

Models for Conductance Measurements in Quality Assurance of Water Analysis

Stephen G. Hughes, Elizabeth L. Taylor, and Peter D. Wentzell*

Trace Analysis Research Centre, Department of Chemistry, Dalhousie University,
Halifax, Nova Scotia, B3H 4J3, Canada

Ross F. McCurdy and Robert K. Boss

Fenwick Laboratories Ltd., Suite 200, 5595 Fenwick Street, Halifax, Nova Scotia, B3H 4M2, Canada

Models for the prediction of conductance in nonbrine water samples through the measurement of ionic concentrations and other parameters are compared. Such predictions are often used for quality assurance purposes by comparing them with actual measurements to determine whether gross analysis errors have been made. A currently recommended method for making such predictions is a semiempirical relation that adapts the Debye–Hückel–Onsager equation to mixed electrolyte systems by incorporating modified definitions of ionic charge and concentration. The limitations of this model are examined, and extensions to it are considered. Other predictive methods, including multiple linear regression (MLR), principal components regression (PCR), partial least-squares (PLS) regression, continuum regression (CR), and neural networks (NN), are also considered. Models employ the concentrations of 10 ions as well as, in some cases, additional water quality measurements. Best results were obtained with an extended form of the Debye–Hückel–Onsager equation and an optimized MLR model. PCR, PLS, CR, and NN did not offer significant advantages.

Perhaps no resource is as important to a region as its water, and this is increasingly the case. The analysis of nonbrine waters is critical in a number of areas including public health (e.g., the assessment of potability), industrial monitoring (e.g., boiler or cooling water), and pollution control (e.g., wastewater management). Procedures for making chemical and physical measurements on such samples are well-established,^{1–5} and an enormous number of routine analyses are conducted worldwide each day. Because of the importance of the results of such analyses and the large number that are carried out, quality assurance becomes an especially important concern. Various methods are commonly used for this purpose, including the use of control samples. While these can test the reliability of the analytical methods employed, it is still possible to obtain erroneous results for individual samples due to errors in measurement, data entry, or other steps in the procedure.

One way to circumvent this problem is to run replicates for each sample, but this can be time-consuming and costly. In the analysis of inorganic ions, another method that can be used is measurement of the conductivity of the sample and comparison of this result with that predicted based on the concentrations of major ions. Additionally, this approach can indicate the presence of major ions not included in the model. While this technique has limitations in identifying erroneous results for all ions, it can serve as a part of the overall quality control program.

One of the recommended approaches to calculating the conductivity in nonbrine water samples is a semiempirical method that will be referred to as the Rossum model (ref 1, pp 1–12). Details of this model were reported in 1975,⁶ and despite its acceptance as a standard procedure in 1989 and its obvious practical utility, very little has been written concerning its validity or general applicability, not to mention possible alternative models. Furthermore, the initial report was limited to samples with conductivities not exceeding about 1400 $\mu\text{S cm}^{-1}$. The goal of the work presented here was to conduct a more rigorous evaluation of the Rossum model through a comparison with a range of other models. Included in this comparison were empirical models based on multiple linear regression (MLR), principal components regression (PCR), partial least squares (PLS), continuum regression (CR), and neural networks (NN), as well as semiempirical models based on extensions to the original Rossum model. The data employed were obtained for a wide range of samples processed by a commercial laboratory.

BACKGROUND

Rossum Model. The theoretical basis for predicting conductance in dilute solutions of simple electrolytes has been established for many years and is well-treated in several texts.^{7–11} It was Kohlrausch who first correctly observed that, in dilute aqueous solutions, the equivalent conductance, Λ , varies with the square root of the concentration of a strong

- (1) Greenberg, A. E.; Clesceri, L. S.; Eaton, A. D., Eds. *Standard Methods for the Examination of Water and Wastewater*, 18th ed.; American Public Health Association: Washington, DC, 1992.
- (2) United States Environmental Protection Agency. *Methods for the Chemical Analysis of Water and Wastes*; U.S. Government Printing Office: Washington, DC, 1979.
- (3) American Society for Testing and Materials. *Manual on Water*; The Society: Philadelphia, PA, 1978.
- (4) Hunt, D. T. E. *The Chemical Analysis of Water: General Principles and Techniques*, 2nd ed.; Royal Society of Chemistry: London, 1986.
- (5) Fresenius, W. *Water Analysis*; Springer-Verlag: New York, 1988.

- (6) Rossum, J. R., Jr. *Am. Water Works Assoc.* **1975**, *67*, 204–205.
- (7) Harned, H. S.; Owen, B. B. *The Physical Chemistry of Electrolytic Solutions*, 3rd ed.; Reinhold: New York, 1958.
- (8) Fuoss, R. M.; Accascina, F. *Electrolytic Conductance*; Interscience Publishers: New York, 1959.
- (9) Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*, 2nd ed.; Butterworth: London, 1959.
- (10) MacInnes, D. A. *The Principles of Electrochemistry*; Reinhold: New York, 1961.
- (11) Atkins, P. W. *Physical Chemistry*, 4th ed.; Freeman: New York, 1990; p 760.

Table 1. List of Symbols and Their Units

c, c_+, c_-	equivalent concentration of an ion (mM)
C	equivalent concentration of a salt = (molar concentration)/ $\sum z_+$ or $\sum z_-$ (M)
C'	composite equivalent concentration for a mixed electrolyte solution (Rossum model) (mM)
D	dielectric constant (relative permittivity) of solution (no units)
e	proton charge (C)
ϵ_0	vacuum permittivity ($J^{-1} C^2 m^{-1}$)
F	Faraday's constant ($C mol^{-1}$)
G	solution conductivity ($\mu S cm^{-1}$)
G_+, G_-	composite ionic conductivities in a mixed electrolyte solution (Rossum model) ($\mu S cm^{-1}$)
η	solution viscosity (P)
$\lambda, \lambda_+, \lambda_-$	equivalent conductance of an ion ($S cm^2 mol^{-1}$)
$\lambda^\circ, \lambda_+^\circ, \lambda_-^\circ$	limiting equivalent conductance of an ion ($S cm^2 mol^{-1}$)
λ_+', λ_-'	composite limiting equivalent conductance of ions in a mixed electrolyte solution (Rossum model) ($S cm^2 mol^{-1}$)
Λ	equivalent conductance of a solution ($S cm^2 mol^{-1}$)
Λ°	limiting equivalent conductance of a solution ($S cm^2 mol^{-1}$)
R	gas constant ($J mol^{-1} K^{-1}$)
z, z_+, z_-	absolute value of ionic charge (no units)
Z_+, Z_-	composite ionic charges for a mixed electrolyte solution (Rossum model) (no units)

electrolyte. He also demonstrated that the limiting equivalent conductance of a salt, Λ° , could be expressed as the sum of the limiting equivalent conductances of the individual ions, λ_+° and λ_-° . The theoretical basis for the concentration dependence was derived later and resulted in the Debye-Hückel-Onsager equation for a single electrolyte in dilute solution:

$$\Lambda = \Lambda^\circ - (A + B\lambda^\circ)C^{1/2} \quad (1)$$

where C is the equivalent concentration of the electrolyte and A and B are constants which depend on solution temperature, dielectric constant, and viscosity, as well as other physical constants. The constant A accounts for relaxation of the counterion cloud around the moving ion, and B compensates for the electrophoretic effect as the ion moves through solution. For 1:1 electrolytes in aqueous solution at 25 °C, $A = 60.20 S cm^2 mol^{-1} M^{-1/2}$ and $B = 0.229 M^{-1/2}$.¹¹ A more general form of the equation, adopted by Shedlovsky,¹² gives the equivalent conductance of a single ion formed from a binary salt in dilute solution as

$$\lambda = \lambda^\circ - \left[\frac{\alpha}{(DT)^{3/2}} w \lambda^\circ + \frac{\beta z}{(DT)^{1/2} \eta} \right] ((z_+ + z_-)C)^{1/2} \quad (2)$$

where

$$w = z_+ z_- \frac{2q}{1 + q^{1/2}} \quad q = \frac{z_+ z_- (\lambda_+^\circ + \lambda_-^\circ)}{(z_+ + z_-)(z_+ \lambda_-^\circ + z_- \lambda_+^\circ)}$$

$$\alpha = \frac{(10^3)^{1/2} e F^2}{24\pi(\epsilon_0 R)^{3/2}} \quad \beta = \frac{(10^3)^{1/2} e F^2}{6\pi(\epsilon_0 R)^{1/2}}$$

and the other symbols are defined in Table 1. Note that conversion factors have been included with α and β that are valid if the units used are as given in Table 1. Shedlovsky gives the constants as $\alpha = 9.838 \times 10^5 M^{-1/2} K^{3/2}$ and $\beta =$

$28.95 S cm^2 mol^{-1} M^{-1/2} K^{1/2}$ (note the latter value will be 2.895 when SI units of $kg m^{-1} s^{-1}$ are used for viscosity). To calculate the conductivity, G , of a solution consisting of a single salt, the equivalent conductance of the anion and cation are added and multiplied by the equivalent concentration. This equation is reasonably accurate for the prediction of conductance in relatively dilute samples, especially for 1:1 electrolytes. For example, on the basis of data from Robinson and Stokes,⁹ 5 mM solutions of NaCl, BaCl₂, and LaCl₃ give errors of -0.4%, -1.5%, and -5.3%, respectively. Errors arise from the truncation of higher-order terms in the derivation of the Debye-Hückel-Onsager equation and, at high concentrations, from ion association.

The prediction of conductance measurements in water samples is complicated by the fact that they are multielectrolyte systems. The problem of nonbinary electrolytes has been addressed,^{7,13} but the solution is algebraically complex. Rossum⁶ treated the problem of multielectrolyte systems semiempirically by considering them as binary salts in which parameters used in the Debye-Hückel-Onsager equation are essentially "average" or composite values. His equation takes the same form as eq 2 but is presented in terms of conductivity (in $\mu S cm^{-1}$):

$$G = G_+^\circ + G_-^\circ - \left[\frac{W\lambda^\circ}{K_1(Z_+ + Z_-)} + K_2 \right] [(Z_+ + Z_-)C]^{3/2} \quad (3)$$

where

$$W = Z_+ Z_- \frac{2Q}{1 + Q^{1/2}} \quad Q = \frac{Z_+ Z_- \Lambda^\circ}{(Z_+ + Z_-)(Z_+ \lambda_-' + Z_- \lambda_+')}$$

$$K_1 = \frac{24\pi(\epsilon_0 DRT)^{3/2}}{eF^2} \quad K_2 = \frac{10^5 e F^2}{6\pi(\epsilon_0 DRT)^{1/2} \eta}$$

The constants K_1 and K_2 are derived from those in eq 2 with adjustments for dimensions. Rossum gave values of $K_1 = 115.2 mM^{1/2}$ and $K_2 = 0.668 S cm^2 mol^{-1} mM^{-1/2}$ (or $\mu S cm^{-1} mM^{-3/2}$) at 25 °C. The composite equivalent concentration, C' , is in millimolar, and G will be in microsiemens per centimeter. The composite parameters are given as

$$G_+^\circ = \sum c_+ \lambda_+^\circ \quad G_-^\circ = \sum c_- \lambda_-^\circ$$

$$Z_+ = \sum c_+ z_+^2 / \sum c_+ z_+ \quad Z_- = \sum c_- z_-^2 / \sum c_- z_-$$

$$\lambda_+' = G_+^\circ / \sum c_+ \quad \lambda_-' = G_-^\circ / \sum c_-$$

$$\Lambda^\circ = \lambda_+' + \lambda_-' \quad C' = (\sum c_+ + \sum c_-) / 2$$

The summations are over all positive or negative ions. Rossum did not provide justification for these definitions other than that they appear logical and seem to work for real samples. The absence of a rigorous theoretical foundation, however, makes the recommendation of this approach as a standard method somewhat questionable. Furthermore, the Debye-Hückel-Onsager equation is known to deviate as the ionic concentration increases.^{8-10,12} Rossum limited his tests to samples below ca. 1400 $\mu S cm^{-1}$. Additional terms have been used by other workers to extend the range of the original Debye-Hückel-Onsager model and the use of these for the

(12) Shedlovsky, T. *J. Am. Chem. Soc.* **1932**, *54*, 1405-1411.

(13) Onsager, L.; Fuoss, R. M. *J. Phys. Chem.* **1932**, *36*, 2689-2778.

Table 2. Statistical Summary of Water Quality Parameters

parameter	units	mean value	min value	max value	SD	median value	percentile 25th-75th
Measured Quantities							
sodium	mM	2.09	0.01	71.7	6.17	0.66	0.24-1.71
potassium	mM	0.36	0.00	22.5	1.55	0.06	0.03-0.18
calcium	mM	2.17	0.00	17.5	2.42	1.66	0.61-2.74
magnesium	mM	0.70	0.00	8.52	0.86	0.55	0.15-0.93
sulfate	mM	0.82	0.01	16.2	1.97	0.23	0.09-0.52
chloride	mM	1.94	0.00	90.3	7.42	0.62	0.20-1.43
nitrate + nitrite	mM	0.39	0.00	7.43	1.18	0.01	0.00-0.05
ammonia	mM	0.20	0.00	14.6	1.03	0.00	0.00-0.01
orthophosphate	mM	0.01	0.00	1.19	0.08	0.00	0.00-0.00
reactive silica	mM	0.37	0.00	1.80	0.25	0.36	0.18-0.49
iron	mM	0.05	0.00	1.42	0.21	0.00	0.00-0.01
manganese	mM	0.01	0.00	0.29	0.02	0.00	0.00-0.00
copper	mM	0.00	0.00	0.02	0.00	0.00	0.00-0.00
zinc	mM	0.01	0.00	2.49	0.12	0.00	0.00-0.00
pH		7.54	4.3	12.0	0.7	7.6	7.2-8.0
alkalinity	mg/L CaCO ₃	214	0	2500	228	200	70-280
TOC	mg/L C	11.6	0.5	850	64.6	1.7	0.8-4.9
conductivity	μS cm ⁻¹	801	4.1	10400	1070	528	288-956
Calculated Quantities							
carbonate	mM	0.02	0.00	0.76	0.05	0.01	0.00-0.02
bicarbonate	mM	4.24	0.00	50	4.56	3.94	0.67-2.73
hardness	mg/L CaCO ₃	287	0.0	2100	301	236	79.4-360
Langelier index		-0.65	-9.15	4.47	1.73	0.03	-1.17-0.38
saturated pH		8.18	6.04	15.8	1.26	7.75	7.39-8.61

Rossum model is investigated here, along with other empirical models.

Other Models. Given the need to predict conductance based on ionic concentration of water samples and the recommendation of the Rossum model as a standard method, it was felt that a more rigorous examination of alternative models should be investigated. These models included extensions and modifications to the original Rossum model, as well as several empirical calibration methods that have appeared in the chemometrics literature. The database used for this investigation included 23 parameters, which are listed in Table 2. Some of these parameters are measured directly, while others are derived. Not all of the parameters measured were found to be useful. For methods based on the Rossum model, only the ionic concentrations were used. For other models, additional parameters were also investigated. The following is a list of the models tested in this work. For ease of reference, each of the models has been assigned an acronym.

RM: This is the original Rossum model.

MRM: The modified Rossum model uses an alternative method for the calculation of the composite charge. Specifically,

$$Z_+ = \sum c_+ z_+ / \sum c_+ \quad (4)$$

and likewise for Z_- . Rossum claimed that his definition of the composite charge provided better results. This conclusion was examined.

ERM: The extended Rossum model includes an additional term in the composite equivalent ionic concentration, C' . The parameter associated with this term is determined empirically.

EMRM: The extended modified Rossum model incorporates the ERM and MRM.

MLR: This represents a class of models generated empirically using multiple linear regression. MLRa uses the ionic concentrations directly, MLRb uses an optimized set of

ionic concentrations, and MLRc employs variables beyond the simple ionic concentrations.

PCR: This represents the best model obtained by principal components regression.

PLS: This is the best model obtained by partial least squares.

CR: This is the best model determined by continuum regression.

NN: This is the best model obtained using back-propagation neural networks.

The principles of principal components regression, partial least squares, continuum regression, and neural networks as they apply to calibration problems in chemistry have been described elsewhere and will not be discussed here.

EXPERIMENTAL SECTION

Sample Parameters. All of the water samples used in this work were analyzed at Fenwick Laboratories Ltd. (Halifax, NS) using their Rapid Chemical Analysis package (RCAP). Methods used were based on standard procedures for water analysis^{1,2} and will only be summarized here. Metallic ions (Na, K, Ca, Mg, Fe, Mn, Cu, Zn) were determined by flame atomic absorption on a Varian SpectraAA 40 spectrometer (Varian Canada, Georgetown, ON). Colorimetric methods were used for chloride, ammonia, nitrate + nitrite, orthophosphate, reactive silica, and alkalinity. These were carried out on an automated COBAS FARA centrifugal analyzer (Roche Diagnostics, Mississauga, ON). Chloride was determined using mercuric thiocyanate in the presence of Fe^{2+} . The method for ammonia utilized the formation of the blue indophenol by reaction with hypochlorite and alkaline phenol (phenate method). Nitrate was reduced to nitrite with hydrazine sulfate and total nitrate plus nitrite was then determined by reaction with sulfanilamide and *N*-(1-naphthyl)ethylenediamine. Orthophosphate and reactive silica

were determined by reaction with ammonium molybdate and ascorbic acid (heteropoly blue method). Selectivity for phosphate was achieved kinetically, while, for silica, the phosphate interference was removed through the addition of oxalic acid. Alkalinity, which is "the sum of all titratable bases",¹ was measured colorimetrically using an automated methyl orange procedure. In a typical water sample, alkalinity depends predominantly on the concentration of carbonate, bicarbonate, and hydroxide ions, so the first two can be estimated indirectly from the pH and alkalinity,¹ as was done here. Sulfate was determined by a turbidimetric method on a nephelometer (Turner Designs Model 40, Mountain View, CA) using BaCl_2 to precipitate BaSO_4 . Total organic carbon (TOC) was determined using the persulfate ultraviolet oxidation method implemented on a Technicon autoanalyzer (Technicon, Tarrytown, NY) with a phenolphthalein indicator. A Radiometer GK2401C glass electrode (Radiometer America, Westlake, OH) was used to measure pH and a flow-through conductance electrode (Radiometer Model CDC314) was used to measure conductivity, which was corrected to 25 °C.

The data used in this study were obtained from 483 typical samples and represented a variety of sources. The bulk of samples analyzed by Fenwick Laboratories are groundwater samples (ca. 70%), with surface water (ca. 25%) and wastewater (ca. 5%) making up the majority of the rest. Conductivities ranged from about 4 to 11 000 $\mu\text{S cm}^{-1}$. For the purposes of this study, a subset of 422 samples with conductivities below 1400 $\mu\text{S cm}^{-1}$ was also considered. The samples were analyzed during the period August 1991 to March 1992.

Software. Most calculations were performed on a 486-compatible computer under DOS 6.0 using the Excel 4.0 spreadsheet program (Microsoft, Redmond, WA) and programs written in MATLAB 4.0 (The Math Works Inc., Natick, MA) for DOS and Microsoft Windows. Some of the MATLAB programs were written by the authors and others were provided by Dr. Barry Wise.¹⁴ Programs for the artificial neural networks were written in MATLAB (486 PC) and C (Sun SparcStation 10/41).

RESULTS AND DISCUSSION

Rossum Models. A statistical summary of most of the RCap parameters is given in Table 2. Because the distributions of measurements are generally not Gaussian, the median and upper and lower quartile ranges have been given in addition to the mean and standard deviation. For the Rossum model and its variations, the first seven measured ionic concentrations were used, along with ammonium, carbonate, and bicarbonate concentrations. It was assumed that there was no nitrite present (as is normally the case), only nitrate. The concentration of NH_4^+ was calculated from the total ammonia concentration and the pH. For the samples analyzed, the bulk of the total ammonia was generally present as the ammonium ion. The carbonate and bicarbonate were calculated from the alkalinity and pH, with the amount of carbonate generally being negligible. Unfortunately, the pH was reported to only one decimal place for these calculations,

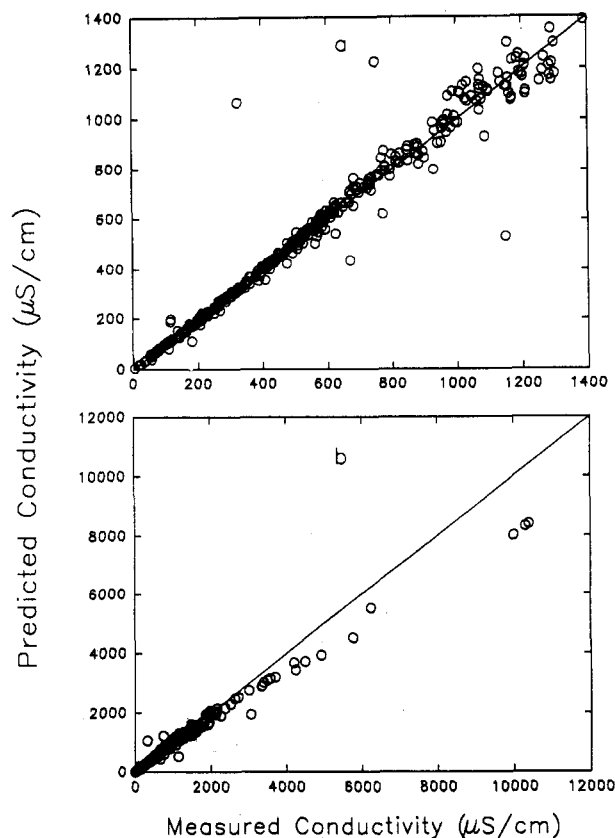


Figure 1. Comparison of measured and predicted conductivities for test data using the Rossum model over two ranges: (a) <1400 $\mu\text{S cm}^{-1}$; (b) all samples.

but this did not have a significant effect as one species normally dominated for both equilibria.

The ions employed for the Rossum model in this work were the same as those employed in the original ref 6, with the exception of NH_4^+ . It was found that the inclusion of the ammonium ion made a significant difference to model performance. On the other hand, the inclusion of phosphate, silicate, iron, manganese, copper, and zinc did not significantly alter the results. These ions were excluded because of their low concentrations and the fact that information on speciation was unavailable. Hydroxide and hydronium ions were also excluded because their effect was generally negligible for the pH range used. The constants used in this work were the same as those employed by Rossum⁶ with the exception of the limiting equivalent conductance of carbonate, for which a more recent value was obtained ($\lambda^\circ = 69.3$).¹⁵

The performance of the Rossum model over two ranges of conductance is illustrated in Figure 1, in which the predicted vs observed conductivities are plotted. The straight line represents the ideal case of perfect agreement. It is evident that prediction errors become larger with conductivity and that there is a systematic variation that produces significant negative deviations of predicted values at higher conductivities, although the agreement below 1400 $\mu\text{S cm}^{-1}$ (the region originally used by Rossum) is quite good. The performance of the Rossum model (RM) is summarized in Table 3, which gives the standard error of prediction (SEP) over two different

(14) Wise, B. M. *PLS Toolbox*, v 1.2, Wise Software, 1991.

(15) Kaye, G. W. C., Laby, T. H., Eds. *Tables of Chemical and Physical Constants*; Longman: London, 1973.

Table 3. Standard Errors of Prediction for Models Employed
std error of prediction

model	<1400 $\mu\text{S cm}^{-1}$	all samples
RM	64	212
MRM	67	148
ERM	67	109
EMRM	67	95
MLRa	76	99
MLRb	68	89
MLRc	61	88
PCR	69	93
PLS	69	92
CR	69	92
NN	79	283

ranges. This is calculated by eq 5,

$$\text{SEP} = \left(\frac{\sum (G_i - \hat{G}_i)^2}{N_{\text{pred}}} \right)^{1/2} \quad (5)$$

where G_i is the measured conductivity, \hat{G}_i is the predicted conductivity, and N_{pred} is the number of measurements predicted. The summation is made over all of the samples that were used for prediction. The prediction errors for other models are also shown in the table and are discussed in greater detail below.

Several modifications to the original Rossum model were employed to determine whether they resulted in improved performance. One of these, the modified Rossum model (MRM), used a modified definition for the composite charge as given in eq 4. Rossum rejected this definition, claiming that it produced conductivity estimates that were "far too low".⁶ We did not find this to be the case and, as shown in Table 3, obtained similar prediction errors below 1400 $\mu\text{S cm}^{-1}$ and substantially improved results when data for samples of higher conductivity were included. An F -test indicated that the improvement when all samples were included was significant (>99% confidence). The systematic deviation was still apparent, however.

As ionic concentrations become high, it is known that the Debye-Hückel-Onsager equation, on which the Rossum model is based, deviates from experimental results. Its range can be extended by the inclusion of terms that account for higher-order effects.⁸⁻¹⁰ Shedlovsky¹² showed that the range could be extended significantly through a slight modification:

$$\Delta = \Delta^\circ - (A + B\lambda^\circ)C^{1/2} + mC(1 - BC^{1/2}) \quad (6)$$

where m is an adjustable parameter and the other symbols have their previous meaning. In an analogous fashion, an additional term was added to the Rossum model as shown below:

$$G = G_+^\circ + G_-^\circ - \left[\frac{W\lambda^\circ}{K_1(Z_+ + Z_-)} + K_2 \right] [(Z_+ + Z_-)C]^{3/2} + m(C)^2 \left[1 - \frac{W[(Z_+ + Z_-)C]^{1/2}}{K_1} \right] \quad (7)$$

Note that a factor of C' has been included for the conversion from equivalent conductance to conductivity. We call this equation the extended Rossum model (ERM) and it represents the first of the models employed with adjustable parameters. To estimate the SEP with this and other models that were fit

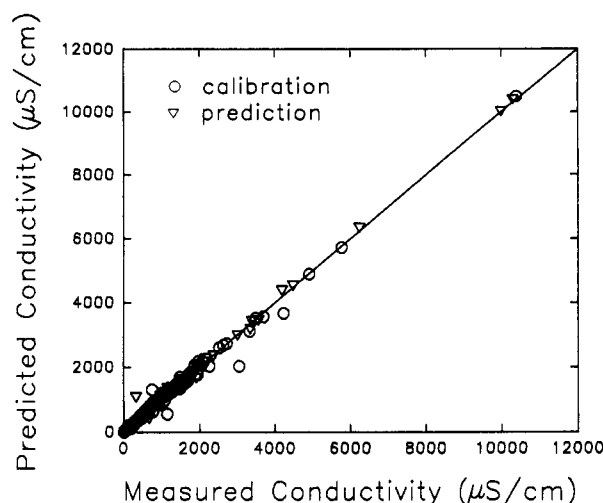


Figure 2. Comparison of measured and predicted conductivities using the extended modified Rossum model.

to the data, a calibration set of 200 samples was first selected at random. These were used to fit the model parameters, and the SEP was calculated with the remaining samples (222 below 1400 $\mu\text{S cm}^{-1}$, 283 when all samples were used). To ensure an unbiased estimate of the true SEP, this procedure was repeated with 50 different calibration/prediction sets and the root mean square of the SEPs was reported. Table 3 shows again that there was a statistically significant improvement (>99%) of the ERM over the original model at high conductivities. Also, as might be expected, the addition of an adjustable parameter significantly reduced systematic deviations in the model. Other extensions to the original equation have also been used, but were not employed here.

The combination of the two model modifications above to give the extended modified Rossum model (EMRM) resulted in the best performance of all of the Rossum model variations, as indicated in Table 3. This is also illustrated in Figure 2, which shows a typical set of results. There is generally excellent agreement between predicted and measured conductivities and a substantial improvement over the results shown in Figure 1. The improvement of the ERM over the other three models was statistically significant when the high conductivity samples were considered. The parameter m was found to be 0.142 when all samples were included.

Other Models. The models developed above are considered to be semiempirical since they have the same form as the Debye-Hückel-Onsager equation but are not rigorously derived. An attempt was also made to model the data with a variety of empirical modeling methods to determine whether these exhibited superior predictive abilities. These included several MLR models in which the conductivity was represented as a linear combination of the measurements. The first such model, MLRa, used the same 10 concentrations as were used for the Rossum models, and had the form

$$\hat{G}_i = \sum_{j=1}^{10} a_j c_{ij} \quad (8)$$

The data were autoscaled prior to solution of this equation to avoid the calculation of an intercept. In addition to first-order terms in concentrations, models that employed $c^{3/2}$ were also used (analogous to the Rossum model), but exhibited

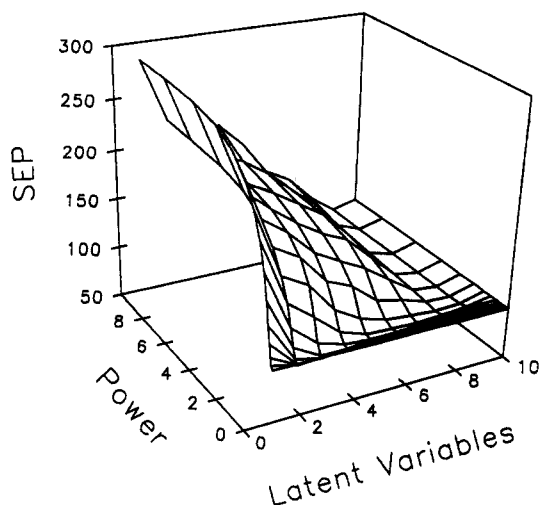


Figure 3. Prediction error surface for continuum regression.

greatly reduced predictive abilities. As before, 50 calibration/prediction sets were used for the estimation of the SEPs, which are listed in Table 3. This model exhibited a significant improvement in predictive ability over the Rossum model at high conductivities, but did not perform as well as the EMRM, which required fewer adjustable parameters. An attempt was also made to develop an MLR model that was more parsimonious.¹⁶ To do this, all combinations of the 10 variables were examined to find the subset with the best predictive ability. Results for the optimum model (MLRb) are given in Table 3. Interestingly, the cation concentrations dominated the best models, although the reason for this is unclear. The MLRb model exhibited significant improvements over the Rossum model at high conductivities, but did not offer a substantial improvement over the EMRM. Likewise MLRc, which was optimized with the addition of four other variables (alkalinity, hardness, total organic carbon, reactive silica), did not exhibit significantly better predictive abilities than the EMRM.

In an attempt to overcome some of the weaknesses often exhibited by MLR, other regression methods (PCR, PLS, CR) were also employed with the same 10 ionic concentrations. These methods have been described elsewhere in the literature^{17–20} and exhibit advantages with certain types of data sets. The results obtained in this case are summarized in Table 3 and Figure 3. Figure 3 is a continuum regression surface showing how the prediction error varies with the number of latent variables included in the model and the power factor, n , used in the continuum regression model. As pointed out by Lorber et al.,¹⁹ MLR, PCR, and PLS are subsets of CR. In Figure 3, MLR corresponds to $n = 0$ (number of latent variables, 10), PCR corresponds to $n = \infty$ ($n = 10$ is a good approximation), and PLS corresponds to $n = 1$. It is clear from the figure that there is not a significant reduction

in the number of latent variables for the best model obtained with any of the methods, indicating that there is not a high degree of correlation among the input variables. Furthermore, Table 3 and Figure 3 show a comparable performance among all of the regression methods. While these results are significantly better than the Rossum model, the improvement over the EMRM is not significant.

Finally, artificial neural networks^{21–24} were applied to the prediction problem with the hope that they might compensate for any nonlinearities causing difficulties with the regression approaches. A back-propagation neural network was used with one hidden layer. Inputs were mean-centered at zero with a span of 2 and the output was mean-centered at 0.5 with a span of 0.5. As before, 200 samples were used for training the neural network. Preliminary trials indicated that two hidden nodes were sufficient for consistent results and 5000 iterations were used. Results given in Table 3 indicate disappointing prediction abilities for the neural network in both concentration ranges. The addition of other water quality parameters as inputs did not improve the performance.

CONCLUSIONS

The results presented here demonstrate that alternatives to the Rossum model exhibit superior predictive abilities at higher conductivities. It was found that Rossum's definition for composite charge did not work as well at high conductivities as did the simpler definition, so the modified definition is recommended. Extension of the model to higher concentrations can also be accomplished through the use of an additional term and an adjustable parameter. The extended modified Rossum model, which incorporates both of these modifications, exhibited the best performance of the semiempirical models at higher conductivities. Improvements in the extended range can also be achieved by employing multivariate regression methods, all of which exhibited performance similar to the EMRM. The use of neural networks was not found to be particularly effective for the prediction of conductance.

Of course, the selection of a model depends not only on its predictive ability but also its robustness. The MRM is completely general but suffers somewhat in its predictive ability for high conductance samples. MLR produced the smallest prediction errors in the extended range, but results were not significantly better than the EMRM. The EMRM represents perhaps the best compromise between maximum predictive ability and the minimum number of adjustable parameters and is recommended on the basis of these results.

ACKNOWLEDGMENT

The authors acknowledge helpful conversations with Dr. J. C. T. Kwak. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Received for review August 11, 1993. Accepted December 3, 1993.*

* Abstract published in *Advance ACS Abstracts*, January 15, 1994.

- (16) Seasholtz, M. B.; Kowalski, B. R. *Anal. Chim. Acta* 1993, 277, 165–177.
- (17) Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: New York, 1987.
- (18) Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* 1987, 59, 1007A–1017A.
- (19) Lorber, A.; Wangen, L.; Kowalski, B. R. *J. Chemom.* 1987, 1, 19–31.
- (20) Kowalski, B. R.; Seasholtz, M. B. *J. Chemom.* 1991, 5, 129–145.
- (21) Wythoff, B. J. *Chemom. Intell. Lab. Syst.* 1993, 18, 115–155.
- (22) Bos, M.; Bos, A.; van der Linden, W. E. *Anal. Chim. Acta* 1990, 233, 31–39.
- (23) Rumelhart, D. E.; McClelland, J. L. *Parallel Distributed Processing*, MIT Press: Cambridge, MA, 1986; Vols. 1 and 2.
- (24) Hecht-Nielsen, R. *Neurocomputing*; Addison-Wesley: New York, 1991.