pH and salinity on organic solubility in the aqueous phase. It was determined, however, that univariate empirical fits were inadequate to describe the multivariate/multicomponent produced-water brine/organic system. In addition, uncertainties in

the PERF solubility data measured at ORNL precluded the use of the model for lumped-parameter properties.

Hence, the produced-water/crude oil chemical system was modeled using a chemical thermodynamic description of liquid–liquid equilibrium for representative organic–aqueous systems. Thermodynamic properties were based on molecular functional group analysis. Two models — the Non-Random Two Liquid (NRTL) and the UNIQUAC (Universal Quasichemical) Functional-group Activity Coefficient (UNIFAC) — were employed to predict the solubility of prototypical organic compounds in brines under conditions of interest to the oil industry. Results of calculations were compared with laboratory data to give the range of various parameters as a function of measured variables (pH, temperature, pressure, and salinity). The model was able to reproduce the increase in solubility with an increase in pH as observed in the water-characterization laboratory experiments conducted at Oak Ridge National Laboratory (ORNL). Temperature dependencies were also modeled.

A statistical partial least-squares algorithm was chosen as the platform for a predictive analysis, with the results of the thermodynamic model being used to create an input data set with representative key organic compounds. Hence, a complex thermodynamic analysis was simplified from a description of ~1000 compounds to two dozen or so representative categories of compounds that have similar behavior. An advantage of a statistical analysis is that qualitative as well as quantitative information can be used in the model, allowing input from field-based variables (geographical location, depth, age of rock formation, etc.) as well as those measured via chemical analysis. Published literature has been reviewed to determine the range and distribution of oil and produced-water compositions, as well as physical conditions (temperature, pressure, pH) expected at the drilling site. These data, as well as the ranges that they cover, were used to formulate the scope of the predictive model. Results from the partial least-squares analysis were compared with produced-water characterization measurements to evaluate the validity of the model.

In summary, the largest fraction of organics in produced water is polar in nature and is comprised primarily of organic acids. Acetic acid was observed in the ORNL produced-water characterization experiments, along with C3 and C4 acids. The difficulty in quantifying oxidized organics in produced water arises because their concentrations not only are dependent on the conditions in the formation but also change upon processing, from formation to downhole to wellhead. These compounds have a large effect on the measured “oil and grease” content of produced water and are also indicative of kerogen breakdown into petroleum. Published data were used to develop correlations for polar organics in produced water based on the key factors of pH, temperature, oil-to-water ratio, and chloride ion concentration. The predictions of the model were compared with ORNL data collected on produced-water brine systems. The model may be used by the petroleum industry to estimate the concentrations of water-soluble organic compounds in produced water associated with drilling in the Gulf of Mexico.

1. INTRODUCTION

National Pollution Discharge Elimination System (NPDES) permits offshore discharge of produced water associated with deep oil wells in the Gulf of Mexico to a daily maximum of $42 \text{ mg}\cdot\text{L}^{-1}$ and average of $29 \text{ mg}\cdot\text{L}^{-1}$ [1]. Although the solubility of most individual oil and grease compounds is lower than these levels, the analytical tests mandated by the U.S. Environmental Protection Agency (EPA) for the measurement of total petroleum hydrocarbons (TPH) also include more-soluble compounds such as carboxylic acids [2], a practice that can result in the apparent TPH value exceeding the permitting level. Crude oil from the Gulf of Mexico has a relatively high concentration of polar molecules, namely those containing oxygen, nitrogen, and sulfur. Many of these are water soluble, making the NPDES regulations particularly stringent and expensive to follow. In other jurisdictions, the concerns are similar. For instance, discharge into the North Sea is limited to a monthly average of $40 \text{ mg}\cdot\text{L}^{-1}$ [3].

Effective remediation of produced water offshore requires an understanding of the types and amounts of soluble hydrocarbons and the ability to predict the solubility as a function of variables that can be measured in the field. Online monitoring of water-soluble organics may be applied to existing wells and would allow quick intervention should problems arise with organic removal. For new wells, prediction of water-soluble organic content in produced water would assist an informed selection of cleanup technologies. Optimized removal methods may be included in the design phase, minimizing retrofit and construction costs. The petroleum industry has been addressing these issues through the Petroleum Energy Research Forum (PERF) and has collaborated with the U.S. Department of Energy (DOE) to support produced-water research at Oak Ridge National Laboratory (ORNL).

The primary focus of produced-water research at ORNL has been to characterize water-soluble organics in produced water and to develop models to explain and predict the concentrations of these organics under conditions similar to those encountered in the field.

1.1 INTERNATIONAL DATA SETS ON PRODUCED WATER COMPOSITION

Because of stronger environmental regulations in Europe, much of the open-literature information on organics in produced water has addressed North Sea oil wells,. Results have been summarized in a review article by Utvik and Hasle (2002) [4]. Analyses from several studies of Norwegian produced water from 1995 onward show the following to be the predominant classes of organic compounds present in the aqueous phase: dispersed oil, BTEX (benzene, toluene, ethylbenzene, and xylene), NPD (naphthalene, phenanthrene, dibenzothiophene, and C1–C3-substituted homologues), PAH (polyaromatic hydrocarbons), organic acids, phenol, and substituted phenols.

Older data come from Barth [5], who measured the concentrations of organic acids in formation and produced waters from wells on the Norwegian continental shelf using isachophoresis — a technique based on ionic mobility. The organic acids accounted for 70–100% of the water-soluble organics in the samples, and acetic acid was the most predominant acid present. In addition, Barth noted that the composition of the produced water (after separation) was more complex than that of the formation water. This suggested to her that the formation water had been subject to bacterial action, or oxidation of organics, in the separation process.

Brendehaug et al. [6] characterized water soluble organics in terms of broad chemical classifications — aliphatic, aromatic, and polar — and performed a detailed analysis to identify compounds of toxicological importance. This group found that the dissolved hydrocarbons were dominated by the volatile components, such as benzene and its derivatives, rather than the oxidized hydrocarbons. Production chemicals, such as flocculants and corrosion and scale inhibitors, were also found to have an effect on organic concentrations.

Brown and coworkers [7] have provided information on water from Gulf of Mexico wells, in particular analyzing content both with and without silica gel treatment. Their results show that the water-soluble fraction is 79 to 98% polar in nature, predominantly organic acids, which are effectively extracted by silica gel treatment. The group also analyzed for PAH, which they found at the extremely low levels of 10 to 30 ppb by weight.

Neff et al. published a detailed analysis of produced water from wells in shallow water on the Louisiana shelf [8]. In comparison with that of deep-water wells, the organic contamination was found to be very high, perhaps arising from a comparatively high level of biological activity.

A survey of produced water from international data was prepared by Tibbetts et al. [9], showing that values for organics in North Sea produced water are similar to worldwide averages. However, the ranges of concentrations are very broad, some varying over 2 or 3 orders of magnitude. Tibbetts also identified many of the chemicals added during the oil-water separation process, which can affect measurements of organic loading.

Although the above-mentioned produced-water data sets show a great degree of variation, a few trends are apparent.

- Unless contamination of the reservoir occurs, the paraffinic oil and grease load in produced water is very low, within NPDES permits for offshore oil and grease. Entrained droplets, which have been measured as contributing up to $40 \text{ mg}\cdot\text{L}^{-1}$ * to the measured TPH, cannot be removed by gravity because of their small size ($<20 \mu\text{m}^3$).
- Oxidized organics are more soluble than aliphatic hydrocarbons. Because oxidation often occurs during separation and treatment, the concentrations of

* Note that $40 \text{ mg}\cdot\text{L}^{-1}$ approaches the daily maximum value for TPH as established by NPDES permits, and is higher than the long-term average level of $29 \text{ mg}\cdot\text{L}^{-1}$.

these organics at the wellhead are difficult to predict from knowledge of the geochemistry of formation waters,. These organics, however, represent a large fraction of the total extractable material (TEM) sampled using EPA methods. Organic acids, which are prevalent in all produced-water fractions, can have concentrations as high as 1000 ppm by weight.

- Other heteronuclear organics are not particularly prevalent in the water-soluble fraction, unless they are introduced to enhance production (e.g., amines).
- Volatile aromatic compounds can represent a significant fraction of the water-soluble organic material, depending on the source of the oil and the degree to which the oil and the water have been degassed before sampling. The concentrations of volatile aromatic compounds are difficult to predict with models based strictly on thermodynamic equilibrium calculations. Nonvolatile polyaromatic hydrocarbons are much less soluble, and these are present in parts-per-billions amounts in the water.
- There are few data on some organic compounds that may be relatively soluble in produced water, for example, ketones, aldehydes, mixed heteronuclear organics, mercaptans, and thiols.

1.2 WATER CHARACTERIZATION AT ORNL

Because more information on Gulf of Mexico wells was needed, the characterization of simulated produced water at ORNL was carried out in previous years using standard EPA analytical methods [10], with the objective to classify the organics into size categories (roughly equivalent to carbon chain length) and into chemical classes (aliphatic, aromatic, and polar). A sensitivity analysis was performed on solubility as a function of independent variables: temperature, pressure, pH, salinity, and water-to-oil ratio. These experiments were carried out to measure the effect of varying physical and chemical conditions on solubility and to derive data for model development.

Contact experiments were carried out with actual crude oil samples, with densities of about $0.85 \text{ g}\cdot\text{cm}^{-3}$. The aqueous phase, however, was prepared in the laboratory to approximate seawater composition. Most contacts were carried out under stirring for 4 days to ensure thermodynamic equilibrium. The oil and, after contact, the aqueous phase, were analyzed by inductively coupled plasma analysis for inorganic ions; by ion chromatography for organic acids; and by gas chromatography after fractionation on a packed column for aliphatic, aromatic, and polar components. Detailed experimental procedures are provided elsewhere [11].

The results of the characterization study [12] showed that of all of the physical variables tested, pH had the greatest effect on the solubility of organic compounds derived from oil. This finding suggested that a large fraction of these compounds are acidic, in agreement with work on North Sea crude oil done elsewhere [13]. Temperature, which was varied from 25 to 75°C, had a slight effect on solubility, which increased for heavier components. A decrease in solubility was observed for lighter components, a phenomenon that was attributed to losses through volatilization. No discernable trends

were observed for the other physical variables: pressure, salinity, or water-to-oil ratio. Because the analyses for the second crude oil sample were done primarily to provide comparative data for the more-detailed study of the first sample, few replicate tests were performed. Consequently the resulting uncertainty was high, $\pm 60\%$.

The data from the tests were examined to see if quantitative relationships could be established from the results and applied to prediction of organic solubilities, in particular those based on pH and temperature. An uncertainty analysis was performed on the data, not only to specify the level of confidence in the results but also to provide the variances required for the development of a statistical model. Results of the uncertainty analysis are summarized in Table 1. *

1.3 APPROACH TO MODELING

Results from the ORNL produced-water characterization studies were intended to support the development of a predictive model for produced-water contamination with organic compounds. Partitioning of organics between nonmiscible crude oil and brine is dependent on a number of factors. Results indicated that the pH of the brine was important in influencing the solubility of polar organic compounds, which make up most of the water-soluble component of crude oil.

* Estimates of the uncertainties in the measurement of the independent variables temperature and pressure relied on the calibrations provided by manufacturers. Temperatures were measured with type T thermocouples, certified to have an accuracy bounded by the greater of $\pm 1^\circ\text{C}$ or 0.75% of the reading (in degrees Celcius). Pressures, measured with a US Gauge, were estimated to be accurate to within ± 25 psi, in this case one-half of the lowest gradation on the scale. Measurements of pH were taken with an Orion model 520A. The calibration of the pH meter was repeated daily using standard phosphate buffers (J. T. Baker) with the following pH values at 25°C : 4.01 ± 0.01 , 6.99 ± 0.01 , and 9.98 ± 0.01 .

Simulated brines were prepared by weighing out salts as received from the manufacturer and dissolving them in deionized water measured using a volumetric flask. The calibration of balances used in the laboratory, Mettler AE-260 and PC4400, was checked monthly using standard weights traceable to the National Institute of Standards and Technology (NIST). Although the weighings were accurate to ± 0.0001 g, this accuracy did not propagate to the resulting salinity of the brine because the resulting solutions were saturated, particularly at high pH. Hence, salinity measurements for $[\text{Cl}^-(\text{aq})]$ were also measured with the Orion meter, using a $\text{Cl}^-(\text{aq})$ -specific electrode. The $\text{Cl}^-(\text{aq})$ concentrations were calibrated using standard solutions. Uncertainty in the salinity measurements was estimated at $\pm 10\%$.

Water cut was measured using the volume of the brine and the weight of the oil — converted to volume using measured density.

Mixing times were accurate to within ± 15 min, as measured on the laboratory wall clock. Although the clock could be read much more precisely than this, procedures in establishing the rotation of large barrels of oil necessitated a more generous assignment of uncertainty.

Concentrations of inorganic anions were measured using an Ion-Pac ICE-AS6 Dionex Corporation ion chromatograph.

Table 1. Results of uncertainty analysis in water-characterization experiments^a

	Measurement method	Uncertainty
Independent variables		
Temperature	Type T thermocouple	±1°C or 0.75%
Pressure	US Gauge	±25 psi
pH	Orion 520A meter	±0.01
Salinity	Orion 520A meter	±10%
Carbonate concentration	DMS Titrino (model 716)	±5 ppm, pH <9
Water cut	Volumetric + gravimetric	±0.01 volume fraction
Time	Wall clock	±15 min
Dependent variables		
Concentration of inorganic anions	ICP-AES	Highly variable ±2% for Ti, V ±130% for Ni
Water content	Brinkman 625F coulometer	±5 µg/L up to 1000 µg/L ±0.5% above 1000 µg/L
Total petroleum hydrocarbon	Gravimetric analysis on CH ₂ Cl ₂ -extracted sample	±20% crude sample 1 ± 4% crude sample 2
Organic fractionation and GC analysis	Open liquid column HP GC	±40% oil components ±60% water-soluble organics

^a GC=gas chromatography; HP GC = high-performance gas chromatography; ICP-AES = inductively coupled plasma–atomic emission spectroscopy.

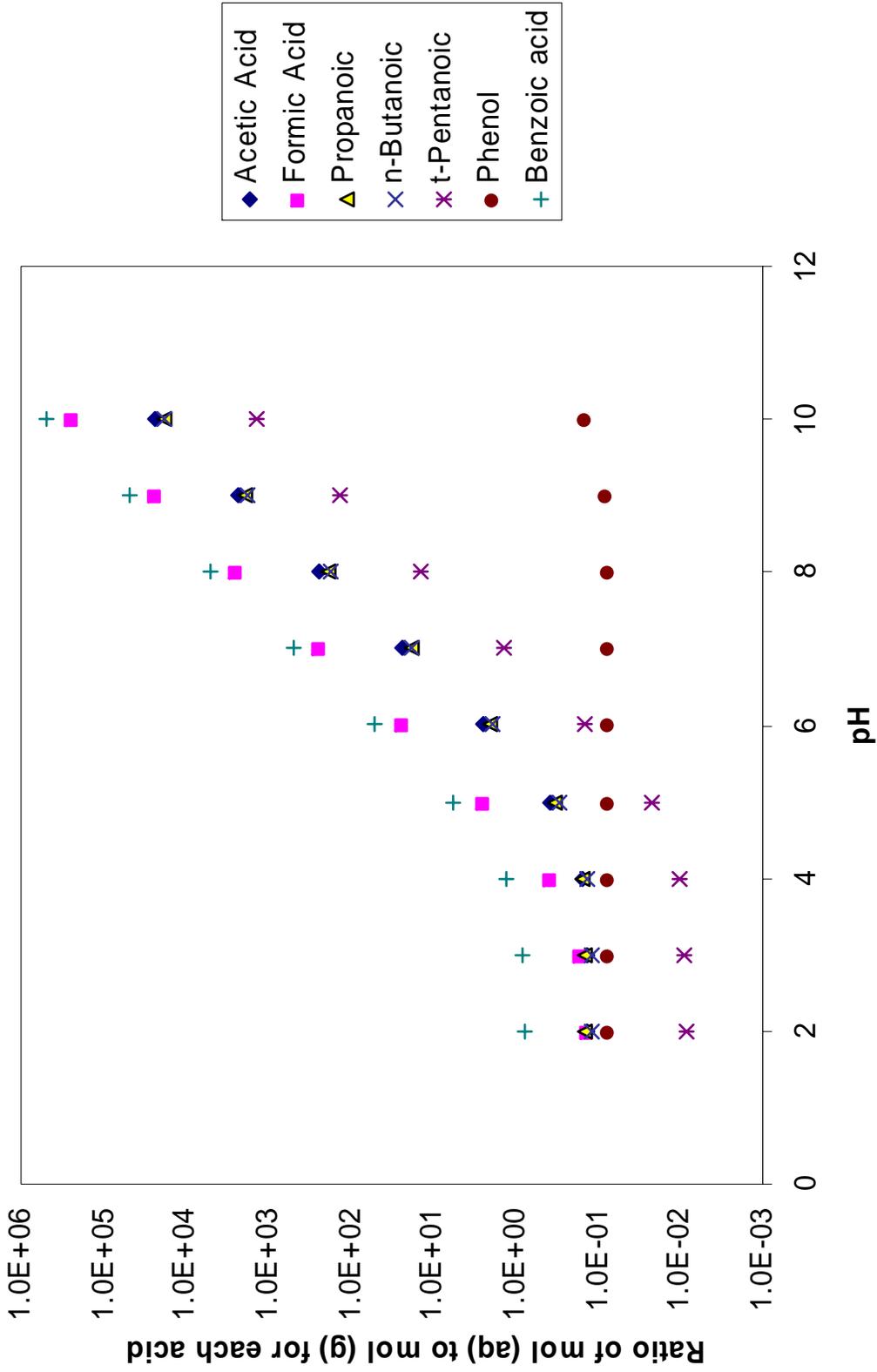


Fig 1. Dependence of Henry's Law constants of polar organics on pH. This graph give the moles of acid in equilibrium with a given number of moles of gas, ratioing the Henry's Law constant with the acid dissociation constant as a function of pH.

Following the method of Khan et al. [14], the relationship between the negative logarithm of the acid-dissociation constant, pKa, and organic acid solubility has been plotted as a function of pH (Fig. 1). In this graph, the ordinate represents the ratio of the moles of acid in the aqueous phase (both dissociated and undissociated) in equilibrium with the moles acid in the head space. As can be seen in the Figure, a clear increase in solubility occurs at the pH corresponding to the pKa, or at the point of increased dissociation in solution. The effect of the Henry's law constant is apparent when comparing the solubility of neovaleric acid with that of the others. Although the pKa of neovaleric acid is comparable with that of the other acids, its solubility curve is shifted downwards because of its low K_h , as defined in Eq(1).

$$K_h = \frac{m_A \gamma_A}{f_A}, \quad (1)$$

where m_A is the molality of the acid in the brine, γ_A is the activity coefficient, and f_A is the fugacity in the gas phase. By selecting the values of pKa and the K_h , one can generate many different solubility curves. It is important to note that simple organic acids, phenols, and ketones have pKa values of ~4.8, ~10, and ~20, respectively — none of which would predict an inflection point close to 7, as observed in the PERF data (Fig. 2). This indicates that the properties of the mixture must be included when modeling organic solubilities, as has been noted for other systems involving carboxylic acids [15].

The work of Khan et al. [14] indicated that the temperature dependence of the organic acid solubility could be accounted for by changes in the Henry's law constant, following an Arrhenius function. The group also investigated the effect of salinity on solubility, using Pitzer-type interaction coefficients to calculate changes in activity coefficient. Salinity, however, was found to have only a major impact on the most highly soluble organic acid, pyruvic acid. This is in agreement with ORNL findings that variations in salinity did not have a measurable effect on the solubility of low concentration organics, within the resolution of the measurements. However, as will be discussed later, ionic strength does affect the dissociation of carboxylic acids and thus does affect the solubility of these compounds.

From these basic concepts, ORNL developed a model for the solubility of semi-volatile organics in produced water, based on thermochemical properties.

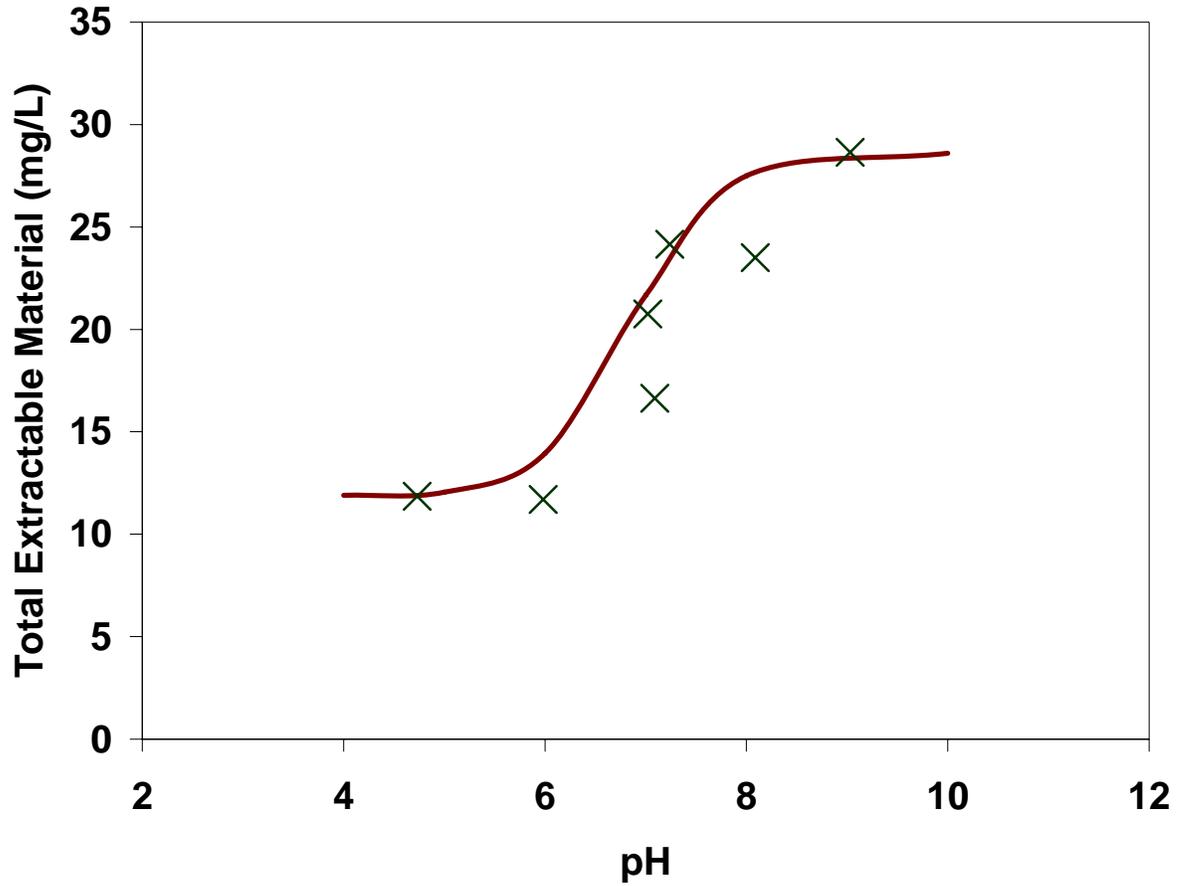


Fig. 2. Comparison of model results with total extractable material data from contact experiments of simulated produced-water brine with actual crude oil.

2. THEORY

2.1 LIQUID-LIQUID EQUILIBRIUM MODEL

The liquid-liquid equilibrium between components of petroleum and brine was modeled using a chemical thermodynamic equilibrium calculation [16], the model being based on an iterative solution of the Rachford-Rice equation.

$$\sum_{i=1}^c (-x_i + u_i + d_i) = 0 \quad (2)$$

and

$$\sum_{i=1}^c z_i \frac{(-H + Kw_i H + Kw_i Ka_i \Gamma_i)}{-\beta H + \beta Kw_i H + \beta Kw_i Ka_i \Gamma + H} = 0 \quad (3)$$

In the equations above, mole fractions of the i^{th} component in the overall system, the hydrocarbon phase, and undissociated and dissociated ions in the aqueous phase are represented by z_i , x_i , u_i , and d_i , respectively. The aqueous–hydrocarbon phase split is represented by β . For each of the c , components the aqueous–hydrocarbon equilibrium constant K_{w_i} , the “acid” dissociation constant Ka_i , and the activity of the dissociated ions in solution Γ_i , are also given. (The inclusion of Γ allows correction for ionic strength.) Obviously, for most compounds other than the organic acids, the terms involving Ka_i will be negligible. The hydrogen ion concentration is given by H . More details can be found in Appendix A, where the complete MAPLE 7 worksheets are given along with representative outputs.

An activity coefficient model was used to describe component behavior in the two nonmiscible liquid phases, hydrocarbon and aqueous. Because the compounds under consideration had little or no volatility under the conditions of the experiment, the vapor phase was not included in the model. The activity coefficients were determined from a liquid–liquid equilibrium database of UNIFAC coefficients based on a functional group analysis [17]. This model incorporates a “combinatorial” component to the activity coefficient based on the size and shape of moieties in the molecule and a “residual” component that incorporates experimentally determined binary interaction parameters between function groups. In particular, UNIFAC parameters for liquid–liquid equilibrium were used in the model [18]. A Non-Random Two-Liquid (NRTL) analysis to generate activity coefficients was also used during testing of the procedure [19]. However, this method is limited to binary or ternary systems and an extension to multicomponent systems is difficult to defend.

Input to the calculation was randomly sampled to demonstrate how uncertainties in the input data were reflected in the results of the computation. The model successfully reproduced parametric studies carried out at ORNL, allowing explanation of changes in solubility observed with variations in pH (refer to the solid line in Fig. 2). The main drawback to the chemical thermodynamic approach (or any phenomenological treatment), however, is that the system must be understood in sufficient detail to ensure that the assumptions of the model are valid. This is problematic in a system comprising ~1000 components, many of which are present in concentrations that are not well known or that contain moieties (e.g., sulfur-functional groups) for which the activity coefficient database is weak.

2.2 PARTIAL LEAST-SQUARES ANALYSIS

The use of statistical simulation to investigate correlations between process variables and outcomes is widely used within the physical and social sciences. Different problems demand different methods, but all are based on a simplification of the true mechanism, with a response surface calculated based on sampling of experimental or input data. Partial least-squares (PLS) analysis is a projection method that is well suited for

dependent, highly correlated, or even qualitative variables [20]. The method has been used by chemists to evaluate the best dispersant to control environmental contamination from an oil spill [21]. Brandvik and Daling were able to predict the effectiveness of surfactant blends based on disparate properties such as viscosity, density, toxicity, and cost. Hence, PLS was selected for predicting solubilities in multidimensional phase space (composition, temperature, pH, salinity, and oil-to-water ratio).

PLS analysis was performed using the Nonlinear Iterative Partial Least-Squares (NIPAL) algorithm, which involves preparing matrices of dependent and independent variables, centered on the mean and scaled by the variance. Model development involves a stepwise breakdown of matrices — in terms of scores, loadings, and weights matrices, as well as regression parameters — to give an alternative description of the data in terms of a set of orthogonal eigenvectors. Predictions were based on the eigenvectors and regression parameters. Plots of scores allowed assessment of the goodness of fit (regression between predictor and response variables) and correlations between predictor variables or responses. The model was applied to the characterization data obtained at ORNL [22]. Predictions for solubility as a function of pH, presented in Fig. 3, show an increase in solubility with increasing pH.^{*} Although the analysis is inherently linear in approach, this technique can be adjusted by applying the fit to a nonlinear set of variables, which in this case involved the use of aqueous-phase dissociation, rather than pH, as one of the input variables. Unlike a thermodynamic model, the analysis can be performed without any assumptions concerning the chemistry of the water or the oil and can even be applied to qualitative factors.

In summary, the primary advantage of the PLS model is that field data can be easily incorporated into the model, and for this reason, it is the favored approach for analysis of environmental data. Of course, the validity of a statistical model depends on the range of input variables, with the objective of sampling as many conditions as possible. Statistics permits an unambiguous sensitivity analysis, allowing a model to focus on key variables [23]. The statistical model can also incorporate the results of work on produced water collected worldwide: offshore and near shore in the Gulf of Mexico, the North Sea, and elsewhere. The nature of the water-soluble organics, as well as production variables (temperature, pH, additives, age of well, and addition of production chemicals), are similar throughout the world[†].

^{*} The plot is actually the first response variable versus the first predictor variable, although the transformation back to physical variables will find them closely aligned with organic acid concentration and degree of dissociation, respectively.

[†] Note that this assumption is not valid when considering the highly localized onshore environments. The author has addressed these issues in a paper titled “Offshore Versus Onshore Produced Water Characterization and Models”, in *Proceedings of GTI's Natural Gas Technologies II Conference and Exhibition*, February 8–11, 2004, Phoenix, Arizona.

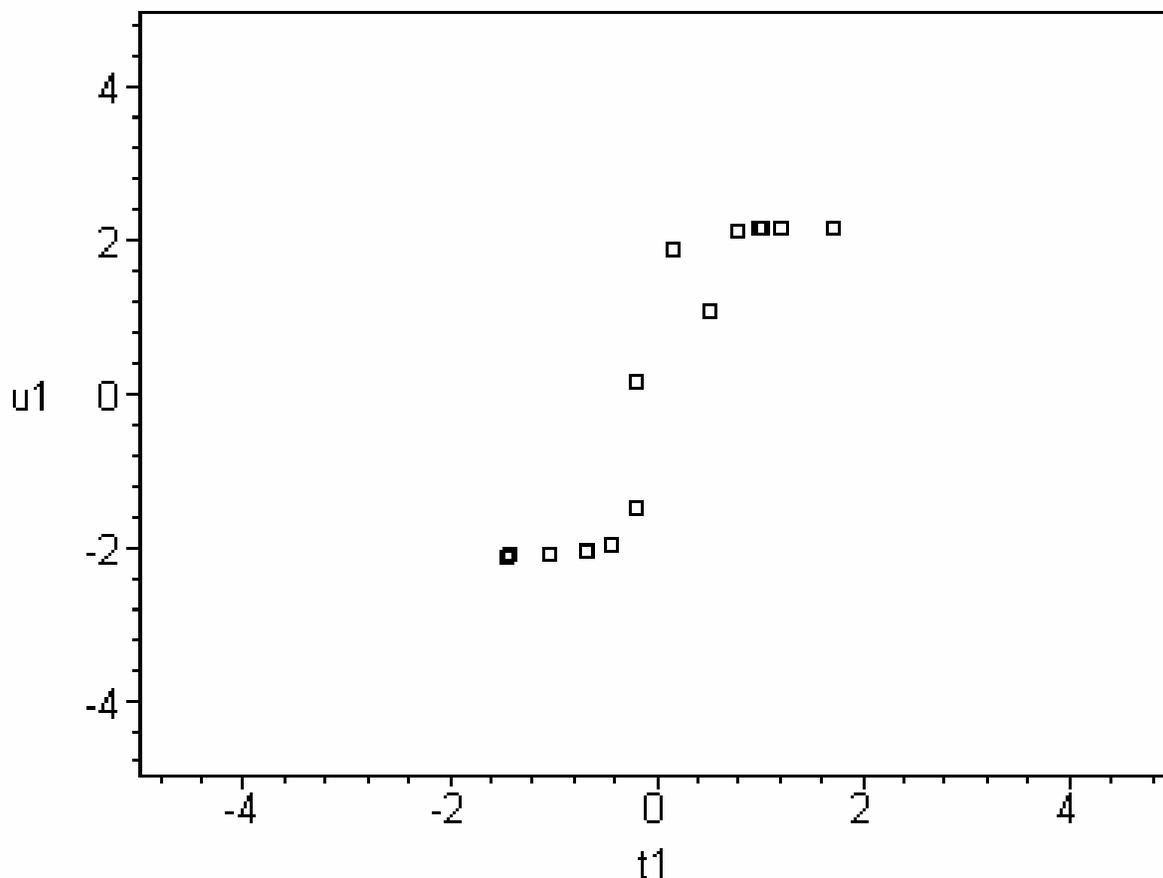


Fig. 3. Results of partial least-squares analysis, showing the leading independent factor (best aligned with concentration) vs the leading dependent factor (best aligned with pH).

3. APPLICATION TO PRODUCED-WATER MODELING

Ultimately, the goal of the project was to construct a predictive model that could be used *a priori* to give organic concentrations in produced water. However, insufficient information exists concerning the produced-water brine to accurately predict concentrations based on a phenomenological model. Not only are some data on chemical composition missing, but many factors important to organic solubility in produced water cannot be studied in the laboratory. Wave action, presence of additives, and turbulence in a downhole separator are examples of these important variables that were not quantified in this study. Hence, to achieve the objective of predictive model, a statistical procedure (i.e., PLS) was implemented to allow the incorporation of independent variables from the laboratory or the field. As the distribution of a particular organic component is dependent on the activities of the component in the aqueous versus hydrocarbon phase, the equilibrium constant Kw_i was first determined for representative organics based on a liquid-liquid equilibrium model. This value was then included as input to a PLS analysis, along with physical variables such as temperature of the system, pH, water-to-

oil ratio, and salinity. Pressure was not included in the analysis because it was found to have a negligible effect on solubility of low-vapor-pressure organics [11].

Water-soluble organics were grouped into classes based on thermodynamic calculations of solubility. A representative system was incorporated into the matrix of predictor variables, allowing the model to be applied to systems for which detailed chemical information was not known. It was found that the physical behavior within each class of organics was similar, and hence, one such component could be representative of the others in the class. Finally, to evaluate the performance, the model was then tested against characterization data measured in the laboratory at ORNL. The only previously established link between the measurements and the model was in the selection of predictor variables. The comparison was encouraging, indicating that the model did a reasonable job of representing the reality of produced-water contamination. Details on the methodology used for model development are provided in this section.

3.1 CONCENTRATIONS OF WATER-SOLUBLE ORGANICS IN CRUDE OIL

Crude oil is described by industry sources in terms of American Petroleum Institute (API) gravity (inversely related to density); sulfur content; pour point; fraction of saturates, aromatics, resins, asphaltenes, waxes, and carbon residue; sulfur content; salt content; nitrogen content; distillation range; and metals content [24]. The fraction of polar organics is not published in industrial databases. Fractionation after distillation is often given, but our focus is on as-produced crude. A statistical principal component analysis (PCA) was performed on these “field-type” variables to determine whether they are correlated with water-soluble organic fractions. A similar approach has been used by Barbieri and coworkers to examine seasonal correlations in freshwater analyses [25].

An EPA publication provides the analysis of a selection of crude oils from diverse geographical locations [26] that were used in a PCA analysis. Table 2 contains this information as well as data from Environment Canada for a selection of crude oils from the Gulf of Mexico and the North Sea [27]. A plot of the vectors from the PCA analysis indicates which components are correlated with the physical properties of the oil (Fig. 4). The numbers in the Figure correspond to the first ten oils listed in Table 2.

Table 2. Properties of selected crude oils									
	API gravity	Sulfur (ppm)	Pour Point (°C)	Oil-water interfacial tension (mN·m ⁻¹)	Saturates (wt %)	Aromatics (wt %)	Resins (wt %)	Asphaltenes (wt %)	Waxes (wt %)
Alaska North Slope	30.89	1.11	-32	22.5	75.0	15.0	6.1	4.0	2.6
Alberta Sweet	35.72	0.63	-18	19.2	77.3	16.8	4.2	1.7	3.2
Arabian Light	31.30	1.93	-21	21.3	75.5	15.2	5.7	3.6	2.7
Sockeye—CA	19.32	4.51	-25	23.0	49.2	17.2	15.1	18.5	1.6
S. Louisiana	32.72	0.49	-41	20.9	80.8	12.6	5.9	0.8	1.7
W. Texas	34.38	0.86	-22	18.8	78.5	14.8	6.0	0.7	2.8
Fuel oil#2—Ont. Diesel	37.52	0.09	-50	21.5	88.2	10.2	1.7	0.0	1.7
Fuel oil#5—NJ	11.55	1.00	-19	21.0	44.2	39.5	8.0	8.4	2.3
Heavy fuel oil 6303—NS	11.47	1.48	-1	21.0	42.5	29.0	15.5	13.0	2.5
Orimulsion—400	8.63	2.00	1	21.0	32.1	19.7	9.6	10.6	0
Venezuela									
Mars	27.6	2.07	-28	18.3	45	40	11	3	1.5
Genesis									
Brae	33.6	0.73	-6					0	
Beatrice	38.7	0.05	13						
Argyll			6						

Source: Adaptation of data from EPA [ref.26] and Environment Canada [ref.27].

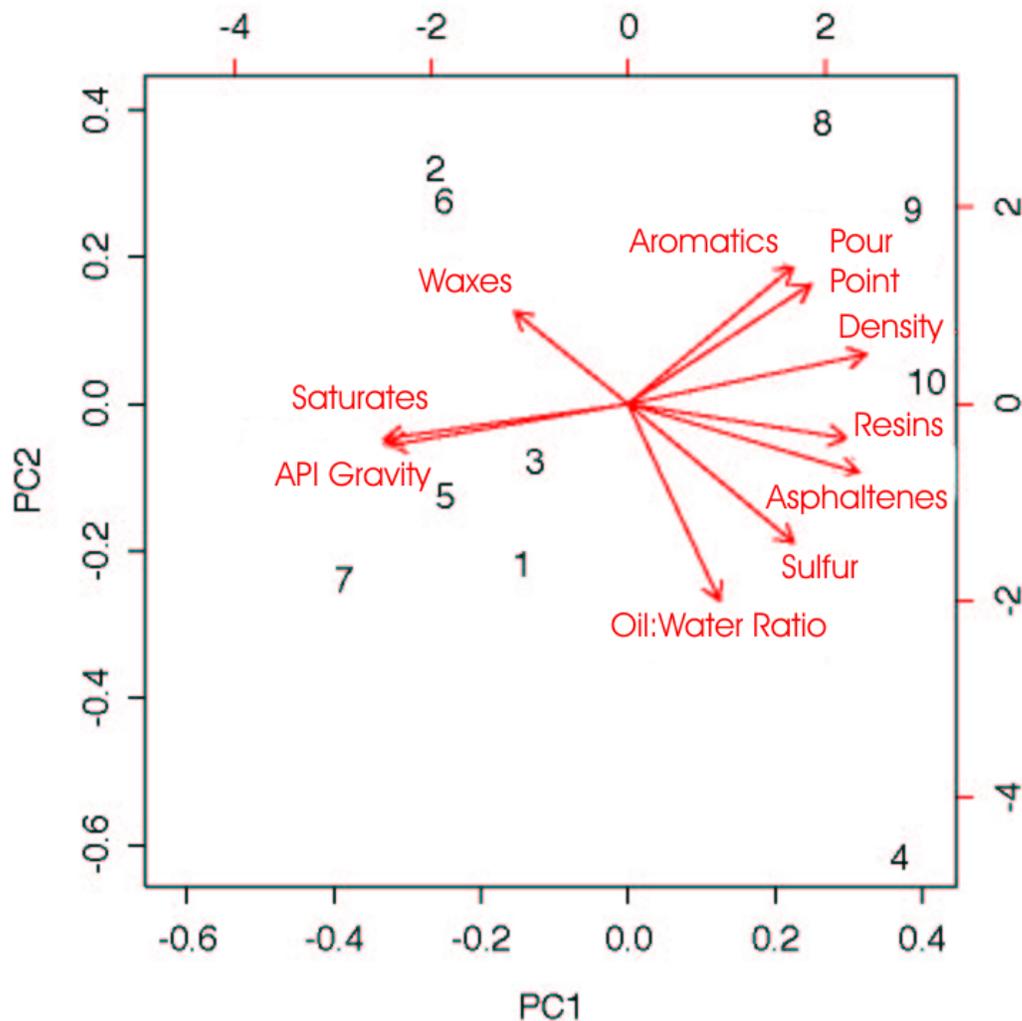


Fig. 4: Results of principal component analysis on crude oil properties.

The analysis revealed the following:

1. The oil compositions clustered according to API gravity or density.
2. Processed oil, diesel, or fuel oil (numbers 7 through 9) formed separate clusters and could also be distinguished by characteristic compositions and physical properties. These were not included in the model of produced-water composition, because processing has a great influence on properties such as pour point and oil–water interfacial tension.
3. High density (low API gravity) was correlated with the mass fraction of saturated hydrocarbons.
4. Sulfur content correlated with relatively high amounts of asphaltenes and resins and low wax content.

The objective of the classification of oil was to associate parameters describing the oil with the content of water-soluble and insoluble organic compounds present in the oil. For this reason, the main classes of organic compounds have been summarized by examples given in Tables 3, 4, and 5: namely, the aliphatic organics, the aromatic organics, and the polar organics. Each of these classes has been examined in turn to determine if representative compounds exhibit phase behavior that can be termed typical of the group. A similar procedure has been carried out on components of oil, in order to simplify representation of the chemical system in a model [28].

Principal component results in Fig. 4 indicate a correlation between API gravity and the distribution of paraffinic organics in crude oil. The physical properties of a number of aliphatic organics are given in Table 3 [29, 30], along with classification in terms of cut point [27]. Figure 5 plots of water/octane distribution coefficients ($D_{w/hc}$) versus density in $\text{g}\cdot\text{mL}^{-1}$ of the organic compound. The plot shows two exponential correlations: one for paraffinic compounds and the other for naphthenic (saturated ring) compounds. When the results are recast in terms of gram molecular mass, the two groups of alkanes fall on the same exponential curve, as shown in Fig. 6.

$$D_{w/hc} = 7.06 e^{-0.100M} \quad (4)$$

The correlation between pour point and aromaticity that is evident in the vectors in Fig. 4 may be described in terms of the Watson characterization factor or the correlation index, both terms that are based on the boiling point [24]. Volatile organic compounds, such as benzene, also have appreciable solubility in water. However, because of losses, this factor is difficult to predict from unpressurized sampling, as was done for the ORNL analysis. In offshore production, it is the less-volatile aromatic compounds that may pose a problem with NPDES permitting, and these are the ones that are considered in a model of organic solubility. Selected aromatic organics are given in Table 4, along with their viscosities at temperatures close to ambient [31].

Table 3. Properties of selected aliphatic, olefinic and naphthenic compounds in crude oil^a

	Chemical formula	Boiling point (K) at Atmospheric pressure (1 bar)	Specific gravity (g·mL ⁻¹) at 293 K ^b	Petroleum fraction TBP ^c
Methane	CH ₄	111.2	Not a liquid	Volatile
Ethane	C ₂ H ₆	184.6		
<i>n</i> -Propane	C ₃ H ₈	231.1		
<i>n</i> -Butane	C ₄ H ₁₀	273		
2,2-Dimethylpropane	(CH ₃) ₂ C ₃ H ₆	282.6	0.591	
1-Pentene	C ₅ H ₁₀	303.1	0.640	
<i>n</i> -Pentane	C ₅ H ₁₂	309.2	0.626	Light gasoline
Cyclopentane	C ₅ H ₁₀	322.4	0.745	
2,2-Dimethylbutane	C ₆ H ₁₄	322.8	0.649	
<i>n</i> -Hexane	C ₆ H ₁₄	341.9	0.659	
cyclohexane	C ₆ H ₁₂	353.8	0.779	
<i>n</i> -Heptane	C ₇ H ₁₆	371.6	0.684	Naptha
Methylcyclohexane	C ₇ H ₁₄	374.1	0.774 (289 K)	
<i>n</i> -Octane	C ₈ H ₁₈	398.8	0.703	
<i>n</i> -Nonane	C ₉ H ₂₀	424.0	0.718	
<i>n</i> -Decane	C ₁₀ H ₂₂	447.3	0.730	
<i>trans</i> decalin	C ₁₀ H ₁₈	460.5	0.870	
<i>n</i> -Dodecane	C ₁₂ H ₂₆	489.5	0.746	Kerosene
<i>n</i> -Tetradecane	C ₁₄ H ₃₀	526.7	0.763	
<i>n</i> -Hexadecane	C ₁₆ H ₃₄	575.2	0.773	Light gas oil
<i>n</i> -Octadecane	C ₁₈ H ₃₈	589.5	0.777 (301 K)	
<i>n</i> -Eicosane	C ₂₀ H ₄₂	617	0.775 (313 K)	Atm. gas oil
<i>n</i> -Octacosane	C ₂₈ H ₅₈	704.8		Vac. gas oil

^a A blank for an entry indicates that no information is available for that organic.

^b Other temperatures are indicated parenthetically.

^c TBP=true boiling point

Table 4. Properties of selected aromatic compounds in crude oil^a

	Chemical formula	Boiling point (K) at atmospheric pressure (1 bar)	Specific gravity (g·mL ⁻¹) at 293 K ^b	Viscosity (cP)
Benzene	C ₆ H ₆	353.3	0.885 (289 K)	0.608
Toluene	C ₇ H ₈	383.8	0.867	0.556
<i>o</i> -Xylene	C ₈ H ₁₀	417.6	0.880	0.760
1-methyl ethylbenzene	C ₉ H ₁₂	435	0.862	0.731
2,3-dihydro-1h-Indene	C ₉ H ₁₀	449.7	0.9639	1.348
Naphthalene	C ₁₀ H ₈	491.1	0.971 (363 K)	
1-Methyl naphthalene	C ₁₁ H ₁₀	517.9	1.020	
Phenanthrene	C ₁₄ H ₁₀	613	1.179 (298 K)	
Anthracene	C ₁₄ H ₁₀	613.1	1.25 (300 K)	

^a A blank for an entry indicates that no information is available for that organic compound.

^b Other temperatures are indicated parenthetically.

Table 5. Properties of selected polar compounds in crude oil^a

	Chemical formula	Boiling point (K) at atmospheric pressure (1 Bar)	Specific gravity (g·mL ⁻¹) at 293 K ^b	pKa
Tetrahydrofuran	C ₄ H ₈ O	304.5	0.938	
Methanol	CH ₄ O	337.7	0.791	15.2
Ethanol	C ₂ H ₆ O	351.4	0.789	16
<i>n</i> -Butyraldehyde	C ₄ H ₈ O	348.0	0.802	16
Ethylacetate	C ₄ H ₈ O ₂	350.3	0.901	24.5
Methylpropionate	C ₄ H ₈ O ₂	352.8	0.915	24.5
Thiophene	C ₄ H ₄ S	357.2	1.071 (289 K)	
Propanol	C ₃ H ₈ O	370.3	0.804	16
Formic acid	CH ₂ O ₂	373.8	1.226 (288 K)	3.751
Valeraldehyde	C ₅ H ₁₀ O	376	0.810	16
Pyridine	C ₅ H ₅ N	388.4	0.983	5.17
<i>n</i> -Butanol	C ₄ H ₁₀ O	390.9	0.810	16
Acetic acid	C ₂ H ₄ O ₂	391.1	1.049	4.756
<i>n</i> -Butylacetate	C ₆ H ₁₂ O ₂	399.3	0.898 (273 K)	24.5
Pyrrole	C ₄ H ₅ N	403.0	0.967 (294 K)	
Propanoic acid	C ₃ H ₆ O ₂	414.5	0.993	4.874
Cyclohexanone	C ₆ H ₁₀ O	428.8	0.951 (288)	19-20
<i>n</i> -Propylisovalerate	C ₈ H ₁₆ O ₂	429.1	0.863	24.5
1-Hexanol	C ₆ H ₁₄ O	430.2	0.819	16
Furfural	C ₅ H ₄ O ₂	434.9	1.159	4.817
Butyric acid	C ₄ H ₈ O ₂	437.2	0.958	4.817
1-Pentanoic acid	C ₅ H ₁₀ O ₂	449.7	0.886 (288 K)	4.842
1-Heptanol	C ₇ H ₁₆ O	449.8	0.822	16
1-Octanol	C ₈ H ₁₈ O	468.3	0.826	16
Hexanoic acid	C ₆ H ₁₂ O ₂	478	0.9265	4.849
Phenol	C ₆ H ₆ O	455.0	1.059 (313 K)	4.849 (293 K)
Aniline	C ₆ H ₇ N	457.6	1.022	9.99
Methylphenylketone	C ₈ H ₈ O	474.9	1.032	4.60
1-Phenylethanone	C ₈ H ₈ O	475	1.033	19-20
<i>o</i> -Ethylphenol	C ₈ H ₁₀ O	491.6	1.037 (273 K)	10.07
Thianaphthene	C ₈ H ₆ S	494		
1-Decanol	C ₁₀ H ₂₂ O	506.1	0.830	16
Quinoline	C ₉ H ₇ N	510.8	1.095	4.80
Ethyldecanoate	C ₁₂ H ₂₄ O ₂	518.2	0.862	24.5
Benzoic acid	C ₇ H ₆ O ₂	523	1.075 (403 K)	4.204
Dodecanol	C ₁₂ H ₂₆ O	533.1	0.835	16
Decanoic acid	C ₁₀ H ₂₀ O ₂	543	0.8782	5
Heptadecanol	C ₁₇ H ₃₆ O	597	0.848 (327 K)	16
Dibenzothiophene	C ₁₂ H ₈ S	605.7		
1-Eicosanol	C ₂₀ H ₄₂ O	629		16

^a A blank for an entry indicates that no information is available for that organic.

^b Other temperatures are indicated parenthetically.

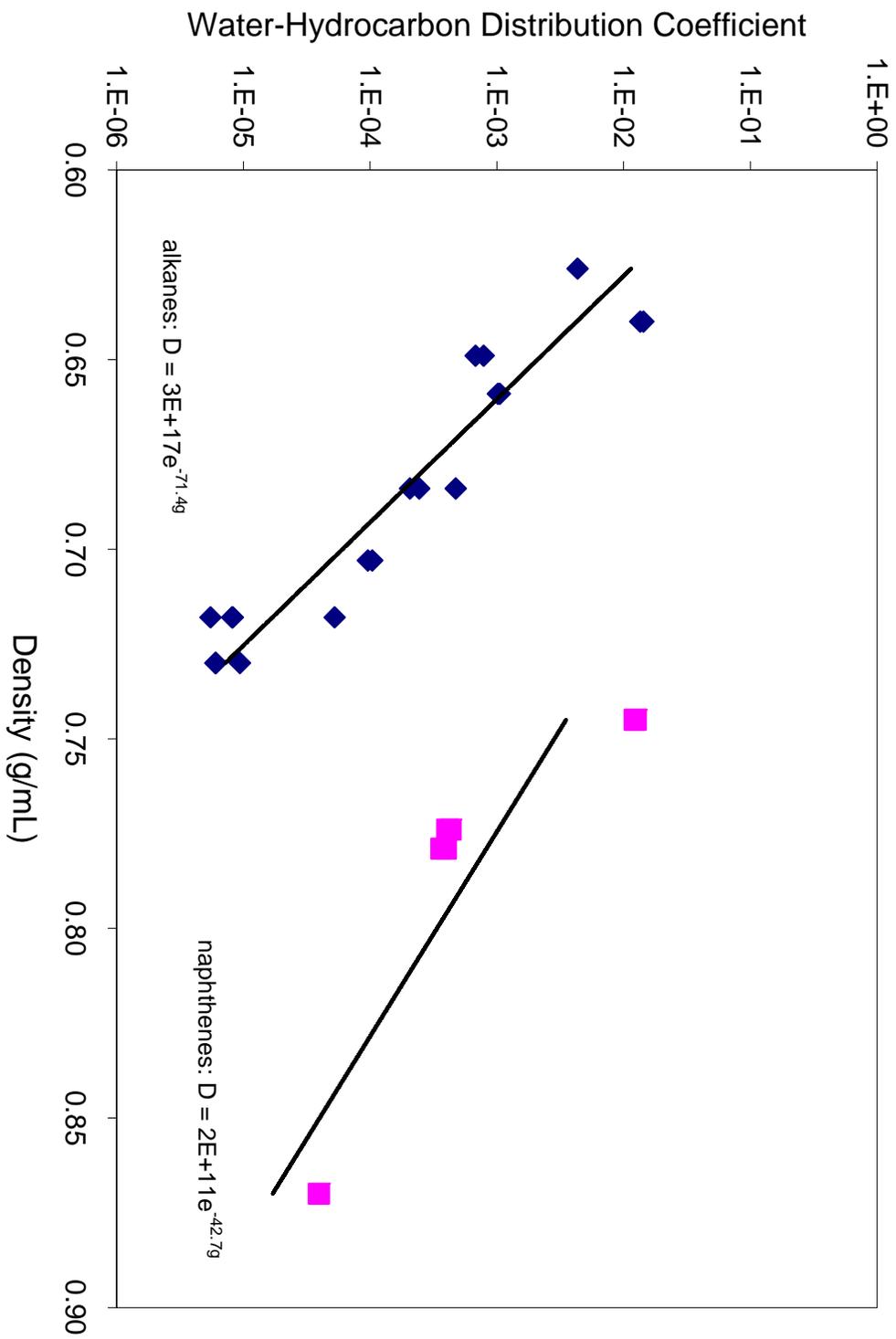


Fig. 5. Correlation of water-hydrocarbon partitioning with density for paraffinic and naphthenic compounds. Paraffinic compounds are indicated by diamonds; naphthenic compounds are indicated by squares. The fits to the results are given on the graph.

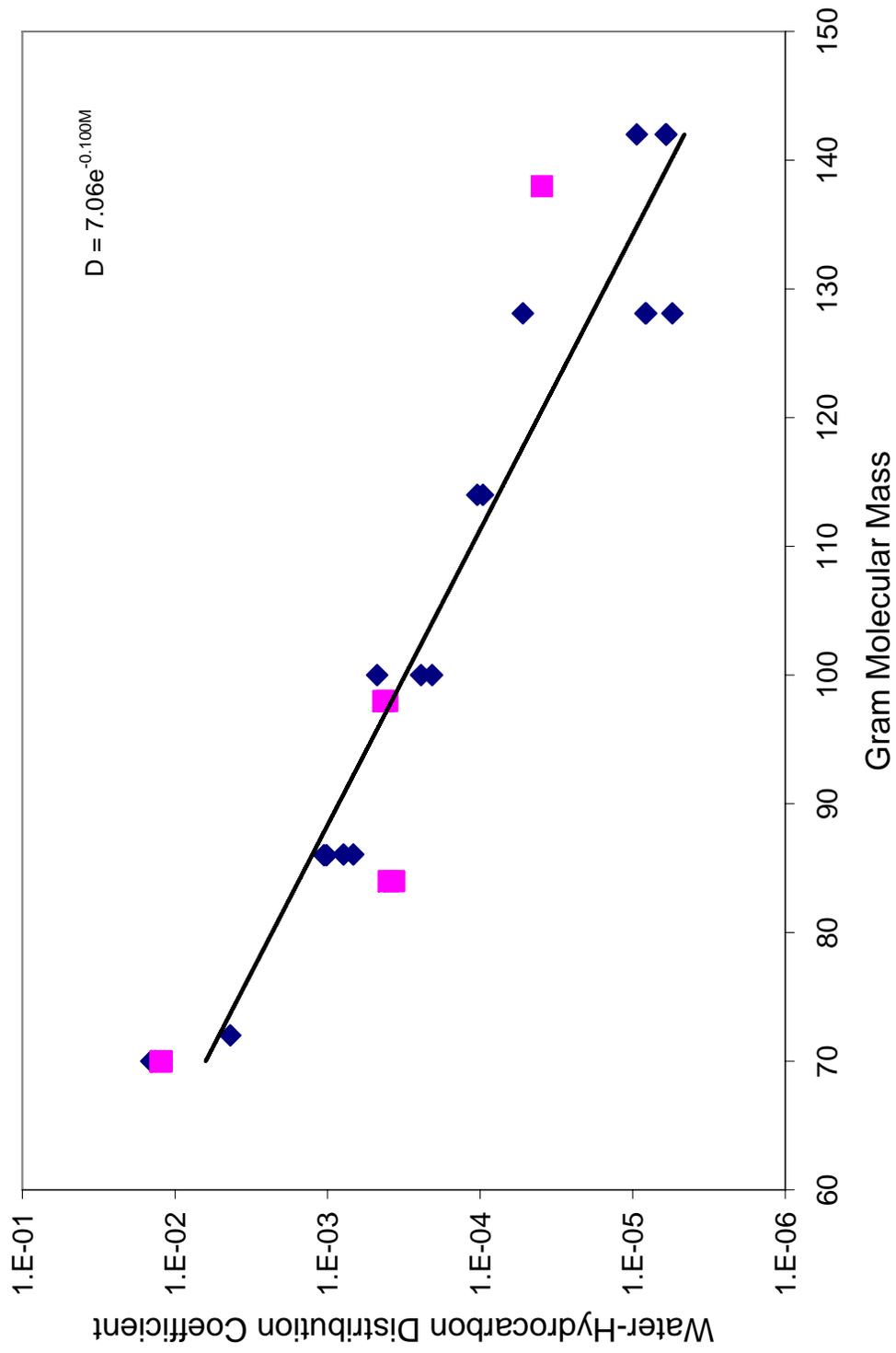


Fig. 6. Correlation of water-hydrocarbon partitioning with gram molecular mass for paraffinic and naphthenic compounds. Paraffinic compounds are indicated by diamonds; naphthenic compounds are indicated by squares. The fit to the results is given on the graph.

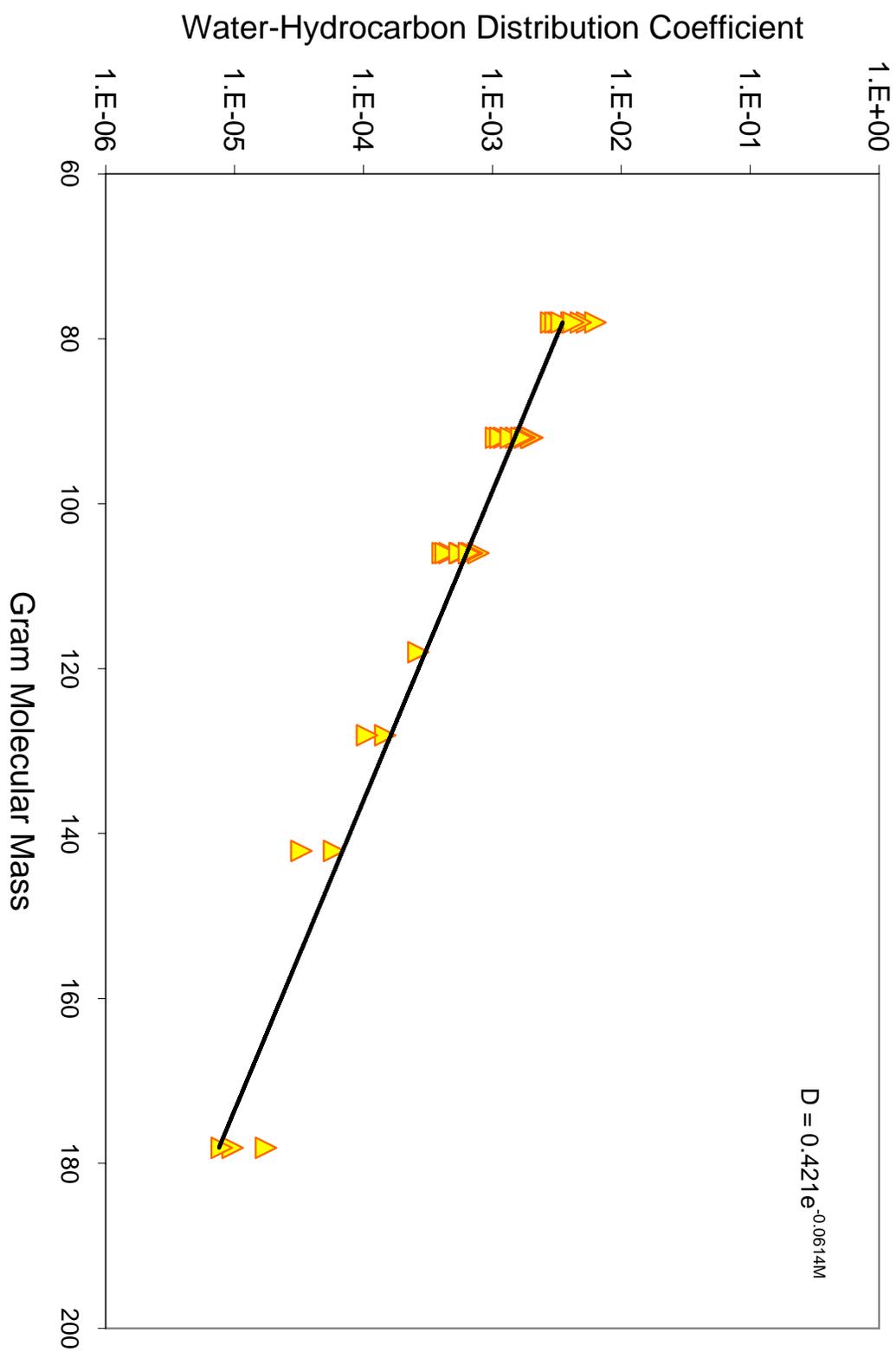


Fig. 7. Correlation of water-hydrocarbon partitioning with gram molecular mass for all aromatic compounds considered. The fit to the results is given on the graph.

Water/octane distribution coefficients as determined by thermodynamic calculations are presented in Fig. 7. The plot of water/octane distribution coefficient versus gram molecular mass (M) shows an exponential correlation between the two for all of the aromatic compounds being considered.

$$D_{w/hc} = 0.421 e^{-0.0614M} \quad (5)$$

The most difficult parameter to quantify is also the most crucial in describing water-soluble organic content, or the amounts and identities of polar organics in water. Although these parameters are somewhat related to sulfur and residue content, as indicated in Fig. 4, a simple relationship is not apparent in the PCA. The composition of aqueous fluids has been reviewed by Kharaka and Hanor [32], with a section devoted to reactive or polar organics*. From the review, ORNL data, and information presented elsewhere in the literature, it is apparent that polar organic content is based primarily on pH, temperature, age of the formation, and salinity. The effect of other factors such as aeration during separation, adjustment of pH, and addition of surfactants, become important during processing [6] but these are beyond the scope of the current project. In this case, we shall base our analysis on the data that have been published on water-soluble organics, a selection of which are given in Table 5, along with associated pKa values [33]. The water/octane distribution coefficient of the polar organics has been plotted in Fig. 8 as a function of normal boiling point in Kelvin, T_b . On the same figure are plotted computed distribution coefficients for aliphatic and aromatic hydrocarbons. The organics can be grouped into the following classes, with a correlation for each.

Organic acids	$D_{w/hc} = 4 \times 10^{13} e^{-0.0699T_b / K}$	(6)
---------------	--	-----

Alcohols and ketones	$D_{w/hc} = 2 \times 10^{10} e^{-0.0621T_b / K}$	(7)
----------------------	--	-----

Aldehydes and esters	$D_{w/hc} = 2 \times 10^7 e^{-0.0567T_b / K}$	(8)
----------------------	---	-----

Aromatic	$D_{w/hc} = 9 e^{-0.0225T_b / K}$	(9)
----------	-----------------------------------	-----

Aliphatic	$D_{w/hc} = 3 \times 10^4 e^{-0.0492T_b / K}$	(10)
-----------	---	------

The concentration of aliphatic organic acids is inversely proportional to temperature because the rate of decarboxylation increases rapidly, a process that converts the organic acid to natural gas and carbon dioxide [34]. However, at wellhead temperatures below 80°C, the rate of decarboxylation is slow and the reaction will not give rise to dramatic changes in organic acid concentration. Hence, the concentration of the organic acids can be assumed to be fixed at the temperature of the formation, in a manner similar to that discussed for the paraffinic and aromatic organic compounds.

* The term “reactive” organics arises from the fact that a mechanism for abiotic formation of natural gas is through the decarboxylation of short-chained aliphatic acids and is thought to be a major route to natural gas formation at temperatures between 200 and 300°C: $CH_3COOH \rightarrow CH_4 + CO_2$.

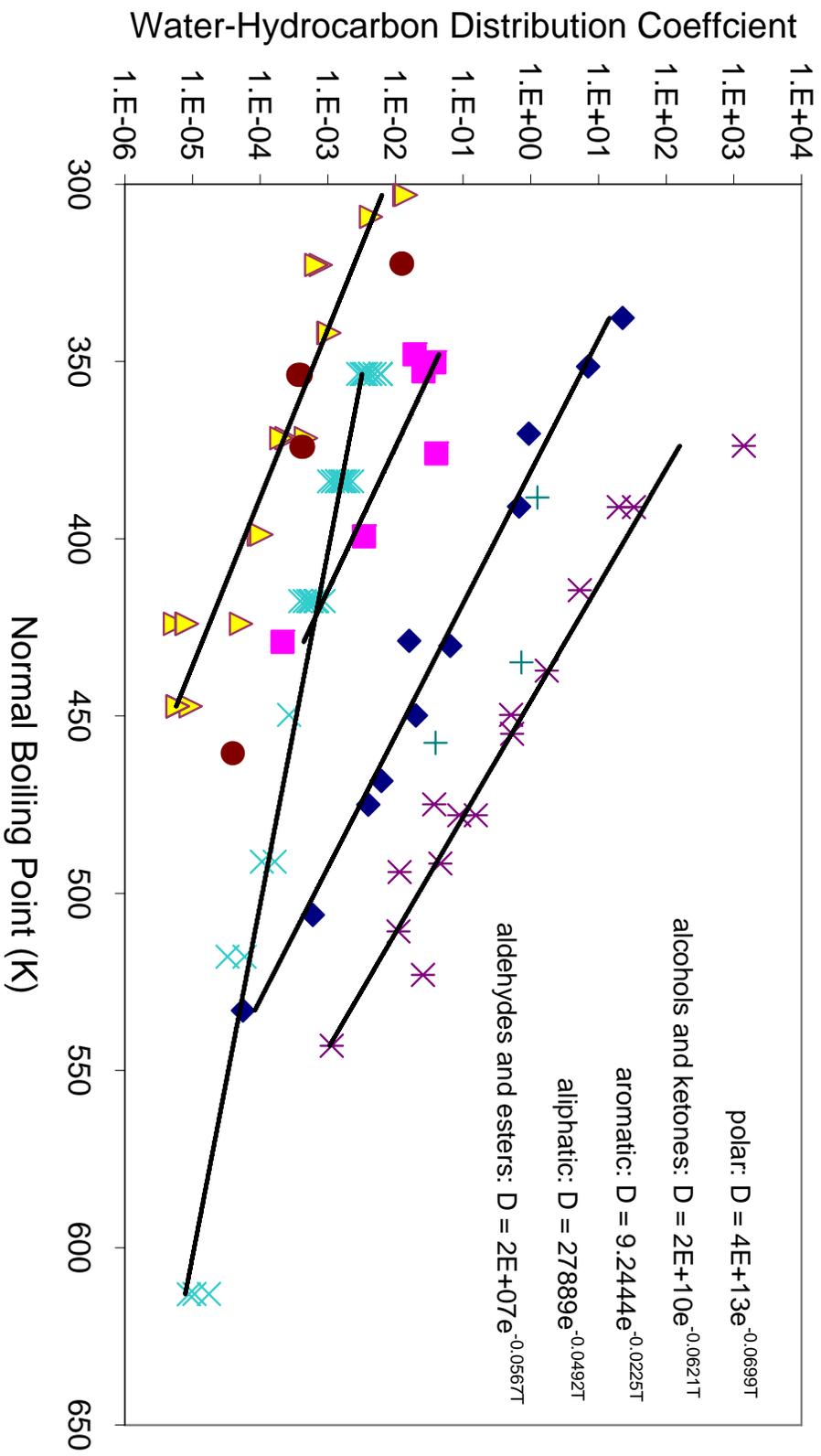


Fig. 8. Correlation of water-hydrocarbon partitioning with normal boiling point for all water-soluble organics considered. Symbols on the graph refer to the following types of organic compounds: triangles = aliphatics; circles = naphthenes; xs = aromatics; squares = aldehydes, ketones, and esters; diamonds = aliphatic alcohols; crosses = heterocyclic rings; and stars = polar compounds. The exponential fits to each series of computational results are given on the graph.

Calculated water/octanol distribution coefficients or K_{ow} , given in Fig. 9, show a comparison with literature values [35] for selected organics. The straight line on the chart indicates a 1:1 relationship. The calculated distributions agree well for most of the paraffinic and aromatic organics, which are indicated as diamonds. The deviation increases as the solubility drops, such as for *n*-heptane in this comparison. Note that the organics that have a pronounced dependence of solubility on pH are included on the chart as triangles for neutral pH, because pH dependence is typically not reported in K_{ow} compendia in the literature.

In summary, we have shown that the thermodynamic calculations based on solution of the Rachford-Rice equation for two-phase liquid–liquid equilibrium agree reasonably well with determinations of distribution coefficient in the literature. In addition, organics can be grouped based on their solubility in water: alkanes (paraffins and naphthenes), aromatics, organic acids, aldehydes and ketones, and alcohols and esters. This grouping will be used in the representation of the water-soluble organics in the model of the crude oil/produced-water brine system.

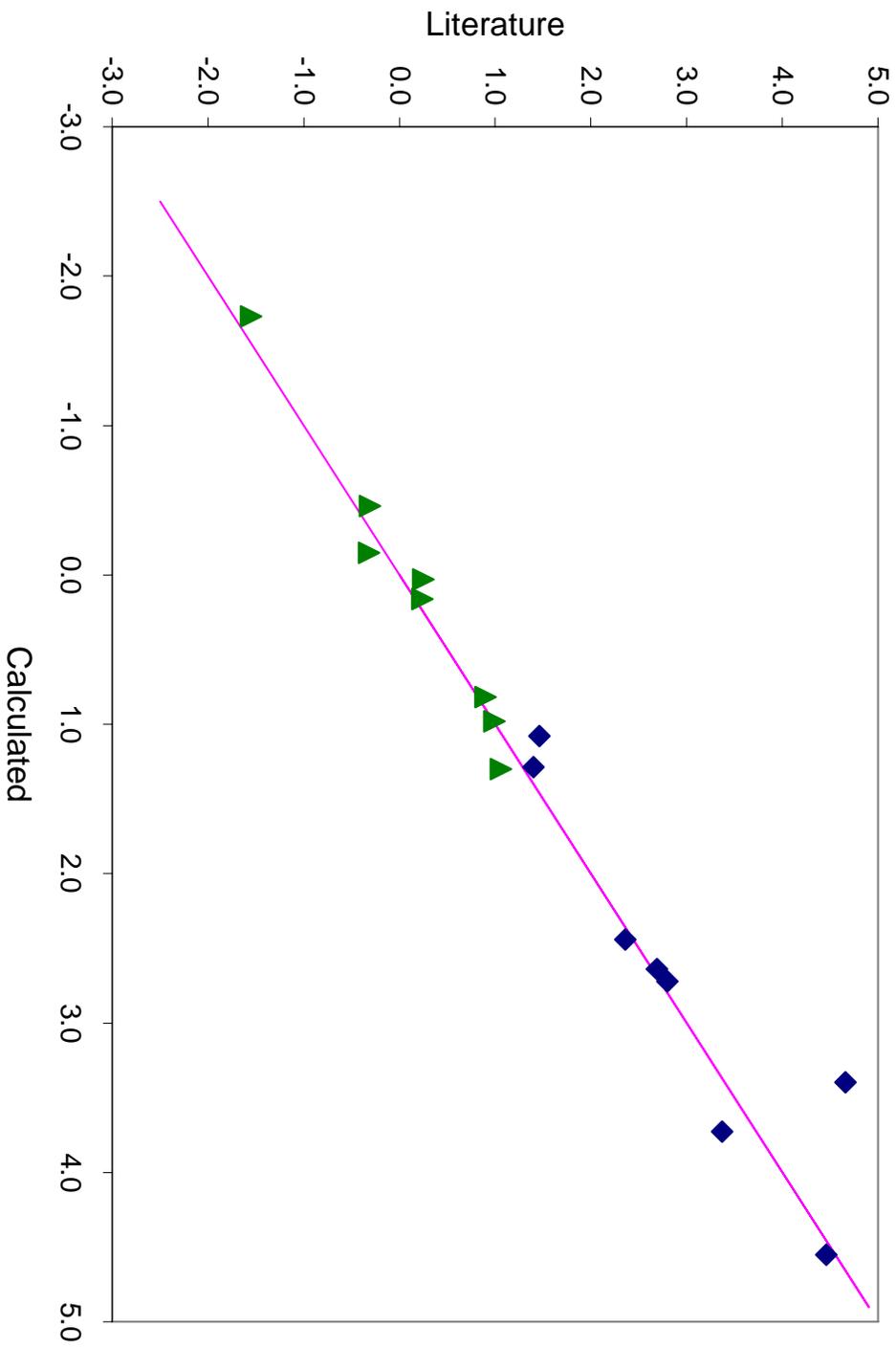


Fig. 9. Comparison of calculated vs literature values for octanol-water partition coefficients. Paraffinic and aromatic organics are indicated by triangles; organics that have a pronounced dependence of solubility on pH are indicated by diamonds.

3.2 RANGES OF PHYSICAL AND CHEMICAL VARIABLES

The physical variables used for the predictive model were selected to be the same as the independent variables in the ORNL water characterization experiments: temperature, pH, salinity, and water-to-oil ratio, because these are the conditions for which validation calculations were to be performed.* The ranges of the variables tested in the laboratory were selected to be representative of production conditions [36]. The temperatures ranged from ambient (25°C) to 75°C, the water-to-oil volumetric ratio from 20 to 80%, the pH from 3 to 10, and the salinity from 0 to 150,000 mg·L⁻¹ NaCl concentration.

The organics used in the calculation are listed in Tables 3 through 5. Aliphatic compounds were grouped by the fraction of oil that contained that particular compound, based on separation by distillation. Aliphatic organics were chosen so that all of the major fractions of crude oil were represented with at least one compound. Compounds that were not readily associated with a particular fraction of oil, particularly the polar organics, were represented by typical alcohols, ketones, aldehydes, esters, and heteronuclear organics. Although the selection was limited to 73 compounds, these compounds were chosen to represent as wide a range of chemical activity and solubility behavior as possible. Preliminary calculations were performed with all 73 compounds, with the subset further reduced to the compounds listed in *italics* in Table 6, each representative of a fraction of crude oil. The main criterion for selection was that the distribution coefficient determined in the first set of calculations be the closest to the mean for that group. Although most analyses available in the public domain do not give fractionation of crude oil into polar components, various polar compounds were included in the calculation to be grouped later for validation. Calculations are not shown for paraffinic organics heavier than the kerosene fraction (e.g., eicosane) because the aqueous mole fraction was found to be negligible. UNIFAC activity coefficients for liquid-phase sulfur compounds (e.g., thiophenes) are not available and thus were not included in the calculation.

* Pressure was not included in the analysis because it was found to have had little effect on the solubility of low-volatility organics that were assessed at ORNL. This finding was not unexpected because the samples had been degassed before arriving at the laboratory.

Table 6. Calculated distribution coefficients ($K_{W/HC}$) of selected organics

	Standard conditions as defined in text	At high temperature, 75°C	At pH = 10	At pH = 3
Light gasoline aliphatics				
<i>Pentane</i>	3×10^{-3}	5×10^{-3}	3×10^{-3}	3×10^{-3}
Light naptha				
<i>Heptane</i>	3×10^{-4}	6×10^{-4}	3×10^{-4}	3×10^{-4}
Heavy naptha				
<i>Nonane</i>	3×10^{-5}	7×10^{-5}	3×10^{-5}	3×10^{-5}
Kerosene				
<i>decalin</i>	2×10^{-5}	5×10^{-5}	2×10^{-5}	2×10^{-5}
BTEX (VOC)				
<i>Toluene</i>	2×10^{-3}	3×10^{-3}	2×10^{-3}	2×10^{-3}
NAP				
<i>Napthalene</i>	1×10^{-4}	2×10^{-4}	1×10^{-4}	1×10^{-4}
PAH				
<i>Phenanthrene</i>	2×10^{-5}	4×10^{-5}	2×10^{-5}	2×10^{-5}
Organic acids				
<i>Formic</i>	$1 \times 10^{+3}$	$7 \times 10^{+2}$	$1 \times 10^{+6}$	$8 \times 10^{+1}$
<i>Acetic</i>	$2 \times 10^{+1}$	$1 \times 10^{+1}$	$9 \times 10^{+3}$	$8 \times 10^{+0}$
<i>Propanoic</i>	$4 \times 10^{+0}$	$4 \times 10^{+0}$	$2 \times 10^{+3}$	$2 \times 10^{+0}$
<i>Butanoic</i>	$1 \times 10^{+0}$	$1 \times 10^{+0}$	$7 \times 10^{+2}$	7×10^{-1}
<i>Pentanoic</i>	4×10^{-1}	5×10^{-1}	$2 \times 10^{+2}$	2×10^{-1}
<i>Octanoic</i>	1×10^{-2}	1×10^{-2}	$7 \times 10^{+0}$	6×10^{-3}
Phenols				
<i>Phenol</i>	2×10^{-1}	2×10^{-1}	3×10^{-1}	2×10^{-1}
Alcohols				
<i>Ethanol</i>	$9 \times 10^{+0}$	$8 \times 10^{+0}$	$1 \times 10^{+1}$	$9 \times 10^{+0}$
<i>Pentanol</i>	3×10^{-1}	3×10^{-1}	3×10^{-1}	3×10^{-1}
<i>Octanol</i>	8×10^{-3}	1×10^{-2}	8×10^{-3}	8×10^{-3}
Ketones				
<i>Cyclohexanone</i>	1×10^{-2}	1×10^{-2}	9×10^{-3}	1×10^{-2}
Esters and aldehydes				
<i>Methylpropanoate</i>	2×10^{-2}	3×10^{-2}	1×10^{-2}	2×10^{-2}
Heterocyclic nitrogen compounds				
<i>Aniline</i>	1×10^{-1}	1×10^{-1}	$5 \times 10^{+1}$	4×10^{-2}
Oxyaromatics				
<i>Ethylphenol</i>	2×10^{-2}	2×10^{-2}	3×10^{-2}	2×10^{-2}

3.3 MODEL BASED ON COMBINATION OF THERMODYNAMIC INPUTS INTO PLS

Computed water/hydrocarbon distribution coefficients based on molar concentrations from thermodynamic calculation are given in Table 6. Calculations given in the first column of the table are for conditions of pH 7; [NaCl] of 10,800 mg·L⁻¹; volumetric ratio of water to water+oil of 50%; and a temperature of 25°C. Results in the second column are for an elevated temperature (75°C) and those in the third and fourth columns represent hydrogen ion concentrations of pH 3 and pH 10, respectively, with the other conditions held constant. The results of the calculation are not unexpected, with the solubility of the organic acids showing the most variation with pH — three orders of magnitude between pH 3 and pH 10 — making this the most important variable influencing overall organic solubility. The assumptions behind the results given in Table 6 were as follows:

- The systems comprised two liquid phases at equilibrium, one mainly hydrocarbon and one aqueous.
- The activities could be described by UNIFAC activity coefficients, modified by Pitzer-type coefficients dependent on ionic strength. Derivation of the UNIFAC activities (note the MAPLE 7 code given in Appendix B) also took temperature into account.

These results are independent of the experimental data collected on produced-water/brine systems at ORNL.

The computed equilibrium constants were then included in a PLS analysis. The reasons for this inclusion were

- to provide a framework for incorporation of field variables;
- to enable predictive analysis based on the measured fractionation of the crude oil, rather than a detailed chemical analysis demanded by a model of thermodynamic equilibrium; and
- to provide a framework to test the model using a cross-correlation method.

The analysis was run several times, for different groups of organics, because a restriction with the method is that the number of response variables must be less than the number of predictor variables. This methodology was examined during the PRESS analysis, which is discussed later.

The PLS algorithm was written in MAPLE 7 and is presented in Appendix C. The code provided in Appendix C shows both the model-building phase of the PLS algorithm and predictions for a group of alcohols. Comments on the MAPLE worksheet are printed in black type, the commands appear in red type, and the output is given in blue type. Details of the algorithm as applied to produced-water analysis have been published elsewhere [22]. However, they are summarized in the bullets below.

- Input for the statistical model development is provided for each of the compounds or classes of compounds and includes solubility as determined by thermodynamic calculation at a particular temperature, pH, salinity, and water-to-(water+oil) volumetric ratio. The predictor matrix comprises the physical conditions, and the response matrix comprises the solubilities. These values are scaled to be mean-

centered and variance-scaled. The calculation was run several times for different groups of compounds because the number of response variables cannot exceed the number of predictor variables.

- The NIPALS algorithm [37] followed in the worksheet involves simultaneous decomposition of the predictor and the response matrices to generate regression coefficients (“b” values), loading and score matrices, and a matrix of weights. If this has been done correctly, the matrices of residuals have very small values.
- The worksheet includes diagnostics that will be used later in the validation procedures. Sum-of-squares is calculated for each sample, as well as the derived factors.
- Plots give the relationship between each predictor and response variable, as well as cross-correlations.
- Once the model has been developed, it is now possible to derive responses for a new set of predictor variables. This is done in the last section of the worksheet. In the same way as done for the model generation, the input predictor variables (or conditions) are recalculated to be meancentered and variance scaled. The response matrix is then generated using the previously derived parameters (matrices and coefficients) from model development. The final step is to reconvert the rescaled variables back into physical measurements.

The use of a statistical approach also allowed testing of model predictions using the PRESS algorithm [38], or predictive residual sum of squares. PRESS is a cross-validation procedure used to test the stability of the PLS regression in which one set of input data is removed from the model and the predicted results are compared with the output of the complete model [39]. The parameters that are calculated during the cross-validation procedure include [40] SS_y , or the sum of squares of the residuals of the calculated values versus the values that have been set aside; PRESS, the accumulated SS_Y divided by the number of samples in the response matrix, s_y ; $XVAL = (PRESS)^{1/2} / s_y$; goodness of fit, $R^2Y = 1 - SS_F/SS_Y$; and predictability $Q^2 = 1 - PRESS/SS_Y$. Results of the diagnostics are given in Table 7 and Fig. 10.

The value of XVAL must be less than one for the model to have any predictive capabilities [41]. As can be seen in the table, XVAL is less than one for all of the compounds tested except for light gas and kerosene. The problems with light gas arose because the model did not include a vapor phase. (This could not be tested against the ORNL validation data, as is discussed later.) The difficulties with predicting kerosene compound solubility resulted from the fact that these solubilities are very low and variation would have been negligible across the range of predictor variables comprising the model. These effects are also demonstrated in the other criteria plotted in Fig. 10 for each compound, given in the same order as in Table 7. The goodness of fit is excellent for both the light gas and kerosene compounds, because the model did not include the factors that influence their partitioning into the aqueous phase. The model performed particularly well for ketones, midrange alcohols, and phenol. Aromatic compounds, mid range aliphatic compounds, and organic acids gave reasonable predictabilities. Quantitative diagnostics, such as are available with PLS, allow an independent scrutiny of the model, showing strengths and weaknesses in the approach.

Table 7. Uncertainties in computed PLS results

Chemical component	Uncertainties in PLS model (%)	XVAL
Ketones	±4	0.375
Aldehydes	±8	0.756
Ethanol	±2	0.865
C ₃ -C ₇ OH	±5	0.603
C ₈ -C ₁₂ OH	±6	0.681
VOC	±8	0.802
NAP	±8	0.793
PAH	±8	0.736
Oxyaromatics	±8	0.909
Light gas	±11	1.12
Light naptha	±9	0.849
Heavy naptha	±10	0.891
Kerosene	±11	1.03
Formic acid	±0.2	0.900
Acetic acid	±2	0.904
Propanoic acid	±3	0.844
Butanoic acid	±6	0.961
Pentanoic acid	±10	0.952
Octanoic acid	±17	0.835
Phenol	±3	0.483
Heterocyclic compounds	±14	0.796

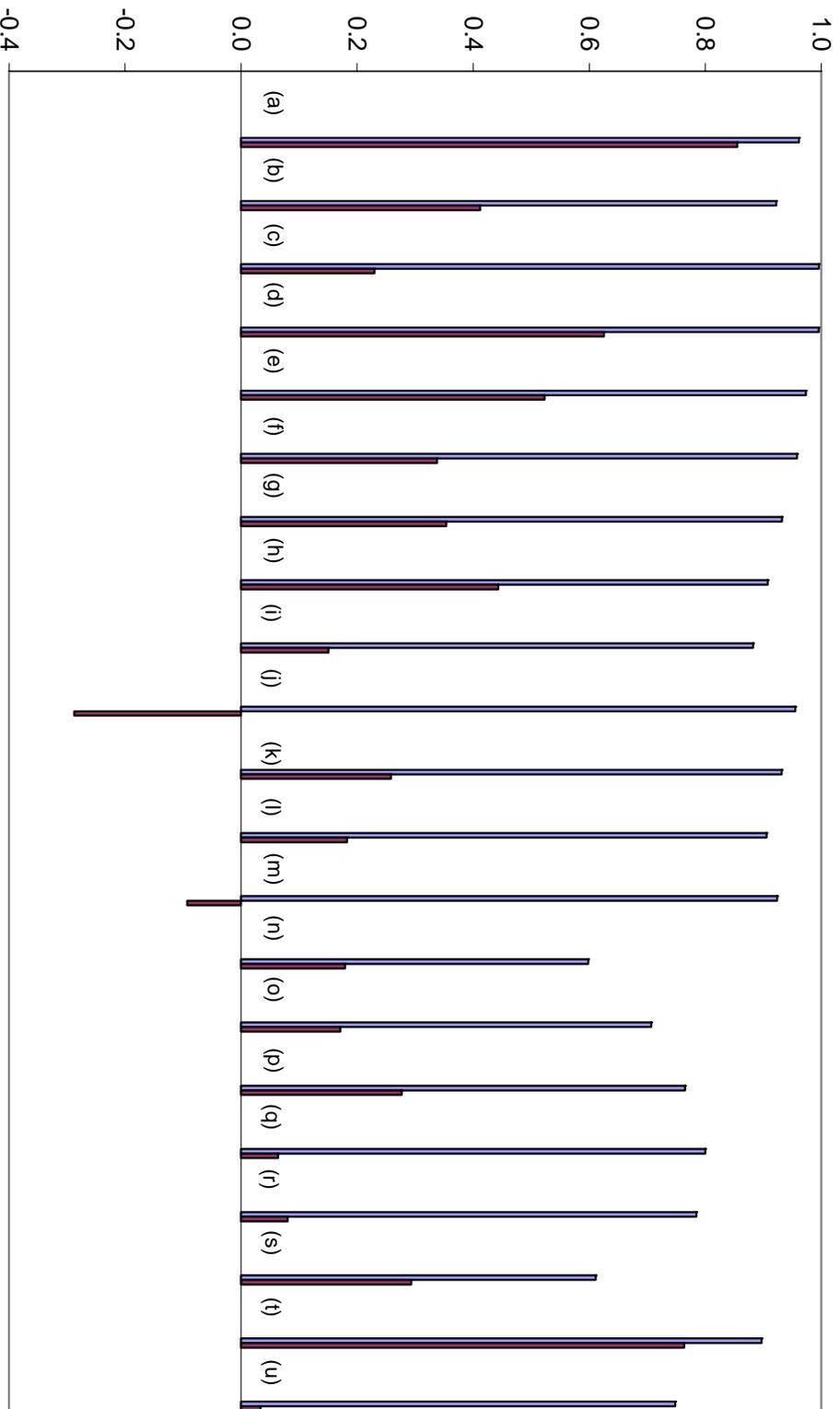


Fig. 10. PLS diagnostics. Each set of bars corresponds to one class of compound as follows (from left to right), with the letter designation appearing to the left of the respective bar: (a) = ketones; (b) = aldehydes; (c) = ethanol; (d) = C₃-C₇OH; (e) = C₈-C₁₂OH; (f) = VOC; (g) = NAP; (h) = PAH; (i) = oxyaromatics; (j) = light gas; (k) = light naphtha; (l) = heavy naphtha; (m) = kerosene; (n) = formic acid; (o) = acetic acid; (p) = propanoic acid; (q) = butanoic acid; (r) = pentanoic acid; (s) = octanoic acid; (t) = phenol; (u) = heterocyclic compounds.

3.4 COMPARISON WITH PRODUCED-WATER DATA

The predictive model was tested against the chemical systems analyzed at ORNL involving simulated produced-water brines and crude oil from the Gulf of Mexico. Calculations were done under the same conditions of pH, temperature, salinity, and phase ratio as established in the laboratory.

In order to compare results of the computation with experimental values, it was necessary to convert information from both sources to a common basis set. Because units of milligrams per liter were what was measured in the laboratory and could easily be derived from the computational results, these units were used for the fractions of aliphatic (light gas and light naptha), aromatics [volatile organic carbons (VOC) and oxyaromatics], polar compounds other than organic acids, and organic acids. However, the information on the Gulf of Mexico crudes had to be converted from a distillation fraction framework, as provided in Tables 2 and 8, to that described by chemical classification. In this way, a distribution coefficient was first calculated and then changed to an aqueous mass concentration based on the relative amounts of organic, or class of organic, in the oil. The results of the comparison are shown in Table 9 for each of the water characterization experiments and plotted in Figs. 11 and 12.

Table 8. Crude oil composition

Fraction	Distribution	Mars (vol %) (Gulf of Mexico)	Genesis (vol %) (Gulf of Mexico)
Aliphatic		45	51
C1-C4	Light ends	3	
C5-C6	Gasoline	7	
C6-C8	Light naptha	9	
C9, C10	Heavy naptha	8	
C10, C14	Kerosene	13	
C16, C18	Heavy gas oil	13	
C20, C28	Vacuum gas oil	25	
	Pitch	23	
Aromatic		40	39
	VOC	15	15
	Other aromatics	25	24
Resins		11	9
Asphaltenes		3	1
Waxes		1.5	0.9

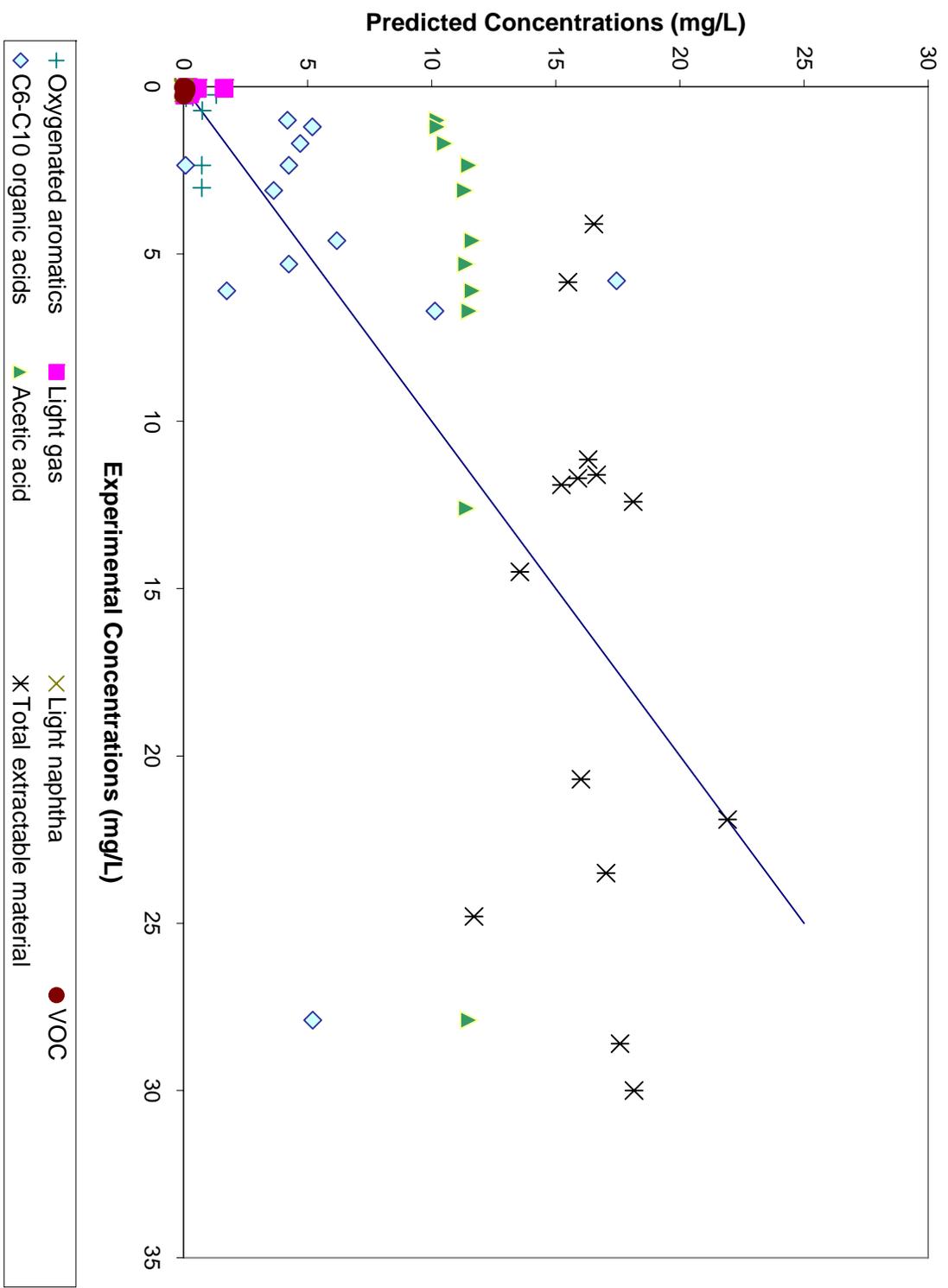


Fig. 11. Calculated vs experimental aqueous concentrations. Physical variables are pH7; temperature = 25°C; ambient pressure, NaCl concentration = 10800 ppm by weight; and water-to-(water + oil) volumetric ratio = 0.8.

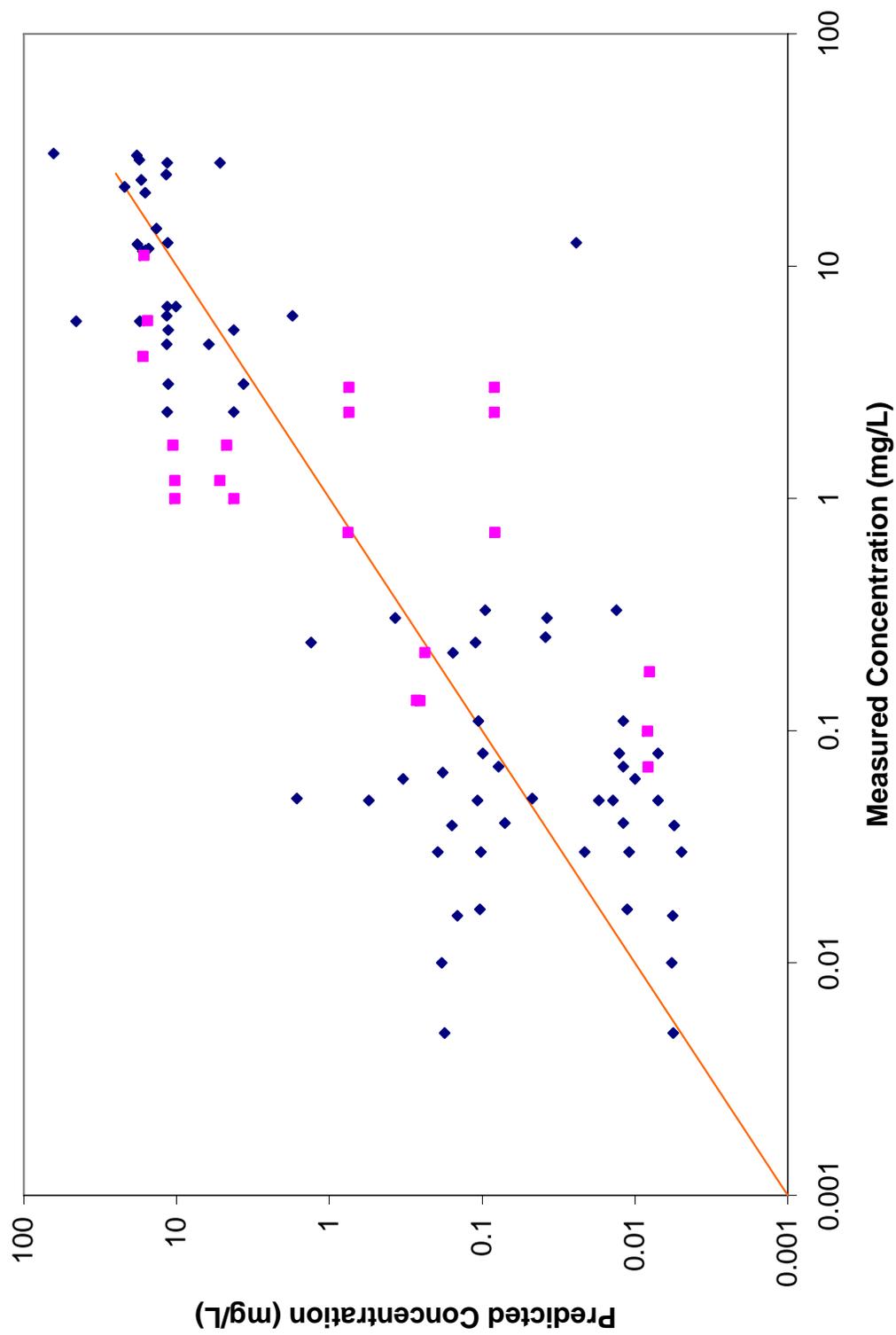


Fig. 12. Logarithmic plot of predicted vs measured concentrations in milligrams per liter. Diamonds are for data from the Mars well; squares are for data from the Genesis well (both in the Gulf of Mexico).

The uncertainties in the computed results are given in Table 8, with one standard deviation ranging from better than $\pm 1\%$ up to $\pm 17\%$ depending on the sensitivity of the results to key variables such as pH. Most of the organics, including key oil and grease compounds, could be predicted to within $\pm 10\%$, without imposing any assumptions regarding the composition of the crude oil. As seen in Fig. 11, the order-of-magnitude results of the calculation generally agree with the characterization results, showing water-soluble organic concentrations of a few milligrams per liter at most.

Figure 12, which presents information on a logarithmic scale, shows there is considerable scatter when prediction is plotted with experimental data. Some of this scatter has arisen because of the wide variety of frameworks used to glean information on the crude oil — for example, distillation fraction, pollutant analysis, and ORNL fractionation experiments — none of which gave much specific information about the chemical speciation of the polar component. Other scatter is a result of the noise in the produced-water experimental data. This scatter is as high as $\pm 60\%$ in some instances because of the multiple steps involved in the characterization procedure. For instance, estimates had to be made for the amount of loss that occurred during each solvent extraction step and each column separation, which increased the uncertainty in the final result. However, even under the adverse conditions in which the samples were collected, transported, and analyzed, it is obvious that the model and the results of the characterization tests show similar trends. The deviations at very low concentration will not cause NPDES permitting levels to be exceeded. Furthermore, in absolute terms, the concentrations of most organics are below levels of concern, except for the organics acids and other oxidized molecules. In addition, the type of oil that was investigated (e.g., the more aromatic Genesis crude or the more aliphatic Mars crude) did not appear to change the validity of the results.

Table 9. Experimental vs calculated distribution coefficients and aqueous concentrations

T (K)	pH	Mars conditions			Aq/Org mass distn coeff.		Aq conc (mg/L)	
		Salinity	Water Cut	WSO	Experiment	Predictions	Experiment	Predictions
298	7.00	108,000	0.80	Aliphatic ,LG	9.6E-06	9.4E-03	0.016	0.15
298	7.00	108,000	0.80	Aliphatic, LN	9.6E-06	3.6E-04	0.016	5.7E-03
298	7.00	108,000	0.80	Aromatic, VOC	2.7E-05	2.4E-03	0.017	0.011
298	7.00	108,000	0.80	Aromatic, oxy	2.7E-05	0.022	0.017	0.10
298	7.00	108,000	0.80	Polar, C6-C10	6.9E-04	0.54	2.4	4.2
298	7.00	108,000	0.80	Acetic	6.9E-04	19	2.4	12
298	7.00	108,000	0.80	TEM	4.8E-04		21	16
298	4.73	108,000	0.80	Aliphatic, LG	2.5E-06	0.010	0.039	0.16
298	4.73	108,000	0.80	Aliphatic, LN	2.5E-06	3.5E-04	0.039	5.5E-03
298	4.73	108,000	0.80	Aromatic, VOC	1.6E-05	2.4E-03	0.070	0.012
298	4.73	108,000	0.80	Aromatic, oxy	1.6E-05	0.016	0.070	0.078
298	4.73	108,000	0.80	Polar, C6-C10	9.2E-04	0.43	3.1	3.6
298	4.73	108,000	0.80	Acetic acid	9.2E-04	15	3.1	11
298	4.73	108,000	0.80	TEM	2.8E-04		12	15
298	5.98	108,000	0.80	Aliphatic, LG	3.2E-07	0.011	5.0E-03	0.18
298	5.98	108,000	0.80	Aliphatic, LN	3.2E-07	3.6E-04	5.0E-03	5.6E-03
298	5.98	108,000	0.80	Aromatic, VOC	8.8E-06	2.4E-03	0.040	0.012

Table 9. Experimental vs calculated distribution coefficients and aqueous concentrations

T (K)	Mars conditions				Aq/Org mass distn coeff.		Aq conc (mg/L)	
	pH	Salinity	Water Cut	WSO	Experiment	Predictions	Experiment	Predictions
298	5.98	108,000	0.80	Aromatic, oxy	8.8E-06	0.015	0.040	0.071
298	5.98	108,000	0.80	Polar, C6-C10	1.6E-03	0.54	5.3	4.2
298	5.98	108,000	0.80	Acetic acid	1.6E-03	16	5.3	11
298	5.98	108,000	0.80	TEM	2.7E-04		12	16
298	8.09	108,000	0.80	Aliphatic, LG	6.4E-07	0.012	0.010	0.18
298	8.09	108,000	0.80	Aliphatic, LN	6.4E-07	3.6E-04	0.010	5.7E-03
298	8.09	108,000	0.80	Aromatic, voc	2.4E-05	2.4E-03	0.11	0.012
298	8.09	108,000	0.80	Aromatic, oxy	2.4E-05	0.022	0.11	0.11
298	8.09	108,000	0.80	Polar	8.2E-03	0.75	28	5.2
298	8.09	108,000	0.80	acetic	8.2E-03	19	28	12
298	8.09	108,000	0.80	TEM	5.5E-04		24	17
298	9.03	108,000	0.80	TEM	6.7E-04		29	18
325	7.00	108,000	0.80	Aliphatic, LG	1.4E-05	0.010	0.22	0.16
325	7.00	108,000	0.80	Aliphatic, C6-C10	4.6E-05	4.5E-04	0.080	7.1E-03
325	7.00	108,000	0.80	Aromatic, VOC	1.8E-05	2.6E-03	0.080	1.3E-02
325	7.00	108,000	0.80	Aromatic, oxy	1.8E-05	0.021	0.080	0.099
325	7.00	108,000	0.80	Polar	2.0E-03	5.1	6.7	10.1
325	7.00	108,000	0.80	Acetic acid	2.0E-03	19	6.7	11.5
325	7.00	108,000	0.80	TEM	5.1E-04		22	22
350	7.00	108,000	0.80	Aromatic, voc	7.3E-05	2.7E-03	0.33	0.013
350	7.00	108,000	0.80	Aromatic, oxy	7.3E-05	0.020	0.33	0.095
350	7.00	108,000	0.80	Polar	3.7E-03	2.0E-03	13	0.024
350	7.00	108,000	0.80	Acetic acid	3.7E-03	16	13	11
350	7.00	108,000	0.80	TEM	5.8E-04		25	12
298	7.00	67,000	0.80	Aliphatic, LG	4.2E-06	0.012	0.066	0.18
298	7.00	67,000	0.80	Aliphatic, C6-C10	1.7E-05	3.2E-04	0.030	5.0E-03
298	7.00	67,000	0.80	Aromatic, VOC	6.6E-06	2.2E-03	0.030	0.011
298	7.00	67,000	0.80	Aromatic, oxy	6.6E-06	2.1E-02	0.030	0.10
298	7.00	67,000	0.80	Polar	1.4E-03	1.0	4.6	6.2
298	7.00	67,000	0.80	Acetic acid	1.4E-03	25	4.6	12
298	7.00	67,000	0.80	TEM	2.9E-04		12	18
298	7.00	190,000	0.80	Aliphatic, LG	1.6E-05	2.4E-03	0.25	0.038
298	7.00	190,000	0.80	Aliphatic, C6-C10	2.9E-05	4.5E-04	0.050	7.1E-03
298	7.00	190,000	0.80	Aromatic, VOC	1.1E-05	2.9E-03	0.050	0.014
298	7.00	190,000	0.80	Aromatic, oxy	1.1E-05	0.025	0.050	0.11
298	7.00	190,000	0.80	Polar	1.8E-03	0.17	6.1	1.7
298	7.00	190,000	0.80	Acetic acid	1.8E-03	25	6.1	12
298	7.00	190,000	0.80	TEM	3.4E-04		14	14
298	7.00	108,000	0.20	Aliphatic, LG	2.0E-07	6.5E-03	0.051	1.6
298	7.00	108,000	0.20	Aliphatic, LN	2.0E-07	1.9E-04	0.051	0.047
298	7.00	108,000	0.20	Aromatic, VOC	3.3E-06	1.4E-03	0.24	0.11
298	7.00	108,000	0.20	Aromatic, oxy	3.3E-06	0.017	0.24	1.3
298	7.00	108,000	0.20	TEM	4.4E-05		30	18
298	7.00	108,000	0.50	Aliphatic, LG	8.0E-07	8.9E-03	0.05	0.55
298	7.00	108,000	0.50	Aliphatic, LN	8.0E-07	2.7E-04	0.05	0.017
298	7.00	108,000	0.50	Aromatic, VOC	1.7E-05	1.9E-03	0.30	0.038
298	7.00	108,000	0.50	Aromatic, oxy	1.7E-05	0.019	0.30	0.37

Table 9. Experimental vs calculated distribution coefficients and aqueous concentrations

Mars conditions					Aq/Org mass distn coeff.		Aq conc (mg/L)	
T (K)	pH	Salinity	Water Cut	WSO	Experiment	Predictions	Experiment	Predictions
298	7.00	108,000	0.50	Polar	4.3E-04	0.56	5.8	17
298	7.00	108,000	0.50	Acetic acid	4.3E-04	15	5.8	45
298	7.00	108,000	0.50	TEM	1.8E-04		31	64
298	7.00	108,000	0.67	Aliphatic, LG	2.0E-06	0.011	0.062	0.33
298	7.00	108,000	0.67	Aliphatic, LN	2.0E-06	3.2E-04	0.062	0.010
298	7.00	108,000	0.67	Aromatic, VOC	3.4E-05	2.2E-03	0.030	0.021
298	7.00	108,000	0.67	Aromatic, oxy	3.4E-05	0.021	0.030	0.20
298	7.00	108,000	0.67	TEM	1.7E-05		12	17
298	7.12	108,000	0.80	Aliphatic, LG	4.8E-04	0.011	0.22	0.24
298	7.12	108,000	0.80	Aliphatic, C6-C10	2.5E-04	3.6E-04	0.18	8.0E-03
298	7.12	108,000	0.80	Aromatic, VOC	4.3E-04	2.4E-03	3.0	0.083
295	7.12	108,000	0.80	Aromatic, oxy	4.3E-04	0.028	3.0	0.74
298	7.12	108,000	0.80	Polar	4.8E-04	0.65	1.0	4.2
298	7.12	108,000	0.80	Acetic acid	4.8E-04	26	1.0	10
298	7.12	108,000	0.80	TEM	3.3E-04		5.8	16
298	8.52	108,000	0.80	Aliphatic, LG	3.0E-05	0.012	0.14	0.25
298	8.52	108,000	0.80	Aliphatic, C6-C10	9.7E-05	3.7E-04	0.070	8.2E-03
298	8.52	108,000	0.80	Aromatic, VOC	3.4E-04	2.4E-03	2.4	0.082
295	8.52	108,000	0.80	Aromatic, oxy	3.4E-04	0.022	2.4	0.74
298	8.52	108,000	0.80	Polar	8.1E-04	0.79	1.7	4.7
298	8.52	108,000	0.80	Acetic acid	8.1E-04	130	1.7	10
298	8.52	108,000	0.80	TEM	6.3E-04		11	16
298	9.48	108,000	0.80	Aliphatic, LG	3.0E-05	0.012	0.14	0.27
298	9.48	108,000	0.80	Aliphatic, C6-C10	1.4E-04	3.7E-04	0.10	8.3E-03
298	9.48	108,000	0.80	Aromatic, VOC	1.0E-04	0.023	0.72	0.082
295	9.48	108,000	0.80	Aromatic, oxy	1.0E-04	2.4E-03	0.72	0.76
298	9.48	108,000	0.80	Polar	5.7E-04	0.95	1.2	5.2
298	9.48	108,000	0.80	Acetic acid	5.7E-04	26	1.2	10
298	9.48	108,000	0.80	TEM	2.3E-04		4.1	16

^aWSO = water-soluble organics; LG = light gas; LN = light naphtha; TEM = total extractable material

4. CONCLUSIONS

This report provides details on the modeling of the organic loading of produced water arising from oil production in the Gulf of Mexico. The model is of interest because of environmental constraints on dumping the produced water back into the ocean. Both the nature and the amount of organics are of interest, because NPDES permitting strictly applies to oil and grease, or paraffinic compounds. However, both at ORNL and elsewhere, it has been determined experimentally that the largest fraction of organic compounds in produced water is polar in nature and is comprised mainly of organic acids. Acetic acid was observed in ORNL produced-water characterization experiments, along with C₃ and C₄ acids. The difficulty in quantifying oxidized organics in produced water arises because their concentrations are dependent not only on the conditions in the formation but also change upon processing, even from downhole to wellhead. However, these compounds have a large effect on the measured “oil and grease” content of

produced water and are also indicative of kerogen breakdown into petroleum and the age of the oil in the formation.

A model for organics in produced water was based on a thermodynamic depiction of liquid–liquid equilibrium based on UNIFAC activity coefficient data. Acid dissociation constants were modified based on the ionic strength of the produced-water brine. However, the effect on other classes of organics was found to be very small, because low concentrations were involved. Temperature effects, from ambient (~25°C) up to 75°C were also incorporated into the model. These variables were selected to allow comparison with produced-water characterizations carried out previously at ORNL. Other factors included the volumetric ratio of oil-to-water and pressure, although these parameters were found to have minor effects. (1) The distribution coefficient did not change with the ratio. (2) Only low-volatility organics were involved in the ORNL analyses, and this type was little affected by small changes in pressure.

Calculations were performed for organics that were chosen to represent the composition of crude oil — both the fractionation from distillation, for which information is readily available from the industry, and polar organics, for which analyses appear only in specialized studies. However, the methodology is established; thus, substitution of different organics should be easy to accomplish (should investigation of the behavior of a specific compound be required).

Use of a phenomenological model, such as the equilibrium model described above, is made more difficult because the system that is being described must be completely characterized. In addition, the activity coefficients are highly unreliable for some condensed-phase organics. In addition, the model currently does not involve nonequilibrium phenomena, i.e., fluid dynamics, turbulence, and the presence of a dispersed phase. The model was broadened from predictions based on equilibrium thermodynamics, by coupling solubilities with a PLS analysis that allowed incorporation of laboratory variables or field variables into the predictor matrix. This avenue was explored for the representative organics in crude oil, with the predictor variables also including temperature, salinity, and water-to-oil volumetric ratio.

The predictions of the model were compared with produced-water characterization data obtained in an earlier project at ORNL and previously reported in the literature. Agreement was reasonable for most of the fractions being examined. A cross-validation analysis was performed on the predictions, and the error in the model ranged from less than 1% to $\pm 17\%$.

It is hoped that the approach of coupling a thermodynamic analysis of the main components of crude oil with a PLS incorporation of laboratory and field variables will be useful to the petroleum industry in estimating the load of water-soluble organic compounds in the produced water associated with drilling in the Gulf of Mexico.

REFERENCES

1. U.S. Environmental Protection Agency (EPA), Compliance Assurance and Enforcement – Water – South Central, “Offshore and Oil & Gas NPDES Permits,” <http://www.epa.gov/region6/gen/w/offshore/>, Sept. 12, 2005.
2. EPA, *Method 1664, Revision A: N-Hexane Extractable Material (HEM; Oil and Grease) and Silica Gel Treated N-Hexane Extractable Material (SGT-HEM; Non-Polar Material) by Extraction and Gravimetry*, EPA-821-R-98-002, U.S. Environmental Protection Agency, 1999.
3. D. Th. Meijer, “Field-proven removal of dissolved and dispersed hydrocarbons at commercial scale from offshore produced water by the macro porous polymer-extraction technology,” presented at the Produced Water Workshop, Aberdeen, Scotland, March 26–27, 2003.
4. T. I. Røe Utvik and J. R. Hasle, “HSE horizons: Recent knowledge about produced water composition and the contribution from different chemicals to risk of harmful environmental effects,” *JPT Online*, http://www.spe.org/spe/jpt/jsp/jptmonthlysection/0.2440.1104_11040_1038813_1040052.00.html, December 2002.
5. T. Barth, “Quantitative determination of volatile carboxylic acids in formation waters by isotachopheresis,” *Analytical Chemistry* **59**, 2232–2237 (1987).
6. J. Brendehaug, S. Johnson, K. H. Bryne, A. L. Giose, T. H. Eide, and E. Aamot, “Toxicity testing and chemical characterization of produced water: A preliminary study,” pp. 245–256 in *Produced Water*, ed. J. P. Ray and F. R. Engelhart, Plenum Press, New York, 1992.
7. J. S. Brown, T. C. Sauer, Jr., M. J. Wade, and J. M. Neff, “Chemical and toxicological characterization of produced water Freon extracts,” in *Produced Water*, ed. J. P. Ray and F. R. Engelhart, Plenum Press, New York, 1992.
8. J. M. Neff, T. C. Sauer, and N. Maciolek, “Composition, fate, and effects of produced water discharges to nearshore marine waters,” pp. 371–385 in *Produced Water*, ed. J. P. Ray and F. R. Engelhart, Plenum Press, New York, 1992.
9. P. J. C. Tibbetts, I. T. Buchanan, L. J. Gawel, and R. Large, “A comprehensive determination of produced water composition,” pp. 97–112 in *Produced Water*, ed. J. P. Ray and F. R. Engelhart, Plenum Press, New York, 1992.
10. EPA, *Nonhalogenated Organics Using GC/FID, Method 8015B, Rev. 2*, U.S. Environmental Protection Agency, 1996.
11. D. T. Bostick, H. Luo, and B. Hindmarsh, *Characterization of Soluble Organics in Produced Water*, ORNL/TM-2001/78, Oak Ridge National Laboratory, Oak Ridge, Tenn., 2002.
12. J. McFarlane, D. T. Bostick, and H. Luo, “Characterization and modeling of produced water,” in *Proceedings of the Groundwater Protection Council Produced Water Conference*, Colorado Springs, Colo., Oct. 15–17, 2002, <http://www.gwpc.org/Meetings/PW2002/Papers-Abstracts.htm>.

13. T. I. Røe Utvik, "Chemical characterization of produced water from four offshore oil production platforms in the North Sea," *Chemosphere* **39**, 2593–2606 (1999).
14. I. Khan, P. Brimblecombe, and S. L. Clegg, "Solubilities of pyruvic acid and the lower carboxylic acids: Experimental determination of equilibrium vapour pressures above pure aqueous and salt solutions," *Journal of Atmospheric Chemistry* **22**, 285–302 (1995).
15. S.-T. Yang, S. A. White, and S.-T. Hsu, "Extraction of carboxylic acids with tertiary and quaternary amines: Effect of pH," *Industrial and Engineering Chemistry Research* **30**, 1335–1342 (1991).
16. J. McFarlane, "New approaches to produced water sampling, analysis and remediation at ORNL," in *Proceedings of the Ground Water Protection Council Annual Forum*, Charleston, S.C., Sept. 19–22, 2004.
17. A. Fredenslund and J. M. Sørensen, "Group contribution estimation methods," pp. 287–361 in *Models for Thermodynamic and Phase Equilibria Calculations*, ed. S. I. Sandler, Marcel Dekker, New York, 1993.
18. T. Magnussen, R. Rasmussen, and A. Fredenslund, "UNIFAC parameter table for prediction of liquid-liquid equilibria," *Industrial and Engineering Chemistry Process Design and Development* **20**, 331–339 (1981).
19. K. S. Pedersen, M. L. Michelsen, and A. O. Fredheim, "Phase equilibrium calculations for unprocessed well streams containing hydrate inhibitors," *Fluid Phase Equilibria* **126**, 13–28 (1996).
20. P. J. Brandvik, "Statistical simulation as an effective tool to evaluate and illustrate the advantage of experimental designs and response surface methods," *Chemometrics and Intelligent Laboratory Systems* **42**, 51–61 (1998).
21. P. J. Brandvik and P. S. Daling, "Optimisation of oil spill dispersant composition by mixture design and response surface methods," *Chemometrics and Intelligent Laboratory Systems* **42**, 63–72 (1998).
22. J. McFarlane, "Application of chemometrics to modeling produced water contamination," *Separations Science and Technology* **40** (1–3), 593–609 (2005).
23. T. H. Andres, "Sampling methods and sensitivity analysis for large parameter sets," *Journal of Statistical Computation and Simulation* **57**, 77–110 (1997).
24. J. H. Gary and G. E. Handwerk, *Petroleum Refining*, vol. 5 of *Chemical Processing and Engineering*, ed. L. F. Albright, R. N. Maddox, and J. J. McKetta, Marcel Dekker, Inc., New York, 1975, pp. 17–19.
25. P. Barbieri, G. Adami, S. Piselli, F. Gemiti, and E. Reisenhofer, "A three-way principal factor analysis for assessing the time variability of freshwaters related to a municipal water supply," *Chemometrics and Intelligent Laboratory Systems* **62**, 89–100 (2002).
26. Z. Wang, B. P. Hollebone, M. Fingas, B. Fieldhouse, L. Sigouin, M. Landriault, P. Smith, J. Noonan, and P. Thouin, *Characteristics of Spilled Oils, Fuel and Petroleum Products, 1: Composition and Properties of Selected Oils*, EPA/600/R-03/072, U.S. Environmental Protection Agency, July 2003.
27. Environment Canada, Environmental Technology Centre (ETC), Spills Technology Databases, Oil Properties Database, <http://www.etccentre.org/>.

28. H. T. McAdams, R. W. Crawford, and G. R. Hadder, *A Vector Approach to Regression Analysis and Its Application to Heavy-Duty Diesel Emissions*, ORNL/TM-2000/5, Oak Ridge National Laboratory, Oak Ridge, Tenn., November 2000.
29. S. Budavari, ed., *The Merck Index*, 12th ed., Merck & Co., Whitehouse Station, N.J., 1996.
30. R. C. Reid, J. M. Prausnitz, and B. E. Poling, eds., *The Properties of Gases and Liquids*, 4th ed., McGraw-Hill, New York, 1987.
31. J. A. Dean, ed., *Lange's Handbook of Chemistry*, 15th ed., McGraw-Hill, New York, 1999.
32. Y. K. Kharaka and J. S. Hanor, "Deep fluids in the continents, I: Sedimentary basins," pp. 499–540 in *Surface and Ground Water, Weathering, and Soils*, vol. 5 of *Treatise on Geochemistry*, ed. J. I. Drever, Elsevier, 2003
33. M. B. Smith and J. March, *March's Advanced Organic Chemistry*, 5th ed., Wiley-Interscience, New York, 2001, pp. 329–331.
34. Y. K. Kharaka, W. W. Carothers, and R. J. Rosenbauer, "Thermal decarboxylation of acetic acid: Implications for origin of natural gas," *Geochimica and Cosmochimica Acta* **47**, 397–402 (1983).
35. D. Mackay, W. Y. Shiu, and K.-C. Ma, *Physical-Chemical Properties and Environmental Fate Handbook*, Chapman & Hall / CRCnet Base, Boca Raton, Fla., 1999.
36. R. P. W. M. Jacobs, R. O. H. Grant, J. Kwant, J. M. Marquenie, and E. Mentzer, "The composition of produced water from Shell-operated oil and gas production in the North Sea," in *Produced Water*, ed. J. P. Ray and F. R. Engelhart, Plenum Press, New York, 1992.
37. P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Analytica Chimica Acta* **185**, 1–17 (1986).
38. S. Wold, A. Ruhe, H. Wold, and W. J. Dunn III. "The collinearity problem in linear regression: The partial least squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing* **5** (3), 735–743 (1984).
39. L. E. Wangen, M. V. Phillips, and L. F. Walker, *Use of Multivariate Calibration for Plutonium Quantification by the Pu(III) Spectrophotometric Method*, LA-11297-MS, Los Alamos National Laboratory, Los Alamos, N.M., May 1988.
40. S. Wold, M. Sjöström, and L. Erikson, "Partial least squares projections to latent structures (PLS) in chemistry," pp. 2006–2121 in vol. 3 M-P of *Encyclopedia of Computational Chemistry*, ed.-in-chief P. von Ragné Schleyer, John Wiley & Sons Ltd., Chichester, UK, 1998.
41. P. Geladi and B. R. Kowalski, "An example of 2-block predictive partial least-squares regression with simulated data," *Analytica Chimica Acta* **185**, 19–32 (1986).

APPENDIX A

MAPLE 7 Code for Thermodynamic Equilibrium Calculations

LIQUID-LIQUID PARTIONING

WITH ACID-BASE EQUILIBRIA AND UNIFAC OR NRTL

MULTIPLE PH

Joanna McFarlane
Oak Ridge National Laboratory
April 2, 2003

The example used in this system involves 2 phases (hydrocarbon-water), where the solute in the aqueous phase has a acid dissociation constant. Either the UNIFAC or NRTL model for activity coefficients can be used.

This model will allow one to model temperature dependence (through the activity coefficients), pH dependence (through dissociation in the aqueous phase), dependence on salinity (through a Debye-Huckel model for dependence of ionic activity coefficient on ionic strength), and water cut (or volume fraction of aqueous to organic phases).

This model will not allow one to model dependence on pressure, as the gas phase is not included in the calculation.

This program allows one to enter a number of pH values. Output from the calculation goes into the whichever directory is specified, e.g., C:\\Maple Worksheets\\data files\\output_multipH.txt.

```
> restart:with(linalg):unprotect(gamma):
```

```
Warning, the protected names norm and trace have been redefined and unprotected
```

```
*****
```

The mathematical model of the phase envelope calculation.

The section starts off by defining the total material balance (TMB), where L is the oil phase, and W is the water phase, and the component material balances (CMB), one for each component:

```
> TMB:=F=L+W: TMB; CMB[i]:=F*z[i]=L*x[i]+W*u[i]+W*d[i]: CMB[i];
```

where

```
> 'F,L,W,z[i],x[i],u[i],d[i]';
```

are respectively, the feed flow, the oil liquid product flow, the aqueous flow, and the mole fraction of component i in the feed, oil phase, and aqueous phase (undissociated and dissociated). The mole fractions sum over

the component index to unity. However, mole fractions do not sum to unity of their own accord and it is necessary to include the mole fraction summation equations as part of the model.

```
> Sumx:=sum(x[i],i=1..c)=1:
Sumz:=sum(z[i],i=1..c)=1:Sumw:=sum(u[i]+d[i],i=1..c)=1: Sumx,Sumz,Sumw;
```

The model is completed by assuming that the L and W phases are in equilibrium with each other.

```
> EQMLW[i]:=u[i]=Kw[i]*x[i]: EQMLW[i];
EQM_acid[i]:=d[i]=u[i]*Ka[i]/H*Gamma[i]:EQM_acid[i];
```

The correction for salinity correction enters into the calculation in Gamma[i] - see below.

```
> EQM_gamma[i]:=Gamma[i]=gamma[H]*gamma[A][i]/gamma[HA][i]:EQM_gamma[i];
```

We can also define alpha, where alpha is the fraction of acid dissociated. It can be shown that alpha is dependent on the acid dissociation constant of the polar molecule, and on the hydrogen ion concentration in solution (from the pH).

```
> alpha[i]:=Ka[i]/(H+Ka[i]);
```

There are $3c+5$ variables in this set of equations:

- The flow rates of the four streams (feed, oil, water)
- c mole fractions in each stream
- The system temperature
- The system pressure

There are only $2c+4$ independent equations, however:

- c material balances ($F=L+W$ for each component)
- c equilibrium equations ($w_i=K_{iw}*x_i$ for each component)
- 3 mole fraction summation equations ($sum x, sum z, sum w$)

This means that $c+2$ variables must be specified before we can carry out any calculations (**degrees of freedom=variables-independent equations**). (also $F=C-P+2$ where #components=#distinct chemical species-#equilibrium equations) A possible (and common) specification includes the feed flow and $c-1$ mole fractions in the feed (the last one is calculated from the mole fraction summation equation for the feed), the temperature and pressure. The calculated quantities include the flow rates, and the mole fractions of the two product streams.

The Rachford - Rice Equation

Our concern here is with the development of a method that permits us to solve simple flash problems. The classical approach is to combine the equations in such a way that flash problems can be solved by a root finding method such as Newton's method. We define the vapor fraction and phase fraction of the aqueous phase respectively:

```
> t0:=W=beta*F: beta=solve(t0,beta);
```

which allows us to solve the total material balance for the organic fraction

```
> t1:=collect(TMB/F,L): t1:=L=solve(t1,L): t1:=subs(t0,t1): t1;
```

The component material balances can be expressed in terms of the aqueous fraction as follows (*collecting terms in F*):

```
> e1:=collect(subs({t1,t0},CMB[i]/F),{F,beta}):
```

The aqueous phase mole fractions can be eliminated using the equilibrium equations and the result rearranged to give the organic phase mole fractions as (*I am not sure why normal was used. Why couldn't mapping have been used instead?*)

```
> e2:=subs(EQM_acid[i],e1): e2a:=subs(EQMLW[i],e2): e3:=solve(e2a,x[i]):
e4:=x[i]=normal(e3):
```

The aqueous phase mole fractions may now be obtained by substituting the above result into the equilibrium equations.

```
> e5:=subs(e4,EQMLW[i]):
```

Rewriting the mole fraction equations

```
> c:='c':RR1:=combine(Sumw-Sumx):RR1;RR2:=subs(EQM_acid[i],RR1):
```

The Rachford Rice Equation is

```
> RReqn:=simplify(subs({e4,e5},RR2)):RReqn;
```

Testing for singularities in the Rachford Rice Equations. One will get a singularity when the denominator in the sum above=0. There will be one singularity for each component.

```
> t2:=denom(op(1,lhs(RReqn)))=0:
> t3:=S[i]=solve(t2,beta):
> RRsing:=unapply(rhs(t3),Kw[i]):
> RRS:=simplify(subs(Kw[i]=solve(t3,Kw[i]),RReqn)):RRS;
```

$$F = L + W$$

$$F z_i = L x_i + W u_i + W d_i$$

$$F, L, W, z_i, x_i, u_i, d_i$$

$$\sum_{i=1}^c x_i = 1, \sum_{i=1}^c z_i = 1, \sum_{i=1}^c (u_i + d_i) = 1$$

$$u_i = Kw_i x_i$$

$$d_i = \frac{u_i K a_i \Gamma_i}{H}$$

$$\Gamma_i = \frac{\gamma_H \gamma_{A_i}}{\gamma_{HA_i}}$$

$$\alpha_i := \frac{K a_i}{H + K a_i}$$

$$\beta = \frac{W}{F}$$

$$L = F - \beta F$$

$$\sum_{i=1}^c (-x_i + u_i + d_i) = 0$$

$$\sum_{i=1}^c \frac{z_i (-H + K w_i H + K w_i K a_i \Gamma_i)}{-\beta H + \beta K w_i H + \beta K w_i K a_i \Gamma_i + H} = 0$$

$$\sum_{i=1}^c \frac{z_i}{\beta - S_i} = 0$$

This procedure calculates the mole fraction in each phase given the phase fraction, the equilibrium constants, the feed concentration, the pH, and the components. The procedure also normalizes the computed mole fractions.

```
> molefraction:=proc(c,z,Kw,Ka,beta,H,components)
>   local i,W,X; global x,u,d,w,x_old,w_old;
   i:='i':x:='x':u:='u':
for i from 1 to c do
   x_old[i]:=x[i]:w_old[i]:=w[i]:
od:
i:='i':
for i from 1 to c do
   x[i]:=z[i]*H/(beta*H*(Kw[i]-1)+beta*Kw[i]*Ka[i]+H);
   u[i]:=z[i]*H*Kw[i]/(beta*H*(Kw[i]-1)+beta*Kw[i]*Ka[i]+H);
   d[i]:=u[i]*Ka[i]/H;
od;
i:='i':W:=sum(u[i]+d[i],i=1..c);X:=sum(x[i],i=1..c);
i:='i':
for i from 1 to c do
   #print(components[i]);
   x[i]:=x[i]/X; #print("Hydrocarbon mole fraction=",x[i]);
   u[i]:=u[i]/W; #print("Undissociated Aqueous mole fraction=",u[i]);
>   d[i]:=d[i]/W; #print("Dissociated Aqueous mole fraction=",d[i]);
   w[i]:=u[i]+d[i]; #print("Total Aqueous mole fraction=",w[i]);
od:
```

```
> end proc:save molefraction,"molefraction.mws":
*****
```

This procedure calculates NRTL parameters for each component in the system. This is where the temperature dependence would become important.

```
> NRTL_setup:=proc(c,Temperature)
  local R,tauT,NRTLparam_alpha,i,j; global NRTLparam_tau,G;
  R:=0.08314;
  tauT:=array(1..c,1..c,[[0,5000,-81],[1561,0,276],[2741,288,0]]);
  NRTLparam_tau:=array(1..c,1..c):
  for i from 1 to c do
    for j from 1 to c do
      NRTLparam_tau[i,j]:=tauT[i,j]/Temperature:od:
    od:
  print(NRTLparam_tau);

  NRTLparam_alpha:=array(1..c,1..c,[[0,0.48,0.15],[.48,0,1.2],[0.15,1.2,0
  ]]):
  G:=array(1..c,1..c):
  > for i from 1 to c do
    for j from 1 to c do G[i,j]:=exp(-
  NRTLparam_tau[i,j]*NRTLparam_alpha[i,j]):od:
    od:
  print(G);
end proc:save NRTL_setup,"NRTL_setup.mws":
*****
```

This procedure calculates activity coefficients in the aqueous and hydrocarbon phases, given the mole fractions of each of the components. The output is the equilibrium constant between the aqueous and hydrocarbon phases.

```
> NRTL_activity_coefficients:=proc(c,x,u,G,NRTLparam_tau)
  local i,j,k,sum1,sum2,sum3,sum4,lngw,lngx,p,q;global Kw;
```

Calculation of ln activity coefficients for aqueous phase.

```
sum1:=0:sum2:=0:sum3:=0:sum4:=0:i:='i':j:='j':k:='k':
sum1:= sum(NRTLparam_tau[j,i]*G[j,i]*u[j], 'j'=1..c);
sum2:= sum(G[j,i]*u[j], 'j'=1..c);
sum3:= sum(u[k]*NRTLparam_tau[k,j]*G[k,j], 'k'=1..c);
sum4:= sum(u[k]*G[k,j], 'k'=1..c);
for i from 1 to c do
  lngw[i]:=eval(sum1/sum2+sum((u[j]*G[i,j]/sum4)*(NRTLparam_tau[i,j]-
sum3/sum4), 'j'=1..c));
  p[i]:=exp(lngw[i]);
od:
> print("Aqueous activity coefficients ",p);
```

Calculation of ln activity coefficients for hydrocarbon phase

```
> sum1:=0:sum2:=0:sum3:=0:sum4:=0:i:='i':j:='j':k:='k':
sum1:= sum(NRTLparam_tau[j,i]*G[j,i]*x[j], 'j'=1..c);
sum2:= sum(G[j,i]*x[j], 'j'=1..c);
sum3:= sum(x[k]*NRTLparam_tau[k,j]*G[k,j], 'k'=1..c);
```

```

sum4:= sum(x[k]*G[k,j],'k'=1..c);
for i from 1 to c do
  lngx[i]:=eval(sum1/sum2+sum((x[j]*G[i,j]/sum4)*(NRTLparam_tau[i,j]-
sum3/sum4),'j'=1..c));
  q[i]:=exp(lngx[i]);
od:
print("Hcarbon activity coefficients ",q);

```

Calculation of Equilibrium Constants for the H-carbon/water system, where $K_w[i]$ is defined as $w[i]/x[i]$.

```

> i:='i':for i from 1 to c do Kw[i]:=exp(lngx[i]-lngw[i]);od;
> print("Distribution coefficients",Kw);

end proc:save
NRTL_activity_coefficients,"NRTL_activity_coefficients.mws":
*****
> read "UNIFAC_LL_activity_coefficients.mws":
*****

```

Example1: 5 components - 2 phases.

Water, acetic acid, phenol, n-octane and nhexane.

Must solve for beta (hydrocarbon-water split). Very tricky to get a reasonable first guess so that the Rachford Rice equation will converge.

Will attempt the procedure recommended by Walas - p.371.

n-octane and water will be the "solvents", and acetic acid will be the "solute". Have to estimate initial K_w , <1 for hydrocarbons, ~ 1 for species with solubility in both phases, and >1 for aqueous components. $K_w[i]=u[i]/x[i]$ - note that in a fully dissociated system, $K_w[i]$ will not give the concentration in the aqueous phase directly. Hence, the distribution coefficients given at the last printout are not true K_w , but ratios of mole fraction in aqueous (dissociated+undissociated) versus organic phases.

- SETUP

Need components of feed, the feed composition, the temperature (K), pKa for the aqueous species (the pKa for the organic phase is chosen to be arbitrarily large), pH, the salinity in terms of chloride ion concentration (mg/L), the charge on the dissociated acid anion and its size in angstroms.

```

>
components:=['hexane','noctane','decane','icosane','benzene','naphthalene',
'phenanthrene','aceticacid','phenol','ethyldecanoate','water'];z:=vector(
[0.01,0.0095,0.0889,0.0028,0.0102,.0233,.0032,.0501,.0011,.0007,0.8]);c:=nops(
components);UNIFAC_flag:=1;Temperature:=300.;Kw:=vector([0.0555,0.0555,0.0555,
.0555,.0555,.0555,.0555,1.,1.,1.,18.]);multipH:=[2,4,5,6,7,8,10];pH_number:=nops(
multipH);pH:=multipH[1];phase_split:=Vector(1..pH_number):salinity:=108000.;
subgroup:=matrix([[1,1,1,2],[1,1,2,4],[2,1,1,2],[2,1,2,6],[3,1,1,2],[3,1,2,8],[4,1,1,2],[4,1,2,18],[5,3,9,6],[6,3,9,8],[6,3,10,1],[7,3,9,10],[7,3,10,2],[8,1,1,1],[8,13,23,1],[9,3,9,5],[9,9,18,1],[10,1,1,2],[10,1,2,8],[10,14,26,1],[11,8,17,1]]);

```

```

pKa:=( [20.,20.,20.,20.,20.,20.,20.,4.5,10.,10.,14.] );H:=10^(-
pH);Anion_charge:=( [1,1,1,1,1,1,1,1,1,1,1] );Anion_size:=( [6,6,6,6,6,6,6
,6,6,6,3.5] );H_carbon_molefraction:=Matrix(1..c,1..pH_number):Aqueous_m
olefraction:=Matrix(1..c,1..pH_number):
Diss_molefraction:=Matrix(1..c,1..pH_number):Undiss_molefraction:=Matri
x(1..c,1..pH_number):Water_Hcarbon_ratio:=Matrix(1..c,1..pH_number):

```

```

components := [ hexane , noctane , decane , icosane , benzene , naphthalene ,
phenanthrene , aceticacid , phenol , ethyldecanoate , water ]

```

```

z := [ .01 , .0095 , .0889 , .0028 , .0102 , .0233 , .0032 , .0501 , .0011 , .0007 , .8 ]

```

```

c := 11

```

```

UNIFAC_flag := 1

```

```

Temperature := 300.

```

```

Kw := [ .0555 , .0555 , .0555 , .0555 , .0555 , .0555 , .0555 , 1., 1., 1., 18. ]

```

```

multipH := [ 2, 4, 5, 6, 7, 8, 10 ]

```

```

pH_number := 7

```

```

pH := 2

```

```

salinity := 108000.

```

$$subgroup := \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 4 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 2 & 6 \\ 3 & 1 & 1 & 2 \\ 3 & 1 & 2 & 8 \\ 4 & 1 & 1 & 2 \\ 4 & 1 & 2 & 18 \\ 5 & 3 & 9 & 6 \\ 6 & 3 & 9 & 8 \\ 6 & 3 & 10 & 1 \\ 7 & 3 & 9 & 10 \\ 7 & 3 & 10 & 2 \\ 8 & 1 & 1 & 1 \\ 8 & 13 & 23 & 1 \\ 9 & 3 & 9 & 5 \\ 9 & 9 & 18 & 1 \\ 10 & 1 & 1 & 2 \\ 10 & 1 & 2 & 8 \\ 10 & 14 & 26 & 1 \\ 11 & 8 & 17 & 1 \end{bmatrix}$$

$$pKa := [20., 20., 20., 20., 20., 20., 20., 4.5, 10., 10., 14.]$$

$$H := \frac{1}{100}$$

$$Anion_charge := [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$$

$$Anion_size := [6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 3.5]$$

We will be using either the Non-Random Two Liquid or UNIFAC model to calculate activity coefficients

```
> if UNIFAC_flag<>1 then NRTL_setup(c, Temperature); fi;
```

Calculation of ratio of ionic activity coefficients. The activity coefficient of the hydrogen ion is assumed to be controlled by the dominant counter ion, the Cl⁻. For simplicity, the activity coefficient of the anion species is also assumed to be the same (see Bowden, Clegg and Brimblecomb, Chemosphere, 32(2), 405-420, 1996).

To incorporate the effect of salinity into the calculation, used gamma to modify Ka. Although the gamma is explicitly written into the Rachford-Rice equation, it is not explicitly included everytime Ka is used to determine the mole fraction. And so, it is easier to "adjust" the Ka in this section.

Even though a Gamma is calculated for the Hcarbon species such as octane, it is not important in the calculations, because the pKa is so large.

```
> Salt_molality:=salinity/35450.;
sqrt_IS:=sqrt(Salt_molality);
gamma_H:=exp(-
((1.167*sqrt_IS)/(1+1.167*0.3282*sqrt_IS)+0.3*Salt_molality));
gamma_A:=gamma_H;
i:='i':
for i from 1 to c do
  gamma_HA[i]:=exp(2*Salt_molality*0.122);
  Gamma[i]:=gamma_A*gamma_H/gamma_HA[i];
  Ka[i]:=10^(-pKa[i])*Gamma[i];
  Gamma[i]:=1;
od;
```

Salt_molality := 3.046544429

gamma_A := .1182728572

*gamma_HA*₁ := 2.102983058

Γ_1 := .006651726792

*Ka*₁ := .6651726792 10⁻²²

Γ_1 := 1

*gamma_HA*₂ := 2.102983058

Γ_2 := .006651726792

*Ka*₂ := .6651726792 10⁻²²

Γ_2 := 1

*gamma_HA*₃ := 2.102983058

Γ_3 := .006651726792

*Ka*₃ := .6651726792 10⁻²²

Γ_3 := 1

*gamma_HA*₄ := 2.102983058

Γ_4 := .006651726792

*Ka*₄ := .6651726792 10⁻²²

Γ_4 := 1

*gamma_HA*₅ := 2.102983058

$\Gamma_5 := .006651726792$
 $Ka_5 := .6651726792 \cdot 10^{-22}$
 $\Gamma_5 := 1$
 $gamma_HA_6 := 2.102983058$
 $\Gamma_6 := .006651726792$
 $Ka_6 := .6651726792 \cdot 10^{-22}$
 $\Gamma_6 := 1$
 $gamma_HA_7 := 2.102983058$
 $\Gamma_7 := .006651726792$
 $Ka_7 := .6651726792 \cdot 10^{-22}$
 $\Gamma_7 := 1$
 $gamma_HA_8 := 2.102983058$
 $\Gamma_8 := .006651726792$
 $Ka_8 := .2103460703 \cdot 10^{-6}$
 $\Gamma_8 := 1$
 $gamma_HA_9 := 2.102983058$
 $\Gamma_9 := .006651726792$
 $Ka_9 := .6651726792 \cdot 10^{-12}$
 $\Gamma_9 := 1$
 $gamma_HA_{10} := 2.102983058$
 $\Gamma_{10} := .006651726792$
 $Ka_{10} := .6651726792 \cdot 10^{-12}$
 $\Gamma_{10} := 1$
 $gamma_HA_{11} := 2.102983058$
 $\Gamma_{11} := .006651726792$

$$Ka_{11} := .6651726792 \cdot 10^{-16}$$

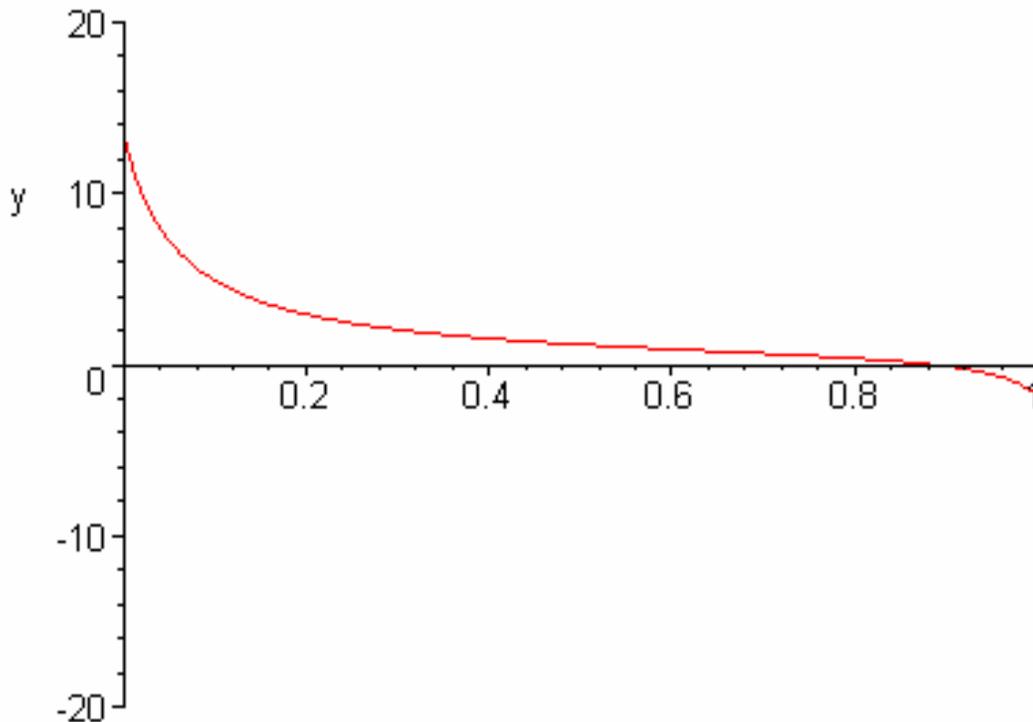
$$\Gamma_{11} := 1$$

The first guess

A plot is used because otherwise the blind application of the RR algorithm can lead to the "wrong" solution.

```
> i:='i':RReqn;
fRR:=unapply(lhs(RReqn),beta):numer(fRR(beta)):plot(fRR,0..1,y=-
20..20,discont=true);
```

$$\begin{aligned}
 & -0.001396915500 \frac{1}{-0.009445000000 \beta + .010000000000} \\
 & + \frac{.1053833812 \cdot 10^{-7}}{.210346 \cdot 10^{-6} \beta + .010000000000} + \frac{.1197310822 \cdot 10^{-14}}{.1 \cdot 10^{-11} \beta + .010000000000} \\
 & + \frac{.1360000000}{.1700000000 \beta + .010000000000} = 0
 \end{aligned}$$



```
➤ RRSoln:=fsolve({RReqn},{beta},0..1);assign(RRSoln);
```

```
➤
```

$$RRSoln := \{ \beta = .8843856660 \}$$

Evaluate the mole fractions in each phase

```
➤ molefraction(c,z,Kw,Ka,beta,H,components);
➤
```

.8982400440

```
> check:=1:
count:=1:
H_carbon_molefraction:=[]:Aqueous_molefraction:=[]:Diss_molefraction:=[]
Undiss_molefraction:=[]:Water_Hcarbon_ratio:=[]:
```

If this was an ideal system - the calculation would end. But will go back and recalculate activity coefficients, solve Rachford-Rice equation, and mole fractions. When the change in mole fraction from one iteration to the next is within a certain number, the loop ends. There is also a loop overrun check. Either the NRTL equation or UNIFAC can be used..

Iterations to convergence

```
> for pH_index from 1 to pH_number do
pH:=multipH[pH_index];H:=10^(-pH):
while check>1E-08 do

if UNIFAC_flag<>1 then print("Non-Random Two
Liquid");NRTL_activity_coefficients(c,x,u,G,NRTLparam_tau);
else
print("UNIFAC");
print("Hydrocarbon activity
coefficients");UNIFAC_LL_activity_coefficients(x,lngx):eval(lngx):
print("Aqueous activity
coefficients");UNIFAC_LL_activity_coefficients(u,lngw):eval(lngw):
Distribution coefficients for the hydrocarbon/water system, where Kw[i] is defined
as w[i]/x[i].
> i:='i':for i from 1 to c do Kw[i]:=exp(lngx[i]-lngw[i]);od;
> #print("Distribution coefficients",Kw); fi;

i:='i':beta:='beta':RRsoln:=fsolve({RReqn},{beta},0..1):assign(RRsoln):
molefraction(c,z,Kw,Ka,beta,H,components):
phase_split[pH_index]:=beta:print("beta is ",beta);
count:=count+1:
i:='i':check:=0:
if count<20 then for i from 1 to c do check:=check+abs(x_old[i]-
x[i])+abs(w_old[i]-w[i]):od:end if:

end do:

i:='i':for i from 1 to c do
H_carbon_molefraction:=[op(H_carbon_molefraction),x[i]]:
Aqueous_molefraction:=[op(Aqueous_molefraction),w[i]]:
Diss_molefraction:=[op(Diss_molefraction),d[i]]:
Undiss_molefraction:=[op(Undiss_molefraction),u[i]]:
Water_Hcarbon_ratio:=[op(Water_Hcarbon_ratio),(u[i]+d[i])/x[i]]:
od:

check:=1:count:=1:
print("*****
*****");end do;
```

pH := 2

"UNIFAC"

"Hydrocarbon activity coefficients"

"UNIFAC activity coefficients " , table([*icosane* = .6436223621 ,
decane = 1.000746839 , *phenanthrene* = 1.426481518 ,
aceticacid = 1.634126211 , *hexane* = .9829980661 , *benzene* = .9236822478 ,
phenol = 8.160924311 , *water* = 401.5303580 , *ethyldecanoate* = .5636498927 ,
noctane = 1.017458743 ,
napthalene = 1.195705052
])(*components*)

"Aqueous activity coefficients"

"UNIFAC activity coefficients " , table([*icosane* = .2564437971 10¹¹ ,
decane = 645556.0839 , *phenanthrene* = 211807.6944 ,
aceticacid = 2.028652836 , *hexane* = 7688.858662 , *benzene* = 1002.212253 ,
phenol = 21.08493731 , *water* = 1.005776656 , *ethyldecanoate* = 737228.5488 ,
noctane = 72273.03509 ,
napthalene = 15176.80698
])(*components*)

"beta is " , .8411842939

i := *i*

check := 1

count := 1

***** \
*****"

pH := 4

"UNIFAC"

"Hydrocarbon activity coefficients"

"UNIFAC activity coefficients " , table([*icosane* = .6436647972 ,
decane = 1.000757118 , *phenanthrene* = 1.426433292 ,
aceticacid = 1.634132222 , *hexane* = .9829862674 , *benzene* = .9236447488 ,
phenol = 8.166376536 , *water* = 401.6269754 , *ethyldecanoate* = .5636914418 ,
noctane = 1.017457874 ,
napthalene = 1.195660574
])(*components*)

"Aqueous activity coefficients"

```
"UNIFAC activity coefficients " , table( [icosane = .2574115886 1011 ,  
decane = 646864.6078 , phenanthrene = 212343.9129 ,  
aceticacid = 2.029469762 , hexane = 7699.077596 , benzene = 1003.647586 ,  
phenol = 21.11710127 , water = 1.005860695 , ethyldecanoate = 739250.4526 ,  
noctane = 72394.30681 ,  
naphthalene = 15206.88370  
])( components )
```

"beta is " , .8411978449

i := *i*

check := 1

count := 1

```
***** \  
*****"
```

pH := 5

"UNIFAC"

"Hydrocarbon activity coefficients"

```
"UNIFAC activity coefficients " , table( [icosane = .6440451324 ,  
decane = 1.000849267 , phenanthrene = 1.426001161 ,  
aceticacid = 1.634182271 , hexane = .9828806874 , benzene = .9233088848 ,  
phenol = 8.215617311 , water = 402.4941325 , ethyldecanoate = .5640599690 ,  
noctane = 1.017450048 ,  
naphthalene = 1.195262070  
])( components )
```

"Aqueous activity coefficients"

```
"UNIFAC activity coefficients " , table( [icosane = .2662343953 1011 ,  
decane = 658689.9914 , phenanthrene = 217203.5693 ,  
aceticacid = 2.036794961 , hexane = 7791.104249 , benzene = 1016.584113 ,  
phenol = 21.40711846 , water = 1.006613041 , ethyldecanoate = 757587.7263 ,  
noctane = 73488.33462 ,  
naphthalene = 15478.71250  
])( components )
```

"beta is " , .8413191984

i := *i*

```

check := 1
count := 1
***** \
*****"

pH := 6
"UNIFAC"
"Hydrocarbon activity coefficients"
"UNIFAC activity coefficients " , table( [icosane = .6473921271 ,
decane = 1.001664030 , phenanthrene = 1.422229503 ,
aceticacid = 1.634308917 , hexane = .9819630182 , benzene = .9203767233 ,
phenol = 8.678937627 , water = 410.2169717 , ethyldecanoate = .5669888408 ,
noctane = 1.017388835 ,
naphthalene = 1.191783448
])(components )

"Aqueous activity coefficients"
"UNIFAC activity coefficients " , table( [icosane = .3560813465 1011,
decane = 770094.2739 , phenanthrene = 264160.4229 ,
aceticacid = 2.101230690 , hexane = 8631.632743 , benzene = 1135.620026 ,
phenol = 24.08600530 , water = 1.013138856 , ethyldecanoate = 935869.5069 ,
noctane = 83636.68093 ,
naphthalene = 18041.80935
])(components )

"beta is " , .8423754496
i := i
check := 1
count := 1
***** \
*****"

pH := 7
"UNIFAC"
"Hydrocarbon activity coefficients"

```

```
"UNIFAC activity coefficients " , table( [icosane = .6617966827 ,
    decane = 1.005197607 , phenanthrene = 1.406494983 ,
    aceticacid = 1.625727874 , hexane = .9781678256 , benzene = .9081369576 ,
    phenol = 11.64313002 , water = 446.4799038 , ethyldecanoate = .5701515027 ,
    noctane = 1.017210341 ,
    naphthalene = 1.177265619
  ])( components )
```

"Aqueous activity coefficients"

```
"UNIFAC activity coefficients " , table( [icosane = .1116187708 1012,
    decane = .1421778115 107, phenanthrene = 575092.5768 ,
    aceticacid = 2.377184060 , hexane = 12894.98544 , benzene = 1762.249354 ,
    phenol = 38.43787340 , water = 1.039511970 ,
    ethyldecanoate = .2144213486 107, noctane = 138900.7891 ,
    naphthalene = 33161.35364
  ])( components )
```

"beta is " , .8467220188

i := *i*

check := 1

count := 1

```
***** \
*****"
```

pH := 8

"UNIFAC"

"Hydrocarbon activity coefficients"

```
"UNIFAC activity coefficients " , table( [icosane = .6732600679 ,
    decane = 1.007876913 , phenanthrene = 1.394130681 ,
    aceticacid = 1.595102166 , hexane = .9750986237 , benzene = .8985417383 ,
    phenol = 16.93953056 , water = 482.5588798 , ethyldecanoate = .5471043300 ,
    noctane = 1.016965883 ,
    naphthalene = 1.165871159
  ])( components )
```

"Aqueous activity coefficients"

```
"UNIFAC activity coefficients " , table( [icosane = .2478004643 1012,
    decane = .2179953077 107, phenanthrene = 999776.9269 ,
    aceticacid = 2.593467953 , hexane = 17050.46941 , benzene = 2406.980595 ,
    phenol = 53.53561846 , water = 1.058916268 ,
    ethyldecanoate = .3820960390 107, noctane = 197774.4146 ,
    naphthalene = 51099.60468
  ])(components )
```

```
"beta is " , .8500513378
```

```
    i := i
```

```
    check := 1
```

```
    count := 1
```

```
***** \
*****"
```

```
    pH := 10
```

```
"UNIFAC"
```

```
"Hydrocarbon activity coefficients"
```

```
"UNIFAC activity coefficients " , table( [icosane = .6754784874 ,
    decane = 1.008335011 , phenanthrene = 1.391651802 ,
    aceticacid = 1.583162481 , hexane = .9744360958 , benzene = .8966241480 ,
    phenol = 18.87583131 , water = 491.5546344 , ethyldecanoate = .5363839196 ,
    noctane = 1.016851341 ,
    naphthalene = 1.163590591
  ])(components )
```

```
"Aqueous activity coefficients"
```

```
"UNIFAC activity coefficients " , table( [icosane = .2864762636 1012,
    decane = .2356064641 107, phenanthrene = .1106721534 107,
    aceticacid = 2.635105879 , hexane = 17938.15673 , benzene = 2548.840225 ,
    phenol = 56.89339338 , water = 1.062580215 ,
    ethyldecanoate = .4244023475 107, noctane = 210892.3747 ,
    naphthalene = 55324.84261
  ])(components )
```

```
"beta is " , .8507077096
```

```
    i := i
```

```
    check := 1
```

```
count := 1
```

```
***** \
*****"

> "UNIFAC activity coefficients ", TABLE([icosane = .6589415015, decane
= 1.006375019, "UNIFAC" = 1.406296572, aceticacid = 1.620696061, hexane
= .9814682234, benzene = .9105293036, phenol = 7.553015791, water =
395.6031712, ethyldecanoate = .6042545034, noctane = 1.019521552,
napthalene = 1.178732104])(components);
```

Print final results

```
> f_out:=fopen("C:\\1Data\\Fossil Energy\\Modeling\\Programs and
Results\\output_multipH.txt",WRITE,TEXT):
Reconfiguring data for output file.

>
H2:=matrix(pH_number,c,H_carbon_molefraction);H1:=Matrix(pH_number,c,H2
):
A2:=matrix(pH_number,c,Aqueous_molefraction);A1:=Matrix(pH_number,c,A2)
:
D2:=matrix(pH_number,c,Diss_molefraction);D1:=Matrix(pH_number,c,D2):
U2:=matrix(pH_number,c,Undiss_molefraction);U1:=Matrix(pH_number,c,U2):
W2:=matrix(pH_number,c,Water_Hcarbon_ratio);W1:=Matrix(pH_number,c,W2):

> fprintf(f_out,"\n Rows pH:%A\n Columns components:%A\n Hydrocarbon
mole fraction\n%10.8f\n Aqueous_molefraction\n%10.8f\n
Diss_molefraction\n%10.8f\n Undiss_molefraction\n%10.8f\n
Water_Hcarbon_ratio\n%10.8f\n",multipH,components,H1,A1,D1,U1,W1):
fclose(f_out):
```

1. pH dependence

With a pKa of about 4.5, do see an inflection at about pH=7, which agrees with the experimental data (30% octane, 6% acetic acid, 64% water). However, at low solute concentrations, the difference is not so apparent. With a pKa of about 10 (octane-phenol-water), the inflection point is between pH 11 and 12.

Also, did not use DECHEMA NRTL coefficients, but ones from Pedersen for methanol (instead of acetic acid or phenol)- which are quite different. If we use the coefficients for acetic acid from DECHEMA, it is highly soluble in water even when not dissociated.

2. Dependence on salinity

This was calculated using Debye-Huckel term for ionic activity coefficients, in combination with Pitzer coefficients. The effect of a high ionic content is to push the inflection curve for the solute to higher pH. It becomes less soluble (or less dissociated) as one increases the salt content.

APPENDIX B

MAPLE 7 Code for Determination of UNIFAC Activity Coefficients

PROCEDURE TO CALCULATE UNIFAC ACTIVITY COEFFICIENTS FOR LIQUID-LIQUID EQUILIBRIUM

Written by Joanna McFarlane, Dec. 18, 2002

Description of UNIFAC comes from "Chemical and Engineering Thermodynamics", by Stanley I. Sandler, 3rd Ed., Wiley, p. 435 (1997), and "Phase Equilibria in Chemical Engineering", by Stanley M. Walas, p.219-221, Butterworth (1985).

A procedure to calculate UNIFAC coefficients based on LLE parameters.

Overall expression for the activity coefficient, including the combinatorial and residual parts (ACE==activity coefficient expression).

```
> restart:with(linalg):
UNIFAC_LL_activity_coefficients:=proc(x_in::anything,lng_UNIFAC::evaln)

local
ACEqn,i,distinct_flag,distinct_subgroup_array,Em,En,Eqnc1,Fm,Fn,gamma_U
NIFAC,gamma_c,gamma_r,groupi,groupj,groupk,Interaction,j,k,l,lnGm,lnGn,
mole_fraction,n_component,n_distinct_subgroup,n_subgroup,phi,q,r,SA,sli
,slj,slk,subgroup_sum,sumxl,theta,theta_m,theta_n,Total_groups,total_th
eta,total_theta_m,UNIFAC,Vol,Xm,Xn,zu;

global components,subgroup,temperature;
```

Have to initialize a number of variables...

```
>
distinct_flag:='distinct_flag':distinct_subgroup_array:='distinct_subgr
oup_array':Xn:='Xn':theta_n:='theta_n':En:='En':Fn:='Fn':
zu:=10;n_distinct_subgroup:=0;sumxl:=0;mole_fraction:=x_in;
```

This array contains the main group, the subgroup, group volume (R), surface area (Q), and gram molar mass. These parameters are for liquid-liquid interactions.

```
>
UNIFAC:=array(1..57,1..7,[[ "CH2",1,"CH3",1,.9011,.848,15.034],[ "CH2",1,
"CH2",2,.6744,.54,14.026],
[ "CH2",1,"CH",3,.4469,.228,13.018],[ "CH2",1,"C",4,.2195,0,12.010],
[ "C=C",2,"CH2=CH",5,1.3454,1.176,27.044],[ "C=C",2,"CH=CH",6,1.1167,.867
,26.036],
[ "C=C",2,"CH=C",7,.8886,.676,25.028],[ "C=C",2,"CH2=C",8,1.1173,.988,26.
036],
[ "ACH",3,"ACH",9,.5313,.4,13.018],[ "ACH",3,"AC",10,.3652,.12,12.01],
[ "ACCH2",4,"ACCH3",11,1.2663,.968,27.044],[ "ACCH2",4,"ACCH2",12,1.0396,
.66,26.036],
[ "ACCH2",4,"ACCH",13,.8121,.348,25.028],[ "OH",5,"OH",14,1,1.2,17.008],
[ "P1",6,"P1",15,3.2499,3.128,60],[ "P2",7,"P2",16,3.2491,3.124,60],
[ "H2O",8,"H2O",17,.92,1.4,18.016],[ "ACOH",9,"ACOH",18,.8952,.68,29.018]
,
[ "CH2CO",10,"CH3CO",19,1.6724,1.488,43.044],[ "CH2CO",10,"CH2CO",20,1.44
57,1.18,42.036],
```

```

["CHO",11,"CHO",21,.998,.948,29.018],["furfural",12,"furfural",22,3.168,2.481,96.090],
["COOH",13,"COOH",23,1.3013,1.224,45.018],["COOH",13,"HCOOH",24,1.528,1.532,46.026],
["CCOO",14,"CH3COO",25,1.9031,1.728,59.044],["CCOO",14,"CH2COO",26,1.6764,1.42,58.036],
["CH2O",15,"CH3O",27,1.145,1.088,31.034],["CH2O",15,"CH2O",28,.9183,.78,30.026],
["CH2O",15,"CH-
O",29,.6908,.468,29.018],["CH2O",15,"FCH2O",30,.9183,1.1,30.026],
["CCl",16,"CH2Cl",31,1.4654,1.264,49.483],["CCl",16,"CHCl",32,1.238,.952,48.475],
["CCl",16,"CCl",33,1.006,.724,47.467],["CCl2",17,"CH2Cl2",34,2.2564,1.988,84.940],
["CCl2",17,"CHCl2",35,2.0606,1.684,83.932],["CCl2",17,"CCl2",36,1.8016,1.448,82.924],
["CCl3",18,"CHCl3",37,2.87,2.41,119.389],["CCl3",18,"CCl3",38,2.6401,2.184,118.381],
["CCl4",19,"CCl4",39,3.39,2.91,153.838],["ACCl",20,"ACCl",40,1.1562,.844,47.467],
["CCN",21,"CH3CN",41,1.8701,1.724,41.052],["CCN",21,"CH2CN",42,1.6434,1.416,40.044],
["ACNH2",22,"ACNH2",43,1.06,.816,28.034],["CNO2",23,"CH3NO2",44,2.0086,1.868,61.042],
["CNO2",23,"CH2NO2",45,1.7818,1.56,60.034],["CNO2",23,"CHNO2",46,1.5544,1.248,59.026],
["ACNO2",24,"ACNO2",47,1.4199,1.104,58.018],["DOH",25,"(CH2OH)2",48,2.4088,2.248,60.052],
["DEOH",26,"(HOCH2CH2)2O",49,4.0013,3.568,106],["pyridine",27,"C5H5N",50,2.9993,2.113,79.098],
["pyridine",27,"C5H4N",51,2.8332,1.833,78.090],["pyridine",27,"C5H3N",52,2.667,1.553,77.082],
["TCE",28,"CCl2CHCl",53,3.3092,2.86,130],["MFA",29,"HCONHCH3",54,2.4317,2.192,59],
["DMFA",30,"HCON(CH3)2",55,3.0856,2.736,73.09],["TMS",31,"(CH2)4SO2",56,4.0358,3.2,120.19],
["Me2SO",32,"Me2SO",57,2.8266,2.472,78.131]]):

```

This array of interaction parameters is taken from the Walas. The parameters for interactions between main groups - column 2 in the list above. These parameters are for liquid-liquid interactions.

```

> Interaction:=array(1..32,1..32,[[0,74.54,-114.8,-
115.7,644.6,329.6,310.7,1300,2255,472.6,158.1,383,139.4,972.4,662.1,42.
14,-243.9,7.5,-5.55,924.8,696.8,902.2,556.7,575.7,527.5,269.2,-300,-
63.6,928.3,331,561.4,956.5],
[292.3,0,340.7,4102,724.4,1731,1731,896,0,343.7,-214.7,0,1647,-
577.5,289.3,99.61,337.1,4583,5831,0,405.9,0,425.7,0,0,0,0,0,500.7,115.4
,784.4,265.4],
[156.5,-
94.78,0,167,703.9,511.5,577.3,859.4,1649,593.7,362.3,31.14,461.8,6,32.1
4,-18.81,0,-231.9,3000,-878.1,29.13,1.64,-1.77,-11.19,358.9,363.5,-
578.2,0,364.2,-58.1,21.97,84.16],
[104.4,-269.7,-
146.8,0,4000,136.6,906.8,5695,292.6,916.7,1218,715.6,339.1,5688,213.1,-

```

114.1,0,-12.14,-141.3,-107.3,1208,689.6,3629,-175.6,337.7,1023,-
390.7,0,0,0,238,132.2],
[328.2,470.7,-9.21,1.27,0,937.3,991.3,28.73,-195.5,67.07,1409,-140.3,-
104,195.6,262.5,62.05,272.2,-61.57,-41.75,-597.1,-189.3,-348.2,-30.7,-
159,536.6,53.37,183.3,-44.4,0,0,0,0],
[-136.7,-135.7,-223,-162.6,-281.1,0,0,-61.29,-153.2,-47.41,-
344.1,299.3,244.4,19.57,1970,-
166.4,128.6,1544,224.6,0,0,0,150.8,0,0,0,0,0,0,0,0],
[-131.9,-135.7,-252,-273.6,-268.8,0,0,5.89,-153.2,353.8,-338.6,-241.8,-
57.98,487.1,1970,-166.4,507.8,1544,-207,0,0,0,150.8,0,0,0,0,0,0,0,0],
[342.4,220.6,372.8,203.7,-122.4,247,104.9,0,344.5,-171.8,-349.9,66.95,-
465.7,-6.32,64.42,315.9,370.7,356.8,502.9,-97.27,198.3,-
109.8,1539,32.92,-269.2,0,-873.6,1429,-364.2,-117.4,18.41,0],
[-159.8,0,-473.2,-470.4,-63.15,-547.2,-547.2,-595.9,0,-825.7,0,0,0,-
898.3,0,0,0,4894,0,0,-851.6,0,-16.13,-538.6,0,-637.3,0,0,0,0,0],
[66.56,306.1,-78.31,-73.87,216,401.7,-127.6,634.8,-568,0,-
37.36,120.3,1247,258.7,5.202,1000,-301,12.01,-
10.88,902.6,430.6,1010,400,-328.6,211.6,0,0,148,0,0,0,0],
[146.1,517,-75.3,223.2,-431.3,643.4,231.4,623.7,0,128,0,1724,0.75,-
245.8,0,751.8,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
[14.78,0,-10.44,-184.9,444.7,-94.64,732.3,211.6,0,48.93,-
311.6,0,1919,57.7,0,0,-347.9,-249.3,61.59,0,0,0,0,0,-278.2,0,-208.4,-
13.91,0,173.8,0,0],
[1744,-48.52,75.49,147.3,118.4,728.7,349.1,652.3,0,-101.3,1051,-
115.7,0,-117.6,-
96.62,19.77,1670,48.15,43.83,874.3,0,942.2,446.3,0,572.7,0,0,-
2.16,0,0,0,0],
[-320.1,485.6,114.8,-170.2,180.6,-76.64,-152.8,385.9,-
337.3,58.84,1090,-46.13,1417,0,-235.7,0,108.9,-209.7,54.57,629,-
149.2,0,0,0,343.1,0,0,0,0,0,0],
[1571,76.44,52.13,65.69,137.1,-218.1,-
218.1,212.8,0,52.38,0,0,1402,461.3,0,301.1,137.8,-
154.3,47.67,0,0,0,95.18,0,0,0,0,0,0,0,0,0],
[73.8,-24.36,4.68,122.9,455.1,351.5,351.5,770,0,483.9,-
47.51,0,337.1,0,225.4,0,110.5,249.2,62.42,0,0,0,0,0,0,0,0,0,0,0,0],
[27.9,-52.71,0,0,669.2,-186.1,-401.6,740.4,0,550.6,0,808.8,437.7,-
132.9,-197.7,-21.35,0,0,56.33,0,0,0,0,0,0,0,0,0,0,0,0],
[21.23,-185.1,288.5,33.61,418.4,-465.7,-
465.7,793.2,0,342.2,0,203.1,370.4,176.5,-20.93,-157.1,0,0,-
30.1,0,70.04,-75.5,0,0,0,0,18.98,0,0,0,0,0],
[89.97,-293.7,-4.7,134.7,713.5,-
260.3,512.2,1205,1616,550,0,70.14,438.1,129.5,113.9,11.8,17.97,51.9,0,4
75.8,492,1302,490.9,534.7,0,0,0,0,0,0,0,0,0],
[-59.06,0,777.8,-47.13,1989,0,0,390.7,0,190.5,0,0,1349,-246.3,0,0,0,0,-
255.4,0,346.2,0,-154.5,0,124.8,0,-387.7,0,0,0,0,0],
[29.08,34.78,56.41,-53.29,2011,0,0,63.48,0,-349.2,0,0,0,2.41,0,0,0,-
15.62,-54.86,-465.2,0,0,0,0,134.3,0,0,0,0,0],
[175.8,0,-218.9,-15.41,529,0,0,-239.8,-860.3,857.7,0,0,681.4,0,0,0,0,-
216.3,8455,0,0,0,0,179.9,125.3,0,924.5,0,0,0,0,0],
[94.34,375.4,113.6,-97.05,483.8,264.7,264.7,13.32,0,377,0,0,152.4,0,-
94.49,0,0,0,-34.68,794.4,0,0,0,0,139.8,0,0,0,0,0,0,0],
[193.6,0,7.18,-127.1,332.6,0,0,439.9,-
230.4,211.6,0,0,0,0,0,0,0,514.6,0,0,175.8,0,0,963,0,0,0,0,0,0,0],
[108.5,0,247.3,453.4,-289.3,0,0,-424.3,523,82.77,0,-75.23,-
1707,29.86,0,0,0,0,0,-241.7,0,164.4,481.3,-246,0,0,0,0,0,0,0,0],
[81.49,0,-50.71,-30.28,-
99.56,0],


```

    #print("Combinatorial part of activity coefficient
for",components[i],ln(gamma_c[i])):
od;

```

Residual contribution

Group mole fractions.

```

> i:='i':for i from 1 to n_subgroup do Xm[subgroup[i,3]]:=0: od:i:='i':
i:='i':k:='k':for i from 1 to n_subgroup do for k from 1 to n_component
do Xn[subgroup[i,3],k]:=0;od:od:

```

```

k:='k':for k from 1 to n_subgroup do

```

```

Xm[subgroup[k,3]]:=Xm[subgroup[k,3]]+mole_fraction[subgroup[k,1]]*subgr
oup[k,4];

```

```

Xn[subgroup[k,3],subgroup[k,1]]:=subgroup[k,4]/subgroup_sum[subgroup[k,
1]];
od:

```

```

k:='k': for k from 1 to n_subgroup do
distinct_flag[subgroup[k,3]]=0:od:

```

```

k:='k':
for k from 1 to n_subgroup do
    if distinct_flag[subgroup[k,3]]<>1
        then Xm[subgroup[k,3]]:=Xm[subgroup[k,3]]/Total_groups:
        distinct_flag[subgroup[k,3]]:=1:
        n_distinct_subgroup:=n_distinct_subgroup+1:
        distinct_subgroup_array[n_distinct_subgroup]:=subgroup[k,3]:
    end if;
od:

```

```

#print(Xn);print(Xm);print(distinct_subgroup_array);

```

Group surface area fraction.

```

>
k:='k':total_theta_m:=sum(Xm[distinct_subgroup_array[k]]*UNIFAC[distinct
_subgroup_array[k],6],k=1..n_distinct_subgroup):

```

```

k:='k':for k from 1 to n_component do total_theta[k]:=0:od:
i:='i':k:='k':for i from 1 to n_subgroup do for k from 1 to n_component
do theta_n[subgroup[i,3],k]:=0;od:od:

```

```

k:='k':for k from 1 to n_subgroup do
total_theta[subgroup[k,1]]:=total_theta[subgroup[k,1]]+Xn[subgroup[k,3]
,subgroup[k,1]]*UNIFAC[subgroup[k,3],6]:od:

```

```

k:='k':for k from 1 to n_subgroup do
theta_m[subgroup[k,3]]:=Xm[subgroup[k,3]]*UNIFAC[subgroup[k,3],6]/total
_theta_m:
theta_n[subgroup[k,3],subgroup[k,1]]:=Xn[subgroup[k,3],subgroup[k,1]]*U
NIFAC[subgroup[k,3],6]/total_theta[subgroup[k,1]]:
od:

```

```
#print("surface area fraction for each group within mixture ",theta_m);
#print("surface area fraction for each group within each component
",theta_n);
```

Interaction parameters, Em, En

```
> i:='i':for i from 1 to n_subgroup do Em[subgroup[i,3]]:=0;od:
i:='i':k:='k':for i from 1 to n_subgroup do for k from 1 to n_component
do En[subgroup[i,3],k]:=0;od:od:
```

Calculation for groups in mixture

```
i:='i':k:='k':
> for i from 1 to n_distinct_subgroup do
  sli:=distinct_subgroup_array[i];
  groupi:=UNIFAC[sli,2];
  for k from 1 to n_distinct_subgroup do
    slk:=distinct_subgroup_array[k];
    groupk:=UNIFAC[slk,2];
    Em[sli]:=Em[sli]+theta_m[slk]*exp(-
Interaction[groupk,groupi]/Temperature);
  od;
od:#print("Interaction parameters for each subgroup in the mixture
",Em);
```

Calculations for groups within pure components.

```
i:='i':j:='j':k:='k':
for i from 1 to n_component do
#print(components[i]);
for j from 1 to n_distinct_subgroup do
slj:=distinct_subgroup_array[j];
groupj:=UNIFAC[slj,2];
if (theta_n[slj,i]<>0) then
for k from 1 to n_distinct_subgroup do
  slk:=distinct_subgroup_array[k];
  groupk:=UNIFAC[slk,2];
  En[slj,i]:=En[slj,i] + theta_n[slk,i]*exp(-
Interaction[groupk,groupj]/Temperature);
od;
else En[slj,i]:=0;
end if;
#print(slj,En[slj,i]);
od:od:
>
```

Calculation of auxiliary functions, Fm and Fn

```
> i:='i':for i from 1 to n_subgroup do Fm[subgroup[i,3]]:=0;od:
i:='i':k:='k':
for i from 1 to n_subgroup do for k from 1 to n_component do
Fn[subgroup[i,3],k]:=0;od:od:
```

Calculations for groups in the mixture.

```
i:='i':k:='k':
for i from 1 to n_distinct_subgroup do
```

```

sli:=distinct_subgroup_array[i];
groupi:=UNIFAC[sli,2];
for k from 1 to n_distinct_subgroup do
  slk:=distinct_subgroup_array[k];
  groupk:=UNIFAC[slk,2];
  Fm[sli]:=Fm[sli]+theta_m[slk]*exp(-
Interaction[groupi,groupk]/Temperature)/Em[slk];
od;
od:#print("Auxiliary function Fm for each subgroup in the mixture
",Fm);

```

Calculations for groups within pure components

```

i:='i':j:='j':k:='k':
for i from 1 to n_component do
#print(components[i]);
for j from 1 to n_distinct_subgroup do
slj:=distinct_subgroup_array[j];
groupj:=UNIFAC[slj,2];
if (theta_n[slj,i]<>0) then
(required to zero out contributions from groups that don't exist in a molecule)
for k from 1 to n_distinct_subgroup do
  slk:=distinct_subgroup_array[k];
  if En[slk,i]<>0 then (required to avoid division by zero)
    groupk:=UNIFAC[slk,2];
    Fn[slj,i]:=Fn[slj,i] + theta_n[slk,i]*exp(-
Interaction[groupj,groupk]/Temperature)/En[slk,i];
  end if;
od;
else Fn[slj,i]:=0;
end if;
#print(slj,Fn[slj,i]);
od:od:

```

Big Gamma - no change from first system

```

i:='i':for i from 1 to n_subgroup do lnGm[subgroup[i,2]]:=0;od:
i:='i':k:='k':
for i from 1 to n_subgroup do for k from 1 to n_component do
lnGn[subgroup[i,2],k]:=0;od:od:

i:='i':
for i from 1 to n_distinct_subgroup do
  sli:=distinct_subgroup_array[i];
  lnGm[sli]:=UNIFAC[sli,6]*(1-ln(Em[sli])-Fm[sli]);
od;

i:='i':k:='k':
for i from 1 to n_component do
>for k from 1 to n_distinct_subgroup do
slk:=distinct_subgroup_array[k];
if En[slk,i]<>0 then lnGn[slk,i]:=UNIFAC[slk,6]*(1-ln(En[slk,i])-
Fn[slk,i]) else lnGn[slk,i]:=0 end if;
#print(components[i],slk,lnGn[slk,i]);
od;od;

```

Calculation of the residual component. In this case, the activity coefficient for the averaged component is the same as those averaged over each component. It breaks down however, when one calculates the over activity coefficient, and logarithms are taken.

```
> i:='i':k:='k':
for i from 1 to n_component do ln(gamma_r[i]):=0:od:i:='i':

for k from 1 to n_subgroup do
ln(gamma_r[subgroup[k,1]]):=ln(gamma_r[subgroup[k,1]])+subgroup[k,4]*(1
nGm[subgroup[k,3]]-lnGn[subgroup[k,3],subgroup[k,1]]);
od;
#i:='i':for i from 1 to n_component do print("Residual part of activity
coefficient for",components[i],ln(gamma_r[i])):od:
> i:='i':for i from 1 to n_component do
lng_UNIFAC[i]:=ln(gamma_r[i])+ln(gamma_c[i]):gamma_UNIFAC[components[i]
]:=exp(lng_UNIFAC[i]):od;
print("UNIFAC activity coefficients ",gamma_UNIFAC(components));
> end proc:
```

Warning, the protected names norm and trace have been redefined and unprotected

```
*****
> save UNIFAC_LL_activity_coefficients,
"UNIFAC_LL_activity_coefficients.mws":
*****
```

First test system:

```
>
components:=["ethane","hexane","octanol"];Temperature:=300;u:= [.2
5,.25,.5];
subgroup:=matrix([[1,1,1,2],[2,1,1,2],[2,1,2,4],[3,1,1,1],[3,1,2,
7],[3,5,14,1]]);
```

components := ["ethane", "hexane", "octanol"]

Temperature := 300

u := [.25, .25, .5]

subgroup :=
$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 2 & 4 \\ 3 & 1 & 1 & 1 \\ 3 & 1 & 2 & 7 \\ 3 & 5 & 14 & 1 \end{bmatrix}$$

```
> UNIFAC_LL_activity_coefficients(u,a);print("Ln activity coefficients
");eval(a);
```

```
"UNIFAC activity coefficients " , table( ["ethane" = .8102104066 ,
    "hexane" = 1.371848082 ,
    "octanol" = 1.026048572
  ])( components )
```

"Ln activity coefficients "

```
table( [1 = -.2104613038 , 2 = .3161587958 , 3 = .0257150869 ])
```

Second test system:

Note that the value for the octanol hasn't changed with the transformation, but the butane is not the average of ethane and hexane. Hence, it looks as if we will have to use a distribution to represent the water soluble organic classes, rather than "mean" compounds.

```
> components := ["butane", "octanol"]; Temperature := 300; x := [.5, .5];
a := 'a': subgroup := matrix( [[1,1,1,2], [1,1,2,2], [2,1,1,1], [2,1,2,7], [2,5,14,1]] );
```

```
components := ["butane", "octanol"]
```

```
Temperature := 300
```

```
x := [.5, .5]
```

```
subgroup := 
$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 2 \\ 2 & 1 & 1 & 1 \\ 2 & 1 & 2 & 7 \\ 2 & 5 & 14 & 1 \end{bmatrix}$$

```

```
> UNIFAC_LL_activity_coefficients(x,a); print("Ln activity coefficients "); eval(a);
```

```
"UNIFAC activity coefficients " ,
```

```
table( ["butane" = 1.151757942 , "octanol" = 1.026048572 ])( components )
```

"Ln activity coefficients "

```
table( [1 = .1412894208 , 2 = .0257150869 ])
```

Third test system: The calculated activity coefficients should be compared to those of Sandler's program, which are 1.0857 for phenol and 1.5672 for water respectively, using VL parameters.

```
> components := ["phenol", "water"]; Temperature := 323.15; y := [0.5, 0.5];
subgroup := matrix( [[1,3,9,5], [1,9,18,1], [2,8,17,1]] );
UNIFAC_LL_activity_coefficients(y,a); print("Ln activity coefficients "); eval(a);
```

```
components := ["phenol", "water"]
```

```
Temperature := 323.15
```

$y := [.5, .5]$

$subgroup := \begin{bmatrix} 1 & 3 & 9 & 5 \\ 1 & 9 & 18 & 1 \\ 2 & 8 & 17 & 1 \end{bmatrix}$

"UNIFAC activity coefficients " ,

table(["phenol" = 1.112600164 , "water" = 1.817483760])(*components*)

"Ln activity coefficients "

table([1 = .1066997665 , 2 = .5974529951])

Fourth test system:

>

```
components=['noctane','phenol','water'];Temperature:=279.5;x=[0.959,0.001,0.040];subgroup:=matrix([[1,1,1,2],[1,1,2,6],[2,3,9,5],[2,9,18,1],[3,8,17,1]]);
```

```
UNIFAC_LL_activity_coefficients(x,a);print("Ln activity coefficients ");eval(a);
```

components := [noctane , phenol , water]

Temperature := 279.5

x := [.959, .001, .040]

$subgroup := \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 6 \\ 2 & 3 & 9 & 5 \\ 2 & 9 & 18 & 1 \\ 3 & 8 & 17 & 1 \end{bmatrix}$

"UNIFAC activity coefficients " ,

table([*water* = 265.0827433 , *phenol* = 2.287250216 , *noctane* = 1.034266528])(*components*)

"Ln activity coefficients "

table([1 = .03369250659 , 2 = .8273503172 , 3 = 5.580042016])

Fifth test system. Note that temperature must be input as a real- not an integer!

>

```
components=['noctane','aceticacid','water'];Temperature:=300.0;x=[0.6,0.006,0.394];subgroup:=matrix([[1,1,1,2],[1,1,2,6],[2,1,1,1],[2,13,23,1],[3,8,17,1]]);UNIFAC_LL_activity_coefficients(x,a);print("Ln activity coefficients ");eval(a);
```

components := [noctane , aceticacid , water]

Temperature := 300.0

x := [.6, .006, .394]

$$subgroup := \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 6 \\ 2 & 1 & 1 & 1 \\ 2 & 13 & 23 & 1 \\ 3 & 8 & 17 & 1 \end{bmatrix}$$

```
"UNIFAC activity coefficients " , table( [water = 9.740263386 ,
  aceticacid = .8577199907 ,
  noctane = 2.114382384
  ])(components )
```

"Ln activity coefficients "

table([1 = .7487627527 , 2 = -.1534775839 , 3 = 2.276268159])

>

```
components:=['noctane','napthoicacid','water'];Temperature:=300.0;x:=[0
.6,0.006,0.394];subgroup:=matrix([[1,1,1,3],[1,1,2,4],[1,1,3,1],[2,3,9,
7],[2,3,10,3],[2,13,23,1],[3,8,17,1]]);
;UNIFAC_LL_activity_coefficients(x,a);print("Ln activity coefficients
");eval(a);
```

components := [noctane , napthoicacid , water]

Temperature := 300.0

x := [.6, .006, .394]

$$subgroup := \begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 1 & 2 & 4 \\ 1 & 1 & 3 & 1 \\ 2 & 3 & 9 & 7 \\ 2 & 3 & 10 & 3 \\ 2 & 13 & 23 & 1 \\ 3 & 8 & 17 & 1 \end{bmatrix}$$

```
"UNIFAC activity coefficients " , table( [water = 9.811328505 ,
  napthoicacid = 1.712477373 ,
  noctane = 2.105777302
  ])(components )
```

"Ln activity coefficients "

table([1 = .7446846635 , 2 = .5379410784 , 3 = 2.283537688])

APPENDIX C

MAPLE 7 Code for Partial Least-Squares Analysis

PARTIAL LEAST SQUARES

TRAINING DATA FROM THERMODYNAMIC RUNS

(ethanol, C3-C7OH, C8-C12OH - not pH sensitive)

by Joanna McFarlane

April 15, 2005

Introduction

The object is to build a model that predicts an outcome, Y, based on a response, 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). 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):  
Warning, the protected names norm and trace have been redefined and  
unprotected  
  
Warning, the name changecoords has been redefined
```

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, 15 separate samples (or objects) were used for model development. The X data are the measured variables (reponses) and the Y data are the concentrations (or outcomes). Note that one needs at least as many predictor variables as response variables or else the system does not converge.

```
>  
X:=[[300,2,108000,50],[325,3,108000,50],[350,4,108000,50],[300,5,108000,  
,10],[325,6,108000,10],  
[350,7,108000,10],[300,8,108000,30],[325,9,108000,30],[350,10,108000,30  
],  
[300,2,0,50],[325,3,0,50],[350,4,0,50],[300,5,0,10],[325,6,0,10],[350,7  
,0,10],  
[300,8,0,30],[325,9,0,30],[350,10,0,30],[300,2,50000,50],[325,3,50000,5  
0],[350,4,50000,50],  
[300,5,50000,10],[325,6,50000,10],[350,7,50000,10],[300,8,50000,30],[32  
5,9,50000,30],
```

```

[350,10,50000,30],[300,2,150000,50],[325,3,150000,50],[350,4,150000,50]
,[300,5,150000,10],
[325,6,150000,10],[350,7,150000,10],[300,8,150000,30],[325,9,150000,30]
,[350,10,150000,30]]:
array(X);
Y:=[0.899054867,0.208721139,0.007755868],
[0.890939845,0.218342029,0.009459123],
[0.883061198,0.226620046,0.011245807],
[0.638900958,0.05524475,0.001928888],
[0.616561058,0.056760117,0.002246529],
[0.595928249,0.057343625,0.002502255],
[0.845819554,0.137420873,0.00460774],
[0.838369123,0.14717009,0.005710004],
[0.830271166,0.155951259,0.006931545],
[0.899049279,0.208704376,0.00775028],
[0.890889561,0.218174413,0.009447948],
[0.883017877,0.225819811,0.011147663],
[0.635657084,0.052467572,0.001754256],
[0.621711803,0.054434838,0.002012995],
[0.611004807,0.057848397,0.002394652],
[0.854799178,0.14488589,0.004851204],
[0.843284453,0.151793517,0.00591536],
[0.833183954,0.158534636,0.007054386],
[0.899054867,0.208721139,0.007755868],
[0.890934258,0.218336441,0.009459123],
[0.883038852,0.226547421,0.011234634],
[0.637694615,0.054688058,0.001898099],
[0.615156912,0.055539651,0.002159033],
[0.599241207,0.056468701,0.002389593],
[0.849874925,0.140210806,0.004674824],
[0.84153497,0.150064487,0.00583518],
[0.831276767,0.156876665,0.006979711],
[0.899054867,0.208721139,0.007755868],
[0.890939845,0.218347616,0.009459123],
[0.883061198,0.22663122,0.011245807],
[0.639181996,0.055353102,0.001934531],
[0.617644382,0.057270375,0.002279401],
[0.596355838,0.0581807,0.002576005],
[0.842718148,0.136143974,0.004613107],
[0.836258576,0.145570021,0.005654356],
[0.828630538,0.154440198,0.006853131]]:array(Y);
num_samples:=rowdim(array(X));num_ind_var:=coldim(array(X));num_dep_var
:=coldim(array(Y));

```

300	2	108000	50
325	3	108000	50
350	4	108000	50
300	5	108000	10
325	6	108000	10
350	7	108000	10
300	8	108000	30
325	9	108000	30
350	10	108000	30
300	2	0	50
325	3	0	50
350	4	0	50
300	5	0	10
325	6	0	10
350	7	0	10
300	8	0	30
325	9	0	30
350	10	0	30
300	2	50000	50
325	3	50000	50
350	4	50000	50
300	5	50000	10
325	6	50000	10
350	7	50000	10
300	8	50000	30
325	9	50000	30
350	10	50000	30
300	2	150000	50
325	3	150000	50
350	4	150000	50
300	5	150000	10
325	6	150000	10
350	7	150000	10
300	8	150000	30
325	9	150000	30
350	10	150000	30

.899054867	.208721139	.007755868
.890939845	.218342029	.009459123
.883061198	.226620046	.011245807
.638900958	.05524475	.001928888
.616561058	.056760117	.002246529
.595928249	.057343625	.002502255
.845819554	.137420873	.00460774
.838369123	.14717009	.005710004
.830271166	.155951259	.006931545
.899049279	.208704376	.00775028
.890889561	.218174413	.009447948
.883017877	.225819811	.011147663
.635657084	.052467572	.001754256
.621711803	.054434838	.002012995
.611004807	.057848397	.002394652
.854799178	.14488589	.004851204
.843284453	.151793517	.00591536
.833183954	.158534636	.007054386
.899054867	.208721139	.007755868
.890934258	.218336441	.009459123
.883038852	.226547421	.011234634
.637694615	.054688058	.001898099
.615156912	.055539651	.002159033
.599241207	.056468701	.002389593
.849874925	.140210806	.004674824
.84153497	.150064487	.00583518
.831276767	.156876665	.006979711
.899054867	.208721139	.007755868
.890939845	.218347616	.009459123
.883061198	.22663122	.011245807
.639181996	.055353102	.001934531
.617644382	.057270375	.002279401
.596355838	.0581807	.002576005
.842718148	.136143974	.004613107
.836258576	.145570021	.005654356
.828630538	.154440198	.006853131

num_samples := 36

```
num_ind_var := 4
```

```
num_dep_var := 3
```

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);
```

```

E := [-1.207614729  -1.527525232  .5372459074  1.207614729
      0.          -1.145643924  .5372459074  1.207614729
      1.207614729  -.7637626160  .5372459074  1.207614729
      -1.207614729 -.3818813079  .5372459074  -1.207614729
      0.          0.          .5372459074  -1.207614729
      1.207614729 .3818813079  .5372459074  -1.207614729
      -1.207614729 .7637626160  .5372459074  0.
      0.          1.145643924  .5372459074  0.
      1.207614729 1.527525232  .5372459074  0.
      -1.207614729 -1.527525232  -1.334449512  1.207614729
      0.          -1.145643924  -1.334449512  1.207614729
      1.207614729 -.7637626160  -1.334449512  1.207614729
      -1.207614729 -.3818813079  -1.334449512  -1.207614729
      0.          0.          -1.334449512  -1.207614729
      1.207614729 .3818813079  -1.334449512  -1.207614729
      -1.207614729 .7637626160  -1.334449512  0.
      0.          1.145643924  -1.334449512  0.
      1.207614729 1.527525232  -1.334449512  0.
      -1.207614729 -1.527525232  -.4679238549  1.207614729
      0.          -1.145643924  -.4679238549  1.207614729
      1.207614729 -.7637626160  -.4679238549  1.207614729
      -1.207614729 -.3818813079  -.4679238549  -1.207614729
      0.          0.          -.4679238549  -1.207614729
      1.207614729 .3818813079  -.4679238549  -1.207614729
      -1.207614729 .7637626160  -.4679238549  0.
      0.          1.145643924  -.4679238549  0.
      1.207614729 1.527525232  -.4679238549  0.
      -1.207614729 -1.527525232  1.265127460  1.207614729
      0.          -1.145643924  1.265127460  1.207614729
      1.207614729 -.7637626160  1.265127460  1.207614729
      -1.207614729 -.3818813079  1.265127460  -1.207614729
      0.          0.          1.265127460  -1.207614729
      1.207614729 .3818813079  1.265127460  -1.207614729
      -1.207614729 .7637626160  1.265127460  0.
      0.          1.145643924  1.265127460  0.
      1.207614729 1.527525232  1.265127460  0.]

```

```
> 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);

```

.9635890172	1.008010568	.6080623514
.8961277404	1.150533763	1.142706529
.8306314809	1.273163731	1.703538698
-1.199105659	-1.265577983	-1.221000782
-1.384820273	-1.243129441	-1.121294684
-1.556343608	-1.234485394	-1.041023428
.5210366510	-.04822661277	-.3801210413
.4591002092	.09619760970	-.03412526882
.3917807960	.2262812364	.3493110240
.9635425634	1.007762242	.6063083030
.8957097227	1.148050711	1.139198746
.8302713476	1.261309106	1.672731733
-1.226072422	-1.306718904	-1.275816991
-1.342001432	-1.277575963	-1.194599969
-1.431010140	-1.227007737	-1.074799532
.5956854814	.06235962979	-.3036987733
.4999620131	.1646887357	.03033506723
.4159951971	.2645512036	.3878702681
.9635890172	1.008010568	.6080623514
.8960812949	1.150450983	1.142706529
.8304457156	1.272087870	1.700031543
-1.209134151	-1.273824779	-1.230665313
-1.396493128	-1.261209340	-1.148759289
-1.528802539	-1.247446458	-1.076387530
.5547494994	-.006896739383	-.3590636717
.4854183240	.1390750055	.005166921451
.4001404933	.2399901365	.3644301174
.9635890172	1.008010568	.6080623514
.8961277404	1.150616529	1.142706529
.8306314809	1.273329262	1.703538698
-1.196769352	-1.263972864	-1.219229470
-1.375814453	-1.235570515	-1.110976309
-1.552789003	-1.222085024	-1.017873631
.4952542430	-.06714250574	-.3784363638
.4415549462	.07249429883	-.05159292878
.3781420333	.2038964839	.3246972178

Model Building

Initialization

The dimensions of pm and wm are [ind,1]. The dimensions of qm is [dep,1]. The dimensions of um and tm are [#samples,1]

>

```
check:=1:tstart:=ColumnToList(X2,[1]);ustart:=ColumnToList(Y2,[1]);Xm:=
matrix(X2):Ym:=matrix(Y2):
pm:=matrix(num_ind_var,1,1):qm:=matrix(num_dep_var,1,1):wm:=matrix(num_
ind_var,1,1):tm:=matrix(num_samples,1,1):um:=matrix(num_samples,1,1):bm
:=[]:j:='j':
tstart := [-1.120459090 , -1.120459090 , -1.120459090 , -1.120459090 ,
-1.120459090 , 1.292837411 , 1.292837411 , 1.292837411 , 1.292837411 ,
.08618916074 , .08618916074 , .08618916074 , .08618916074 , .08618916074 ]
tstart := [-1.207614729 , 0., 1.207614729 , -1.207614729 , 0., 1.207614729 ,
-1.207614729 , 0., 1.207614729 , -1.207614729 , 0., 1.207614729 , -1.207614729 ,
0., 1.207614729 , -1.207614729 , 0., 1.207614729 , -1.207614729 , 0.,
1.207614729 , -1.207614729 , 0., 1.207614729 , -1.207614729 , 0., 1.207614729 ,
-1.207614729 , 0., 1.207614729 , -1.207614729 , 0., 1.207614729 , -1.207614729 ,
0., 1.207614729 ]
ustart := [.9635890172 , .8961277404 , .8306314809 , -1.199105659 , -1.384820273 ,
-1.556343608 , .5210366510 , .4591002092 , .3917807960 , .9635425634 ,
.8957097227 , .8302713476 , -1.226072422 , -1.342001432 , -1.431010140 ,
.5956854814 , .4999620131 , .4159951971 , .9635890172 , .8960812949 ,
.8304457156 , -1.209134151 , -1.396493128 , -1.528802539 , .5547494994 ,
.4854183240 , .4001404933 , .9635890172 , .8961277404 , .8306314809 ,
-1.196769352 , -1.375814453 , -1.552789003 , .4952542430 , .4415549462 ,
.3781420333 ]
> 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);wne
w:=scalarmul(wold,worm);
w_ip:=1/innerprod(wnew,wnew);t1:=multiply(Xm,wnew);t2:=scalarmul(t1,w_i
p);
  Y block
t_ip:=1/innerprod(t2,t2);
q:=multiply(t2,Ym);qold:=scalarmul(q,t_ip);qorm:=1/norm(qold,2);qnew:=s
calarmul(qold,qorm);
q_ip:=1/innerprod(qnew,qnew);u1:=multiply(Ym,qnew);u2:=scalarmul(u1,q_i
p);
  Check convergence
check:=norm(t2-tstart);ustart:=u2;tstart:=convert(t2,list);od;
```

Calculate X-loadings and rescale scores and weights

```
print("*****CONVERGENCE*****CONVERGENCE*****CONVERGENCE*****");
t_ip:=1/innerprod(t2,t2);
p:=multiply(t2,Xm);pold:=scalarmul(p,t_ip):porm:=1/norm(pold,2):pnew:=s
calarmul(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 regresion coefficient

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

```
*****CONVERGENCE*****CONVERGENCE*****CONVERGENCE \
*****"
```

"b=", 1.393129283

```
*****CONVERGENCE*****CONVERGENCE*****CONVERGENCE \
*****"
```

"b=", .6316679173

```
*****CONVERGENCE*****CONVERGENCE*****CONVERGENCE \
*****"
```

"b=", .5457312334

Output files (loadings p, q and weights w are used for prediction, b are 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:=delc
ols(wm,1..1);um:=delcols(um,1..1);
```

>

$$pm := \begin{bmatrix} -.004025648884 & -.6534447250 & -.9479292362 \\ -.5559196457 & -.7355818879 & .2607353378 \\ -.004868713296 & .008066545266 & -.03348573448 \\ .8312120291 & -.1785054869 & .1797941947 \end{bmatrix}$$

$$\begin{aligned}
 qm &:= \begin{bmatrix} .5340105602 & -.6309936810 & .9358950730 \\ .6055465724 & -.4955398904 & .3103388825 \\ .5900390412 & -.5968979747 & -.1667038993 \end{bmatrix} \\
 tm &:= \begin{bmatrix} 1.566580883 & 1.080525089 & .5134967537 \\ 1.559912744 & .3044043964 & -.2251154706 \\ 1.553244606 & -.4717162964 & -.9637276950 \\ -1.187080409 & 1.504560350 & .2333108811 \\ -1.193748548 & .7284396576 & -.5053013441 \\ -1.200416686 & -.04768103534 & -1.243913569 \\ -.3702594436 & -.2439362053 & 1.434206025 \\ -.3769275820 & -1.020056898 & .6955937994 \\ -.3835957203 & -1.796177591 & -.04301842494 \\ 1.579080298 & 1.065675489 & .5539363846 \\ 1.572412159 & .2895547965 & -.1846758400 \\ 1.565744021 & -.4865658963 & -.9232880652 \\ -1.174580994 & 1.489710750 & .2737505117 \\ -1.181249133 & .7135900578 & -.4648617135 \\ -1.187917271 & -.06253063499 & -1.203473938 \\ -.3577600289 & -.2587858051 & 1.474645655 \\ -.3644281673 & -1.034906498 & .7360334301 \\ -.3710963056 & -1.811027192 & -.002578795166 \\ 1.573293532 & 1.072550304 & .5352143338 \\ 1.566625393 & .2964296113 & -.2033978914 \\ 1.559957255 & -.4796910817 & -.9420101158 \\ -1.180367761 & 1.496585564 & .2550284606 \\ -1.187035899 & .7204648725 & -.4835837647 \\ -1.193704037 & -.05565582049 & -1.222195990 \\ -.3635467949 & -.2519109903 & 1.455923603 \\ -.3702149333 & -1.028031683 & .7173113786 \\ -.3768830716 & -1.804152376 & -.02130084620 \\ 1.561719999 & 1.086299934 & .4977702312 \\ 1.555051861 & .3101792405 & -.2408419937 \\ 1.548383722 & -.4659414521 & -.9794542182 \\ -1.191941293 & 1.510335194 & .2175843577 \\ -1.198609431 & .7342145016 & -.5210278670 \\ -1.205277570 & -.04190619112 & -1.259640092 \\ -.3751203271 & -.2381613611 & 1.418479501 \\ -.3817884655 & -1.014282054 & .6798672764 \\ -.3884566038 & -1.790402746 & -.05874494858 \end{bmatrix}
 \end{aligned}$$

$$wm := \begin{bmatrix} .09752982840 & -.3299119576 & -.9483712268 \\ -.3258776826 & -.9944656984 & .2609099732 \\ -.006678124331 & .005876984829 & -.01852463784 \\ .9855467344 & -.2961388463 & .1799970558 \end{bmatrix}$$

	1.483744582	.6884499658	.2671839568
	1.849484918	.3320741853	-.1515000645
	2.219680256	-.03131517847	-.5788843196
	-2.127139625	.4766516154	.2760058426
	-2.153889062	.5140083821	-.2181088540
	-2.192886786	.5608518391	-.6999862114
	.02474935910	-.5882393706	.7406712087
	.2832812518	-.8364394702	.3591962400
	.5523460511	-1.096484782	-.03800855055
	1.482534446	.7068749567	.2510449965
	1.845688360	.3528878276	-.1683879374
	2.194132101	.01040075369	-.5940755016
	-2.198796608	.5639997597	.2308274048
	-2.195135302	.5650410018	-.1928155558
	-2.141358541	.5154482595	-.5910468460
	.1766698647	-.7185328669	.8158031769
	.3846105681	-.9174137038	.3916374456
	.6112025039	-1.136518511	-.02620855504
<i>um :=</i>	1.483744582	.6977007707	.2584244389
	1.849409989	.3413953178	-.1603287403
	2.216860213	-.01932061178	-.5875669187
	-2.143191215	.5020856890	.2569125229
	-2.187275900	.5559775628	-.2388253781
	-2.206894293	.5802558883	-.6810971668
	.08020420940	-.6333108629	.7727793012
	.3464835894	-.8884961187	.3818240785
	.5740324507	-1.108326757	-.03721027558
	1.483744582	.6817511066	.2735270565
	1.849535037	.3253343139	-.1451312808
	2.219780492	-.03809606446	-.5724898498
	-2.123874896	.4666258683	.2847383264
	-2.138414333	.4917221429	-.2027115380
	-2.169820304	.5319471022	-.6903272196
	.0005208522000	-.5703016933	.7167334995
	.2492518362	-.8098950312	.3446746816
	.5169846911	-1.068793199	-.04727354372

Model Testing

Sum of Squares, F Test, XVAL, square root PRESS.

Sum of squares. First calculated for samples. Then calculated for each column of the X matrix and Y matrix to show which variables have been minimized. If the sum of squares is large, this means that it did not play a great role in the model. The overall sum of squares for X and Y show how well the model covered for the X data 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;

sum_squares_list := [-.0084862189 , -.0082024280 , -.0079186374 , -.0080842685 ,
  -.0078004780 , -.007516687 , -.008978269 , -.0086944783 , -.00841068770 ,
  .0201946148 , .0204784057 , .0207621963 , .0205965649 , .0208803558 ,
  .021164146 , .019702565 , .0199863554 , .02027014567 , .0069164510 ,
  .0072002418 , .0074840324 , .0073184012 , .0076021921 , .007885983 ,
  .006424400 , .0067081915 , .00699198213 , -.0196398763 , -.0193560858 ,
  -.0190722950 , -.0192379265 , -.0189541356 , -.018670345 , -.020131927 ,
  -.0198481361 , -.01956434524 ]

sum_squares_x_1 := .008225030198

SSx_cols := .008225030198

sum_squares_list := [.0042957928 , .00415213487 , .0040084773 , .00409232360 ,
  .0039486653 , .0038050077 , .0045448705 , .0044012131 , .00425755556 ,
  -.0102226745 , -.01036633227 , -.0105099896 , -.01042614414 , -.0105698017 ,
  -.0107134592 , -.0099735966 , -.0101172542 , -.01026091297 , -.0035011617 ,
  -.00364481958 , -.0037884772 , -.00370463182 , -.0038482891 , -.0039919466 ,
```

-.0032520836 , -.0033957414 , -.003539399671 , .0099418634 , .00979820548 ,
.0096545475 , .00973839300 , .0095947359 , .0094510779 , .0101909412 ,
.0100472834 , .00990362601]

$sum_squares_x_2 := .002107635414$

$SSx_cols := .01033266561$

$sum_squares_list := [.5533518520 , .5348470266 , .5163422013 , .5271423352 ,$
 $.5086375099 , .4901326844 , .5854363848 , .5669315595 , .5484267340 ,$
 $-1.316808776 , -1.335313601 , -1.353818426 , -1.343018292 , -1.361523118 ,$
 $-1.380027944 , -1.284724243 , -1.303229068 , -1.321733894 , -.4509936703 ,$
 $-.4694984957 , -.4880033211 , -.4772031870 , -.4957080124 , -.5142128378 ,$
 $-.4189091374 , -.4374139628 , -.4559187882 , 1.280636541 , 1.262131716 ,$
 $1.243626890 , 1.254427025 , 1.235922198 , 1.217417373 , 1.312721074 ,$
 $1.294216248 , 1.275711423]$

$sum_squares_x_3 := 34.97129998$

$SSx_cols := 34.98163265$

$sum_squares_list := [.00600977579 , .00580880174 , .0056078256 , .00572512232 ,$
 $.00552414784 , .0053231723 , .0063582350 , .0061572601 , .005956284669 ,$
 $-.01430143418 , -.01450240814 , -.0147033840 , -.01458608760 , -.01478706217 ,$
 $-.0149880377 , -.0139529745 , -.0141539497 , -.01435492510 , -.00489809594 ,$
 $-.00509907081 , -.0053000459 , -.00518274920 , -.00538372436 , -.0055846998 ,$
 $-.0045496366 , -.0047506119 , -.004951587111 , .01390858074 , .01370760466 ,$
 $.0135066306 , .01362392653 , .01342295106 , .0132219770 , .0142570389 ,$
 $.0140560637 , .01385508872]$

$sum_squares_x_4 := .004125015151$

$SSx_cols := 34.98575767$

$sum_squares_list := [-.0334551553 , -.0280565867 , -.0206930028 , .1645402457 ,$
 $.0516854773 , -.0469780134 , -.0332547999 , -.0223313968 , -.01679096560 ,$
 $-.0693736539 , -.06434664906 , -.0569251795 , .1017014384 , .0586322740 ,$
 $.0424834103 , .0055219865 , -.0173416370 , -.02844860846 , -.0527197724 ,$
 $-.0473676489 , -.0401433840 , .1352471375 , .0207480057 , -.0387015608 ,$
 $-.0188065675 , -.0152778980 , -.02769588433 , -.0195049155 , -.0141063480 ,$
 $-.0067427630 , .1808267925 , .0746415361 , -.0294731684 , -.0450869683 ,$
 $-.0259264205 , -.01647948821]$

$sum_squares_y_1 := .1360415699$

$SSy := .1360415699$

```

sum_squares_list := [-.06230866744 , -.03200633542 , -.0215972304 , .1672871162 ,
    .07751479652 , -.0260620204 , -.0551295712 , -.0229262118 , -.005063448716 ,
    -.08459865101 , -.05653104484 , -.0554935136 , .1041045375 , .02102661590 ,
    -.0406260218 , .0334150135 , .0235232562 , .01116486053 , -.07414585449 ,
    -.04392630234 , -.0345102784 , .1472031343 , .04759771052 , -.0508602713 ,
    -.0256368843 , .0081139973 , -.003191735156 , -.05373691050 , -.02335181363 ,
    -.0128599435 , .1774639920 , .09364547731 , -.0050898946 , -.0654737082 ,
    -.0380577670 , -.01887644488 ]

```

```

sum_squares_y_2 := .1576447607

```

```

SSy := .2936863306

```

```

sum_squares_list := [-.2255499573 , -.04524975453 , .1612384385 , .3432870762 ,
    .08864919895 , -.1854235212 , -.0372639105 , -.04561211380 , -.01651979678 ,
    -.2394984285 , -.06095196142 , .1182370496 , .2762764438 , .00314949086 ,
    -.2313940483 , .0269639343 , .00665379910 , .009845023557 , -.2320988143 ,
    -.05179861177 , .1511824262 , .3270736886 , .05463573683 , -.2273364802 ,
    -.0227553979 , -.01286878036 , -.007949560257 , -.2208076808 , -.04050747951 ,
    .1659807146 , .3498006641 , .1037098490 , -.1575314482 , -.0308369576 ,
    -.05833749838 , -.03639132721 ]

```

```

sum_squares_y_3 := .9319086109

```

```

SSy := 1.225594942

```

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 on the input variable. Mapping of t scores indicates correlations between independent variables, u scores indicates correlations between dependent variables. In this case, there doesn't appear to be any correlation between the variables - which is a good thing.

```

> 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;

```

```
t1plot := [ 1.566580883 , 1.559912744 , 1.553244606 , -1.187080409 , -1.193748548 ,  
-1.200416686 , -0.3702594436 , -0.3769275820 , -0.3835957203 , 1.579080298 ,  
1.572412159 , 1.565744021 , -1.174580994 , -1.181249133 , -1.187917271 ,  
-0.3577600289 , -0.3644281673 , -0.3710963056 , 1.573293532 , 1.566625393 ,  
1.559957255 , -1.180367761 , -1.187035899 , -1.193704037 , -0.3635467949 ,  
-0.3702149333 , -0.3768830716 , 1.561719999 , 1.555051861 , 1.548383722 ,  
-1.191941293 , -1.198609431 , -1.205277570 , -0.3751203271 , -0.3817884655 ,  
-0.3884566038 ]
```

```
ts1 := "t1"
```

```
t2plot := [ 1.080525089 , .3044043964 , -0.4717162964 , 1.504560350 , .7284396576 ,  
-0.04768103534 , -0.2439362053 , -1.020056898 , -1.796177591 , 1.065675489 ,  
.2895547965 , -0.4865658963 , 1.489710750 , .7135900578 , -0.06253063499 ,  
-0.2587858051 , -1.034906498 , -1.811027192 , 1.072550304 , .2964296113 ,  
-0.4796910817 , 1.496585564 , .7204648725 , -0.05565582049 , -0.2519109903 ,  
-1.028031683 , -1.804152376 , 1.086299934 , .3101792405 , -0.4659414521 ,  
1.510335194 , .7342145016 , -0.04190619112 , -0.2381613611 , -1.014282054 ,  
-1.790402746 ]
```

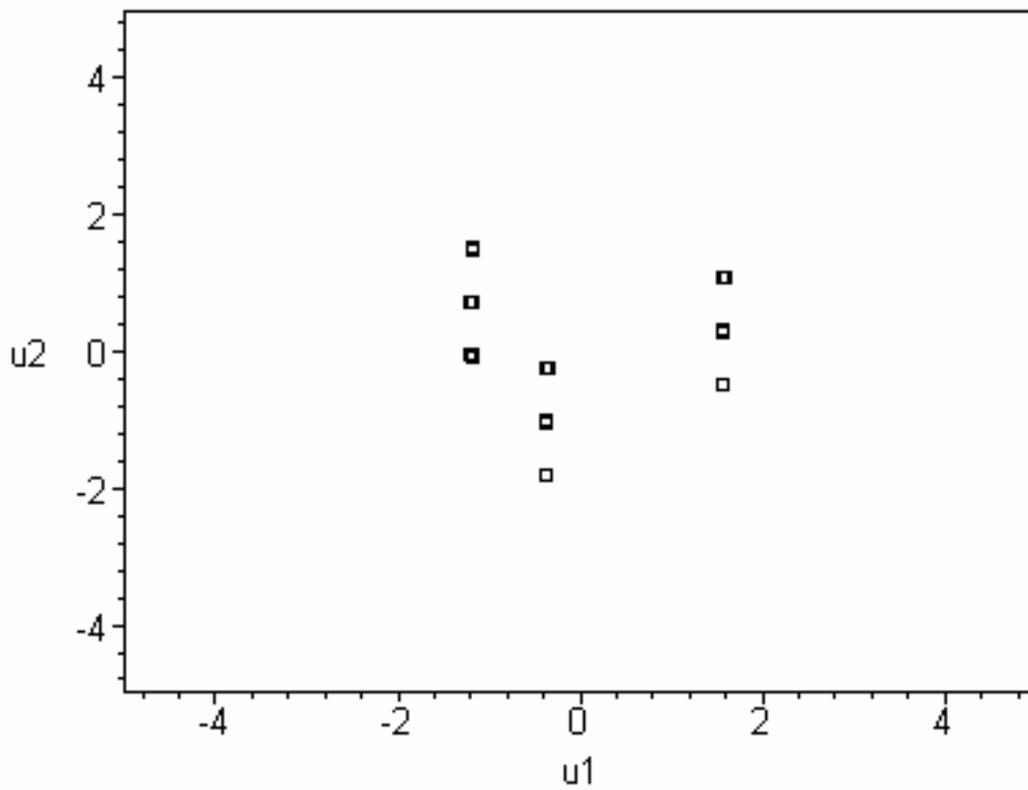
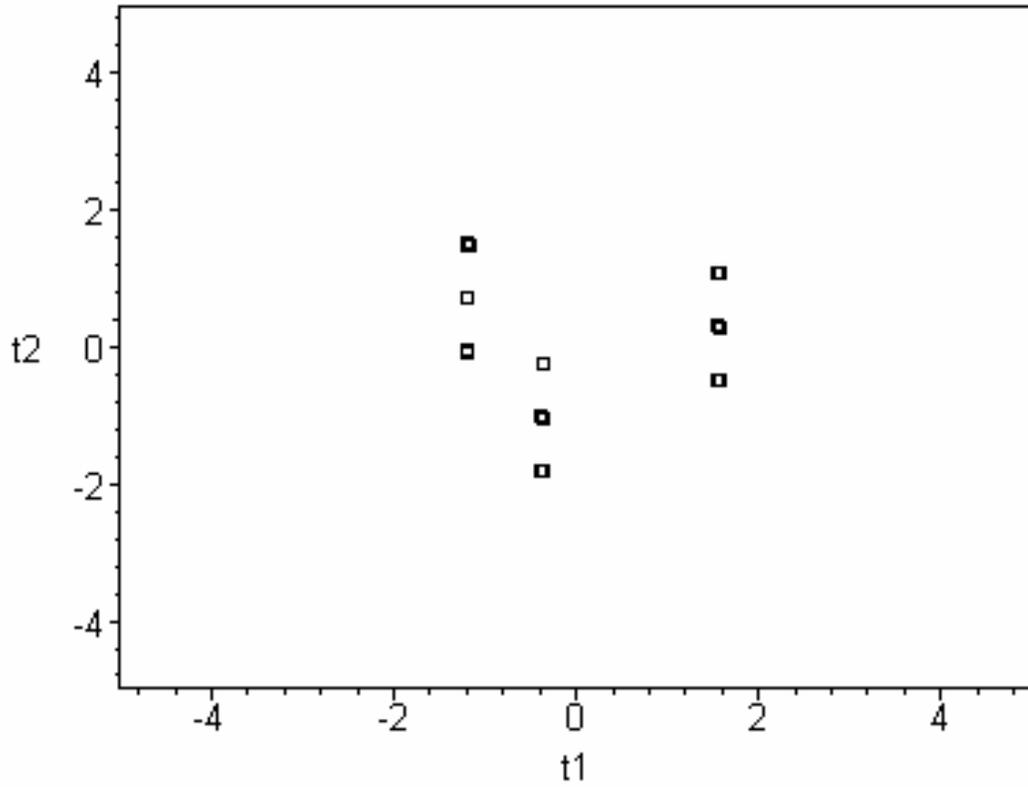
```
ts2 := "t2"
```

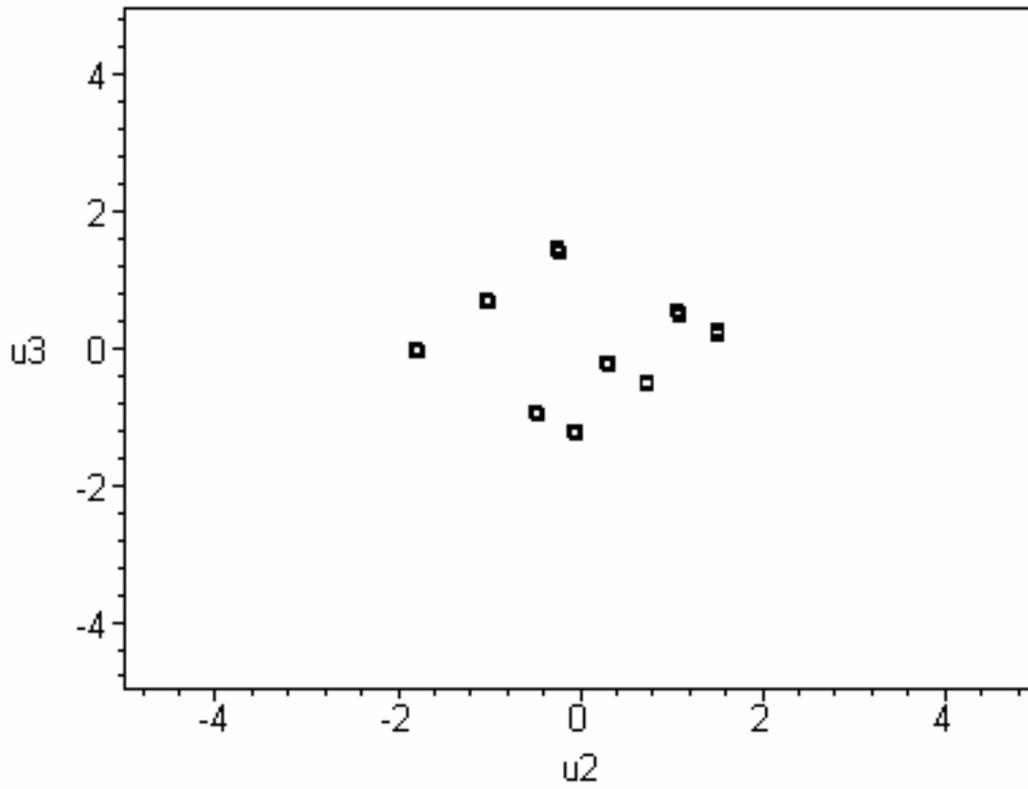
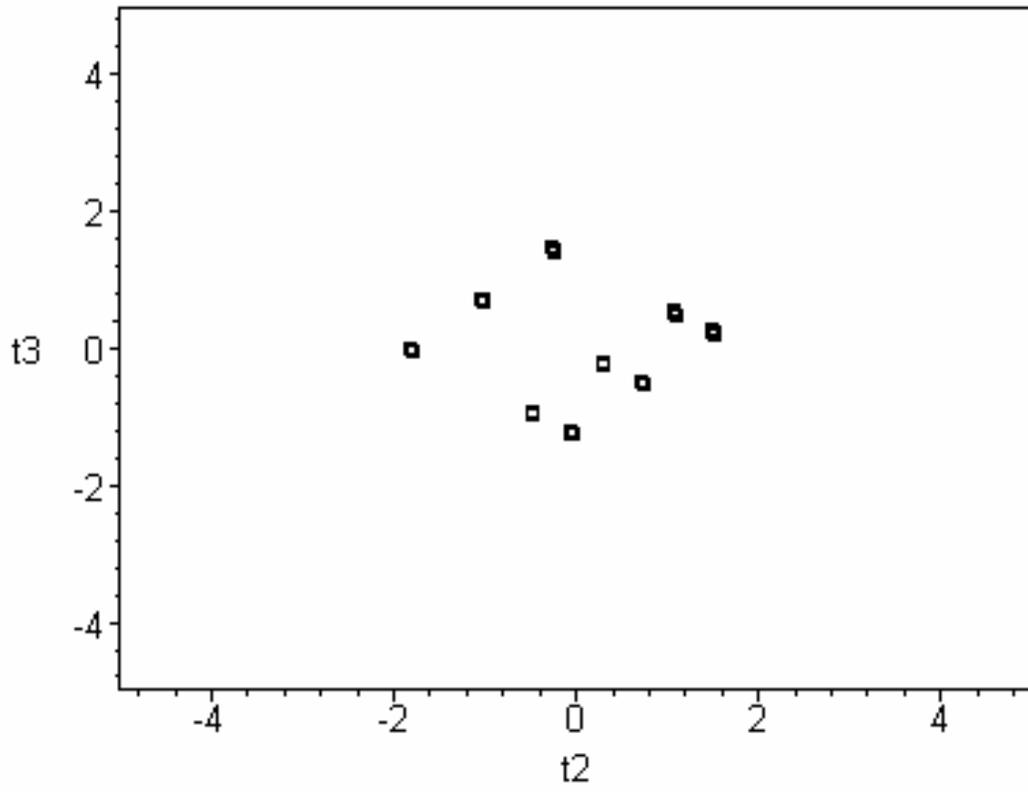
```
set1 := [[ 1.566580883 , 1.080525089 ], [ 1.559912744 , .3044043964 ],  
[ 1.553244606 , -0.4717162964 ], [ -1.187080409 , 1.504560350 ],  
[ -1.193748548 , .7284396576 ], [ -1.200416686 , -0.04768103534 ],  
[ -0.3702594436 , -0.2439362053 ], [ -0.3769275820 , -1.020056898 ],  
[ -0.3835957203 , -1.796177591 ], [ 1.579080298 , 1.065675489 ],  
[ 1.572412159 , .2895547965 ], [ 1.565744021 , -0.4865658963 ],  
[ -1.174580994 , 1.489710750 ], [ -1.181249133 , .7135900578 ],  
[ -1.187917271 , -0.06253063499 ], [ -0.3577600289 , -0.2587858051 ],  
[ -0.3644281673 , -1.034906498 ], [ -0.3710963056 , -1.811027192 ],  
[ 1.573293532 , 1.072550304 ], [ 1.566625393 , .2964296113 ],  
[ 1.559957255 , -0.4796910817 ], [ -1.180367761 , 1.496585564 ],  
[ -1.187035899 , .7204648725 ], [ -1.193704037 , -0.05565582049 ],  
[ -0.3635467949 , -0.2519109903 ], [ -0.3702149333 , -1.028031683 ],  
[ -0.3768830716 , -1.804152376 ], [ 1.561719999 , 1.086299934 ],  
[ 1.555051861 , .3101792405 ], [ 1.548383722 , -0.4659414521 ],  
[ -1.191941293 , 1.510335194 ], [ -1.198609431 , .7342145016 ],  
[ -1.205277570 , -0.04190619112 ], [ -0.3751203271 , -0.2381613611 ],  
[ -0.3817884655 , -1.014282054 ], [ -0.3884566038 , -1.790402746 ]]
```

```

Tplot := PLOT ( CURVES ( [[ 1.56658088300000010 , 1.08052508900000000 ],
[ 1.55991274400000002 , .304404396399999999 ],
[ 1.55324460600000002 , -.471716296400000012 ],
[ -1.18708040900000000 , 1.50456034999999999 ],
[ -1.19374854800000006 , .728439657599999958 ],
[ -1.20041668600000007 , -.0476810353399999970 ],
[ -.370259443600000016 , -.2439362052999999987 ],
[ -.3769275819999999984 , -1.02005689799999999 ],
[ -.3835957202999999998 , -1.796177590999999996 ],
[ 1.57908029800000004 , 1.065675488999999998 ],
[ 1.572412158999999998 , .2895547964999999988 ],
[ 1.565744020999999998 , -.4865658963000000024 ],
[ -1.17458099400000004 , 1.489710749999999998 ],
[ -1.181249133000000012 , .7135900577999999955 ],
[ -1.187917271000000010 , -.06253063499000000004 ],
[ -.3577600289000000026 , -.2587858050999999988 ],
[ -.3644281672999999995 , -1.034906498000000001 ],
[ -.3710963056000000009 , -1.811027192000000006 ],
[ 1.573293532000000010 , 1.072550303999999996 ],
[ 1.566625393000000004 , .2964296113000000006 ],
[ 1.559957255000000004 , -.4796910817000000022 ],
[ -1.180367761000000006 , 1.496585564000000008 ],
[ -1.187035899000000006 , .7204648725000000020 ],
[ -1.193704037000000004 , -.05565582049000000008 ],
[ -.3635467949000000025 , -.2519109903000000026 ],
[ -.3702149332999999993 , -1.028031683000000003 ],
[ -.3768830716000000008 , -1.804152376000000000 ],
[ 1.561719998999999992 , 1.086299933999999992 ],
[ 1.555051860999999992 , .3101792405000000009 ],
[ 1.548383722000000008 , -.4659414520999999986 ],
[ -1.191941292999999994 , 1.510335194000000005 ],
[ -1.198609430999999994 , .7342145016000000014 ],
[ -1.205277570000000002 , -.0419061911199999984 ],
[ -.3751203270999999984 , -.2381613610999999996 ],
[ -.3817884655000000007 , -1.014282053999999993 ],
[ -.3884566038000000022 , -1.790402746000000004 ]], AXESSTYLE (BOX),
STYLE (POINT), COLOUR (RGB, 0, 0, 0), AXESLABELS ("t1", "t2"),
SYMBOL (BOX), VIEW (-5. .. 5., -5. .. 5.))

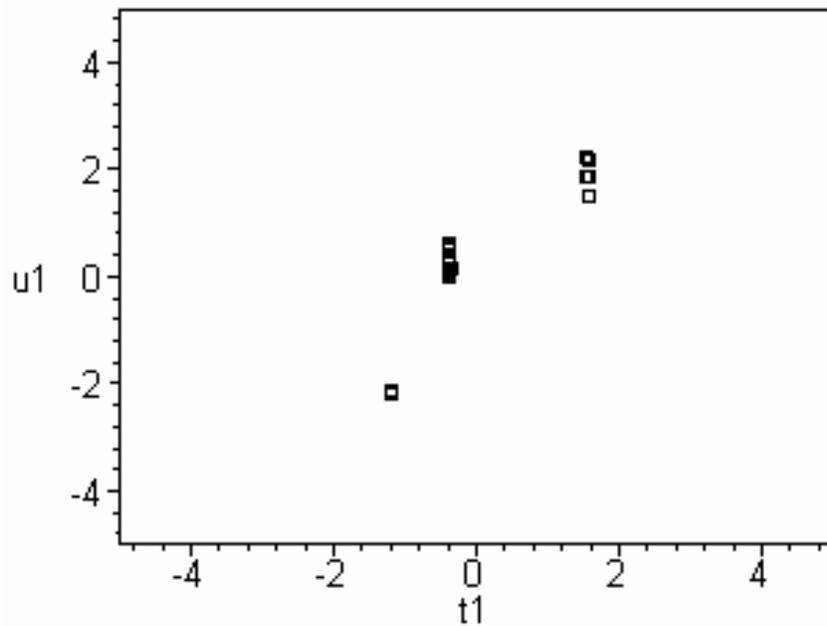
```

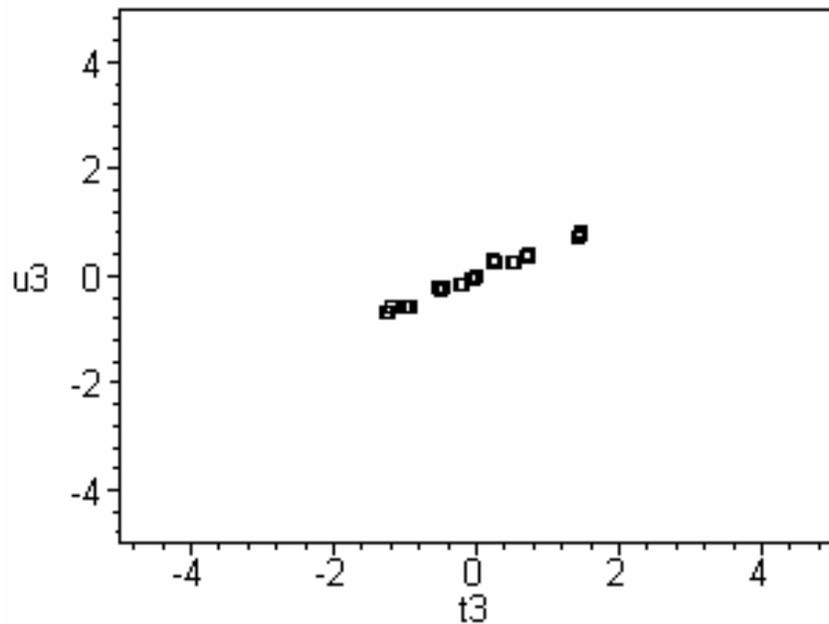
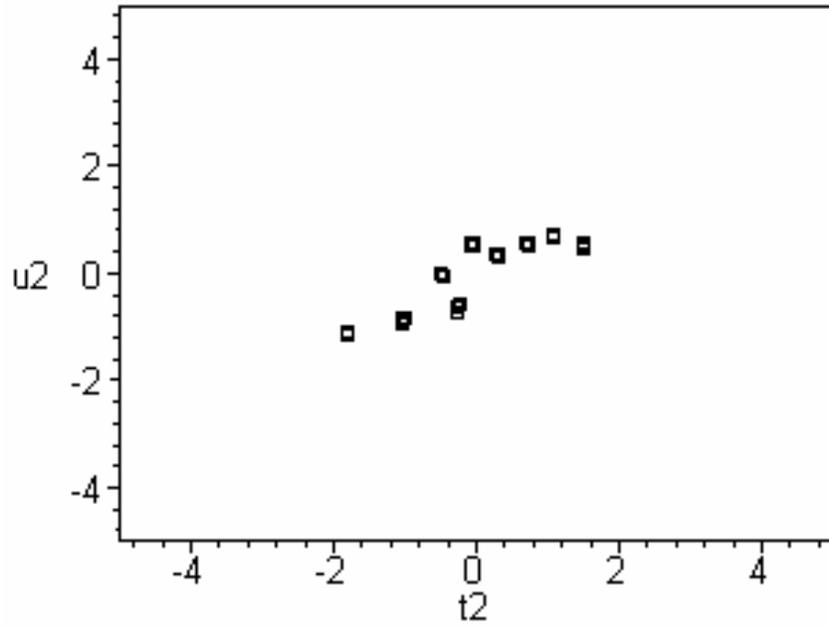




Plots of t versus u gives an idea of how good the regression is. The first plot of t1 versus u1 shows a fair regression, with perhaps one outlier. The data get progressively more scattered at higher variables.

```
> 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. The X matrix are scaled with the mean and variance calculated in the model development.

This section corresponds to one set of conditions input for prediction.

```
> XP:=[300,7,108000,80]:
```

Normalization of the input data.

```
> 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: XP3 := matrix(1, num_ind_var, XP1);
Decomposition of the X matrix.
> ww := matrix(wm); pp := transpose(matrix(pm)); qq := transpose(matrix(qm));
TP := multiply(XP3, ww); TP1 := multiply(TP, pp);
XP2 := matadd(XP3, TP1, 1, -1);
Generation of Y matrix
> UP := multiply(TP, qq);
i := 'i': YP := matrix(1, num_dep_var, 0):
for i from 1 to num_dep_var do
  YP := mulcol(UP, i, bm[i]); UP := YP:
od:

```

Conversion of variables back to physical scales

```

> s := 's': YP1 := matrix(1, num_dep_var, 0):
for i from 1 to num_dep_var do
  s := Columns(Y, [i]);
  s1 := describe[mean](s);
  s2 := describe[standarddeviation[1]](s);
  YP1[1, i] := (evalf(map(x -> s1 + (s2*x), col(YP, i))));
od: print(YP1);

```

$XP3 := [-1.207614729 \quad .3818813079 \quad .5372459074 \quad 3.019036822]$

$ww := \begin{bmatrix} .09752982840 & -.3299119576 & -.9483712268 \\ -.3258776826 & -.9944656984 & .2609099732 \\ -.006678124331 & .005876984829 & -.01852463784 \\ .9855467344 & -.2961388463 & .1799970558 \end{bmatrix}$

$pp := \begin{bmatrix} -.004025648884 & -.5559196457 & -.004868713296 & .8312120291 \\ -.6534447250 & -.7355818879 & .008066545266 & -.1785054869 \\ -.9479292362 & .2607353378 & -.03348573448 & .1797941947 \end{bmatrix}$

$qq := \begin{bmatrix} .5340105602 & .6055465724 & .5900390412 \\ -.6309936810 & -.4955398904 & -.5968979747 \\ .9358950730 & .3103388825 & -.1667038993 \end{bmatrix}$

$TP := [2.729589033 \quad -.8722580176 \quad 1.778369157]$

$TP1 := [-1.126784084 \quad -.4121312858 \quad -.07987569260 \quad 2.744310532]$

$XP2 := [-.080830645 \quad .7940125937 \quad .6171216000 \quad .274726290]$

$UP := [3.672385598 \quad 2.637029022 \quad 1.834752067]$

$[[1.398567829] \quad [.2531196236] \quad [.009008574987]]$

Testing with an array of input data

```

>
XP := [: XP := [[300, 3, 0, 50], [310, 5, 80000, 30], [325, 7, 50000, 10], [340, 9, 13500
0, 25], [350, 4, 75000, 40], [305, 6, 100000, 15], [345, 8, 90000, 35], [320, 10, 30000
, 45], [315, 5.5, 65000, 20], [295, 7, 108000, 80], [295, 4.73, 108000, 80], [295, 5.9
8, 108000, 80], [295, 8.09, 108000, 80], [295, 9.03, 108000, 80], [325, 7, 108000, 80

```

```

],[350,7,108000,80],[295,7,67343,80],[295,7,189743,80],[295,7,108000,20
],[295,7,108000,50],[295,7,108000,67],[295,7.12,108000,80],[295,8.52,10
8000,80],[295,9.58,108000,80]]:
num_pred_samples:=rowdim(array(XP));

```

Changing input values to mean-centred, variance-scaled values using parameters derived from training matrices.

```

> s:='s':XP1:=[]:
for i from 1 to num_ind_var do
s:=Columns(X,[i]);s2:=describe[mean](s);s3:=describe[standarddeviation[
1]](s);XP1:=[op(XP1),(evalf(map(x->(x-
s2)/s3,col(XP,i))))];od:xx:=transpose(matrix(XP1)):

```

num_pred_samples := 24

Iterative decomposition of X matrix and building up of Y matrix for predictions. Run through for each dependent factor (type of organic).

```

> i:='i':F:=matrix(num_pred_samples,num_dep_var,0):
TP:=evalm(xx*ww);
TP1:=evalm(TP*pp);
XP2:=matadd(xx,TP1,1,-1);
F1:=evalm(TP*qq);
for i from 1 to num_dep_var do
  YP:=mulcol(F1,i,bm[i]): F1:=YP:
od:
print(F1):

```

1.470837681	.4028401812	-.01248004882
.2602830067	-.05837183954	-.2378322098
-1.095984630	-.5152344317	-.4042691079
.07500003260	.07904757771	.2644706253
-1.054926564	.06881742313	.3628125224
-.2817549795	-.4167449027	-.5784301763
.2610952453	.2884492788	.4861111105
2.836055708	.8380489897	.6058090293
-.4978945846	-.3267302941	-.3855257988
5.327184516	1.676676496	.9469024008
4.484641906	1.470551000	.7776258404
4.948597086	1.584056229	.8708398048
5.731753431	1.775653055	1.028184978
6.080647727	1.861008987	1.098281880
4.060724922	1.610977302	1.273177050
3.005341931	1.556227975	1.545072590
5.351343509	1.682331254	.9485790039
5.278611608	1.665307313	.9435315014
.8776067873	-.1527051906	-.4929586978
3.102395652	.7619856533	.2269718515
4.363109342	1.280310465	.6349324966
5.371724212	1.687572998	.9558509408
5.891354012	1.814698853	1.060250582
6.284788006	1.910951288	1.139296023

Regenerate "real" concentrations.

```
> s:='s':i:='i':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),(map(x->s1+(s2*x),col(F1,i)))]];
od:
fraction_in_water:=transpose(matrix(Yp1));
```

	.9600725984	.1678696992	.005778960758
	.8144530922	.1367360293	.005061039922
	.6513055470	.1058959633	.004530809741
	.7921651149	.1460123937	.006661262746
	.6562444857	.1453218171	.006974557709
	.7492504991	.1125444034	.003975972106
	.8145507977	.1601478524	.007367359026
	1.124296800	.1972480476	.007748689427
	.7232507296	.1186207522	.004590521669
	1.423958562	.2538587836	.008835335491
	1.322607800	.2399444820	.008296058790
	1.378417691	.2476065424	.008593017324
<i>fraction_in_water</i> :=	1.472624786	.2605401002	.009094283333
	1.514593824	.2663019696	.009317596153
	1.271614166	.2494238233	.009874772468
	1.144660503	.2457280231	.01074096995
	1.426864685	.2542405026	.008840676768
	1.418115651	.2530913180	.008824596570
	.8887119307	.1303681480	.004248264837
	1.156335246	.1921134659	.006541800164
	1.307988459	.2271024793	.007841470184
	1.429316311	.2545943414	.008863843508
	1.491823389	.2631758489	.009196437069
	1.539150176	.2696732762	.009448257903

INTERNAL DISTRIBUTION

1. R.R. Judkins
2. J.A. Zollar
3. D.W. DePaoli
4. C. Tsouris
5. M.J. Sale
6. S.M. Robinson
7. OTIC-RC, OSTI, CRL

EXTERNAL DISTRIBUTION

D. Alleman
National Energy Technology Laboratory
Environmental Research and Analysis,
One West Third, Ste. 1400
Tulsa, OK 74103,
(918) 699-2057
david.alleman@netl.doe.gov

J. Garcia
National Energy Technology Laboratory
Environmental Research and Analysis,
One West Third, Ste. 1400
Tulsa, OK 74103
(918) 699-2036
jesse.garcia@netl.doe.gov

D.K. Schmalzer
ANL, 955 L'Enfant Plaza SW, Suite
6000, Washington, DC 20024
(202) 488-2415
schmalzer@anl.gov

Veronica Blackwell,
Lead Environmental Engineer
ChevronTexaco Energy Research and
Technology Company
100 Chevron Way, Richmond, CA
94802
vblackwell@chevrontexaco.com

Sara McMillen
ChevronTexaco Energy Research and
Technology Company
100 Chevron Way
Richmond CA 94802
(510) 242-3485
SAJM@ChevronTexaco.com

Zara I Khatib,
Tel: 971 50454096
zara.khatib@shell.com

Ileana Rhodes
Shell Global Solutions
Ileana.Rhodes@shell.com

Sung-I Johnson
Manager, Analytical Services
248PL Research Center
ConocoPhillips
Bartlesville, Oklahoma 74004
(918) 661-1508
sung-i.johnson@conocophillips.com

Carolyn Johlman
Phillips Petroleum Company
Research & Development
Bartlesville, OK 74004
(918) 661-7754
carolyn.johlman@conocophillips.com