AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

Centimeter

Inches

2

Conf-930438--8

WSRC-MS-93-222

CANONICAL CORRELATION OF WASTE GLASS COMPOSITIONS AND DURABILITY, INCLUDING pH (U)

by

D. F. Bickford

Westinghouse Savannah River Company
Savannah River Site
Aiken, South Carolina 29808

D. Oksoy
Alfred University
Alfred, NY
L. D. Pye
Alfred, NY
W. G. Ramsey
(WSRC)

A document prepared for:
Annual Meeting - American Ceramic Society
at Cincinnati, OH
from 04/26/93 thru

DOE Contract No. DE-AC09-89SR18035

DISCLAIMER

CANONICAL CORRELATION OF WASTE GLASS COMPOSITIONS AND DURABILITY, INCLUDING pH

by

Dolun Öksoy
College of Business, Alfred University, Alfred, NY 14802
L. David Pye
Industry-University Center for Glass Research
NYS College of Ceramics at Alfred University, Alfred, NY 14802

Dennis F. Bickford
Westinghouse Savannah River Co.,
Aiken, SC 29808

W. Gene Ramsey,
Westinghouse Savannah River Co.,
Aiken, SC 29808

SUMMARY

Control of waste glass durability is a major concern in the immobilization of radioactive and mixed wastes. Leaching rate in standardized laboratory tests is being used as a demonstration of consistency of the response of waste glasses in the final disposal environment. The leaching of silicate and borosilicate glasses containing alkali or alkaline earth elements is known to be autocatalytic, in that the initial ion exchange of alkali in the glass for hydrogen ions in water results in the formation of OH and increases the pH of the leachate. The increased pH then increases the rate of silicate network attack, accelerating the leaching effect. In well formulated glasses this effect reaches a thermodynamic equilibrium when leachate saturation of a critical species, such as silica, or a dynamic equilibrium is reached when the pH shift caused by incremental leaching has negligible effect on pH.

The seven day PCT leach test [1] results of a statistically balanced composition set of thirty samples in the Si-B-Na-Ca-Al-Fe-O system [2] were analyzed using canonical correlation methods. Glass compositions ranged from very durable to those with relatively high solubility: Boron release data examined varied by three orders of magnitude, silicon and sodium releases varied by over two orders of magnitude. Leachate results were analyzed to determine if the inclusion of pH with the leachate composition in the dependent variable set is statistically justifiable, and if it improves the overall correlation when compared to similar analyses excluding the pH term. It was found that the inclusion of the pH term leads to an extremely high correlation coefficient of R= -0.984. Exclusion

of the pH term lead to an overall correlation coefficient of -0.948 indicating that the pH is ------. The leachate final pH could be predicted by glass composition with a R of 0.952.

I. CANONICAL CORRELATION ANALYSIS

Here we will be concerned with finding a vector of weights for each of two sets of variables such that the correlation between the two linear composites using these weights is a maximum. This is the problem of canonical correlation, which was described by its originator Hotelling H. (1935), as a way of determining the most predictable criterion.

Contrary to multiple regression the canonical correlation situation involves at least two dependent variables on the dependent variable side as well as at least two variables on the independent variable side. The variables on the dependent side are weighted in such a way that the linear composite of the dependent variables has a maximum correlation with a linear composite derived from the independent variables. If these dependent variables (i.e. Si, B, Na, Ca, Al, Fe, pH) are considered as criteria , then their unique weights yields the most predictable criterion for linear combination of the dependent variables. Statistically, the problem is to find a vector $\boldsymbol{a}$ and a vector $\boldsymbol{b}$ such that the correlation between the composites $\boldsymbol{a'x}$ and $\boldsymbol{b'y}$ is a maximum where $\boldsymbol{x}$ is a vector of random variables such as Si, B, Na, Ca, Al, Fe, pH and $\boldsymbol{y}$ is another vector of random variables such as SiO_2, Al_2O_3, B_2O_3, Fe_2O_3, FeO, CaO, Na_2O . The correlation between $\boldsymbol{a'x}$ and $\boldsymbol{b'y}$ for an arbitrary $\boldsymbol{a}$ and $\boldsymbol{b}$ can be written as

$$r_{a'x,b'y} = \frac{a'C_{x,y}b}{\sqrt{a'C_x a}\sqrt{b'C_y b}}$$

where C_{xy} is the cross-covariance matrix between the variables in $\boldsymbol{x}$ and the variable in $\boldsymbol{y}$, C_x is the covariance matrix of variables in $\boldsymbol{x}$, and C_y is the covariance matrix of variables in $\boldsymbol{y}$.

This canonical correlation model is merely a tool for examining the interrelationship between two sets of variables. Multivariate normality needs to be assumed for tests of statistical significance. Canonical correlation is closely related to discriminant analysis since interest is focused on the relationship between two sets of continuous variables.

The technique is also related to multivariate regression analysis, where, each continuous dependent is regressed upon a set of continuous independent variables. If we consider one set of variables as dependent variables, then the canonical weights for the dependent variables are the regression weights that would predict the linear combination of the dependent variables (that is, the canonical variate representing the dependent variables obtained from the canonical correlation analysis). As in principal components, it is possible to generate more then one canonical correlation between two variable sets.

2. FORMULATION OF CANONICAL CORRELATION ANALYSIS

Similar to principal components analysis, the problem can be formulated as a maximization problem. Algebraically, the problem is to maximize $a'C_{xy}b$ subject to the constraints that $a'C_x a = 1$ and $b'C_y b = 1$. These constraints are needed to solve for unique weighting vectors a and b and simply indicate that each linear composite is constrained to have a variance of one. Thus the problem is to find the vectors a and b that maximize.

$$z = a'C_{x,y}b - \lambda_1(a'C_x a - 1) - \lambda_2(b'C_y b - 1)$$

where λ_x and λ_y are Lagrange multipliers.

We may standardize the variables and maximize.

$$z = a'R_{x,y}b - \lambda_x(a'R_x a - 1) - \lambda_y(b'R_y b - 1)$$

where R_{xy} represents the cross-correlations between the first set of variables (that is, x) and the second set of variables (that is, y); R_x represents the intercorrelations among the first set of variables (x), and R_y represents the intercorrelations among the second set of variables (y). The standardization is purely a matter of convenience in exposition and the avoidance of unduly complicated notations, and loses no generality.

Applying multivariate differential calculus to the function z results in two characteristic equations that can be solved for λ, a, and b. They are

$$(R_x^{-1}.R_{x,y}.R_x^{-1}.R'_{x,y} - \lambda.I).a = 0$$

and

$$(R_y^{-1}.R'_{x,y}.R_y^{-1}.R_{x,y} - \lambda.I).b = 0$$

There are two latent roots and associated vectors a and b that satisfy these characteristic equations. The largest latent root turns out to be the largest canonical correlation squared and the associated weighting vectors are a and b. The next largest latent root, λ_2, is the next largest canonical correlation squared and has another pair of canonical weight vectors, a_2 and b_2, such that the correlation between a'_2x and b'_2y is maximal given certain conditions discussed below. The number of possible nonzero latent roots or, equivalently, canonical correlations is equal to the dimension of the smallest variable set. For our example, the dimension of the smallest variable set is seven, so that the seven nonzero canonical correlations are possible. The size of the canonical correlation is, of course, a function of the intercorrelations within and between the variable sets. As in principal components and linear discriminant function analysis, some of the canonical correlations might be to small to be of any practical significance. A test for the statistical significance of canonical correlation is required.

Other properties of the solution to the six characteristic equations associated with canonical correlation analysis are discussed below. First of all, note that the dimensions of a and b will be different unless the number of variables in the two variable sets is equal. Let a_i be the i^{th} set of canonical weight vectors associated with one set (x) and b_j be the

j^{th} set of canonical weight vectors associated with the other set (y). Then the $a_i x$ of the first set are uncorrelated with each other, that is, $A'R_xA = \mathbf{I}$ where A has as column vectors the set of canonical weight vectors associated with the set x and R_x is the correlation matrix for x. Similarly, the $b_j y$ of the second set satisfies the property that $B'R_yB = \mathbf{I}$ and, hence, the canonical variates associated with y are uncorrelated with one another. Most important is the fact that the correlation between $a_i x$ and $b'_j y$ is zero for $i \neq j$ and equal to the canonical correlation for $i = j$. This can be summarized as $A'R_{xy}B = D_p$ where R_{xy} is the cross-correlation matrix between x and y and D_p is a diagonal matrix of canonical correlation coefficients.

Note that the maximization of multivariate functions associated with both linear discriminant function analysis and principal component analysis also resulted in characteristic equations for which characteristic (latent) roots and their associated latent vectors needed to be solved. As in the other characteristic equations, there may be more than one nonzero latent root. For principal components, the latent roots turned out to be the variances of the associated principal components. For linear function analysis, the latent roots turned out to be the ratio of the between- to the within-group sum of squares for the associated linear discriminant functions. It is not surprising then, that the latent roots in the two characteristic equations associated with canonical correlation analysis turn out to be the squares of the canonical correlations. The latent roots for both the characteristic equations associated with canonical correlation analysis are identical. Once these are solved for, their associated vectors, the a_i's, associated with the first set of variables (that is, x) can be solved for by substituting, in turn, the latent roots into the first characteristic equation. Similarly, substituting the latent roots, in turn, into the second characteristic equation yields the associated latent vectors, the b_j's, for the second set of variables (y).

3. SIGNIFICANCE TESTS OF CANONICAL VARIATES

Two kinds of significance tests are of interest in canonical correlation analysis. The first is an overall test to decide whether there is any significant linear relationship between the two set of variables. If overall significance is found, we would then want to know how many of the canonical-variate pairs are significant. The significance tests here are closely related to those described for discriminant analysis.

To see this relationship we can compare eigen value μ_i from the basic equation

$$(W^{-1}B - \lambda I)v = 0$$

for discriminant analysis via the discriminant criterion approach and the corresponding eigen value λ_i from the basic equation

$$(R_y^{-1}.R'_{x,y}.R_y^{-1}.R_{x,y} - \lambda.I).b = 0$$

in the canonical correlation approach. Namely, the corresponding eigenvalues resulting from these two basic equations are related by the equality

$$\mu_i = \lambda_i(1 - \mu_i)$$

As a consequence, Wilks' Λ criterion, can be expressible as

$$\Lambda = 1\Big/\prod_{i}^{r}(1+\mu_i)$$

may also be expressed in terms of the λ_i as follows:

$$\Lambda = \prod_{i}^{r}(1-\lambda_i) \qquad (1)$$

The above demonstration, of course, shows only that this alternative expression for Λ holds when the λ_i results from canonical analysis as applied to the problem of discriminant analysis. However, it is quite plausible that Eq. (1) will continue to hold for canonical analysis in general. For each λ_i there is a conditionally maximal value of R_c^2, the squared correlation between corresponding pairs of canonical variates constructed from the two sets of variables. Thus, each factor of the product

$$(1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_r) \qquad (2)$$

is in fact the coefficient of alienation between a particular pair of canonical variates. This is consistent with the fact that Λ is a statistics that is inversely related to the magnitude of differences or strength of relationship: the smaller the value of Λ, the greater the difference or relationship in question.

There is a problem, however, in that the definition of Λ as the ratio $|\boldsymbol{W}|/|\boldsymbol{T}|$ (used in connection with discriminant analysis) does not make sense in context of canonical analysis. The sample in this situation is not composed of several subgroups, and hence there is no such thing as a within-group **SSCP** matrix ***W***. The resolution of this difficulty lies in introducing a more general concept of which the ***W*** matrix is a special instance applicable to multigroup significance tests and discriminant analysis. The general concept is the error **SSCP** matrix, which we will denote $\boldsymbol{S_e}$. Thus, a more general definition of Λ is given as follows:

$$\Lambda = \frac{|S_e|}{|T|}$$

In the application of Λ encountered up to now, the within-group **SSCP** matrix was the appropriate error **SSCP** matrix. In the context of canonical analysis, error **SSCP** matrix is the residual **SSCP** matrix after the effect of the correlations between the canonical-variate pairs have been removed. Of course, this matrix need not actually be computed in order to determine the value of Λ, since Λ may be obtained from Eq. (2) once the required eigenvalues are found.

After Λ has been computed, the overall significance test may be carried out by either the chi-square approximation or the *F*-ratio approximation, with $q+1$. This is consistent with the fact that, in using the canonical correlation approach to discriminant analysis, *K*-1 "dummy criterion variables" were employed; that is, the number of groups is one more than the number of variables in the second set. Thus, Barlett's chi-square approximation becomes

$$
\begin{aligned}
V &= -[N-3/2-(p+q)/2]\ln\Lambda \\
&= -[N-3/2-(p+q)/2]\sum_{j=1}^{r}\ln(1-\lambda_i)
\end{aligned} \tag{3}
$$

with pq degrees of freedom. Similarly, Rao's F-ratio approximation is now written as

$$R_{Rao} = \frac{1-\Lambda^{1/s}}{\Lambda^{1/s}} \cdot \frac{ms - pq/2+1}{pq} \tag{4}$$

where

$$m = N-3/2-(p+q)/2 \text{ and } s = \sqrt{\frac{p^2q^2-4}{p^2+q^2-5}}$$

and R_{Rao} is to be referred to as F-distribution with pq degrees of freedom in the numerator, and $[ms - pq/2 + 1]$ in the denominator. In the case of at least one set consists of single variable in the sense that canonical correlation then reduces to, at most, a multiple correlation.

Beyond the overall significance test described above there are the tests for deciding how many of the canonical correlations should be regarded as significant. The procedure here depends again on the fact that each term of the sum in Eq. (3) for V is itself an approximate chi-square variate. That is for each j,

$$V_j = -[N-3/2-(p+q)/2]\ln(1-\lambda_j) \tag{5}$$

is distributed approximately as chi-square with $p + q - (2j - 1)$ degrees of freedom. Consequently, the cumulative differences between V and V_1, V_2 and so on, are also approximate chi-square variates, and they permit our testing whether a significant (linear) relationship exists between the two sets of variables after the effects of the first, second, and so forth, canonical-variate pairs have cumulatively been removed.

Residual After Removing	Approximate χ^2 statistics	d.f
First canonical pair	$V - V_1$	$(p-1)(q-1)$
First two canonical pairs	$V - V_1 - V_2$	$(p-2)(q-2)$
First three canonical pairs	$V - V_1 - V_2 - V_3$	$(p-3)(q-3)$
....	....	....

As soon as the residual after removing the effects of the first s canonical variate pairs becomes smaller than the prescribed centile point of the appropriate chi-square distribution, we may conclude that only the first s canonical correlations are significant.

4. EXAMPLE WITH AND WITHOUT pH

The data set given in tables 1 and 2 were provided by Savannah River Technology Center, Westinghouse Savannah River Co. We will analyze this data set with and without pH term.

Table 1. Leachate Composition and pH Data of One Week PCT Average of three replicate samples

Glass ID	Si Ave ppm	B Ave ppm	Na Ave ppm	Ca Ave ppm	Al Ave ppm	Fe Ave ppm	Ave pH
1	52.573	48.466	53.089	0	17.533	0	9.05
2	24.807	2.318	42.459	4.127	7.095	0	10.92
4	105.257	9.837	249.357	0	38.482	0	11.76
5	98.434	30.634	77.591	18.516	0	0	10.00
6	88.960	66.617	106.893	0	0	6.814	9.53
7	8343.100	2679.600	9494.067	0	0	0	11.39
8	58.534	7.541	126.130	0.562	0	1.395	11.45
9	8924.700	1180.567	10499.333	1.133	0	0	12.63
10	383.510	43.230	572.957	0	0	54.190	11.95
11	20.983	24.327	64.396	3.445	8.091	0	10.18
13	39.754	183.253	405.597	0	18.745	0	10.26
15	59.070	4.896	272.367	5.272	24.712	0	12.00
16	102.617	13.486	214.450	0	42.116	41.627	11.69
17	33.395	15.328	95.899	0.807	14.621	0	11.18
18	3501.000	2766.067	8780.533	2.600	0	0	12.01
19	159.351	410.934	1061.800	0	0	17.382	10.97
20	164.523	23.710	638.113	0.100	0	1.153	12.32
21	67.103	14.711	104.683	0.884	17.441	6.390	11.06
22	36.865	9.990	65.900	0.435	23.049	2.236	10.75
23	40.831	7.296	54.823	0.519	27.380	4.213	10.47
24	63.451	26.088	150.317	1.036	0	6.582	11.16
25	50.296	35.125	90.120	0.587	17.751	6.720	10.05
26	63.941	6.808	137.417	0.894	23.359	4.405	11.46
27	27.797	6.769	71.170	1.141	17.094	1.132	11.04
28	67.661	17.911	89.273	0	25.124	21.198	10.65
29	60.108	13.116	93.032	0.280	16.300	3.211	10.95
30	71.352	59.421	245.347	1.248	10.972	0	11.47
31	70.137	14.925	160.140	0.973	37.169	7.290	11.49
32	24.751	7.202	42.069	0.557	15.925	1.568	10.42
33	41.955	9.152	63.038	0.624	21.693	3.957	10.72
Min	20.983	2.318	42.069	0	0	0	9.05
Max	8924.7	2766.067	10499.333	18.516	42.116	54.19	12.63
μ	761.561	258.044	1137.412	1.525	14.155	6.382	11.033

Table 2. Glass Composition Molar Fractions

Glass ID	SiO_2	Al_2O_3	B_2O_3	Fe_2O_3	FeO	CaO	Na_2O
1	0.6049	0.1046	0.1407	0.0000	0.0000	0.0000	0.1499
2	0.5903	0.1038	0.0495	0.0000	0.0000	0.1035	0.1529
4	0.6009	0.1107	0.0430	0.0000	0.0000	0.0000	0.2454
5	0.5887	0.0349	0.1365	0.0000	0.0000	0.0961	0.1438
6	0.5857	0.0296	0.1312	0.0871	0.0256	0.0000	0.1407
7	0.5938	0.0062	0.1446	0.0000	0.0000	0.0000	0.2555
8	0.5800	0.0252	0.0473	0.0828	0.0290	0.0934	0.1424
9	0.5953	0.0012	0.0491	0.0000	0.0000	0.1002	0.2543
10	0.5777	0.0165	0.0409	0.0794	0.0294	0.0000	0.2561
11	0.4817	0.1084	0.1543	0.0000	0.0000	0.1040	0.1516
13	0.5000	0.1198	0.1390	0.0000	0.0000	0.0000	0.2412
15	0.5036	0.1121	0.0453	0.0000	0.0000	0.0980	0.2410
16	0.5046	0.1084	0.0440	0.0874	0.0225	0.0000	0.2331
17	0.4551	0.0810	0.1344	0.0715	0.0340	0.0896	0.1344
18	0.4912	0.0057	0.1429	0.0000	0.0000	0.1026	0.2576
19	0.4643	0.0260	0.1319	0.0901	0.0146	0.0000	0.2730
20	0.4876	0.0067	0.0496	0.0975	0.0133	0.0992	0.2461
21	0.5684	0.0751	0.0861	0.0374	0.0080	0.0384	0.1865
22	0.4694	0.1295	0.0974	0.0489	0.0166	0.0532	0.1850
23	0.5047	0.1568	0.0774	0.0295	0.0162	0.0352	0.1802
24	0.5302	0.0500	0.1036	0.0527	0.0166	0.0560	0.1909
25	0.5212	0.0876	0.1357	0.0355	0.0111	0.0370	0.1720
26	0.5553	0.0755	0.0432	0.0493	0.0244	0.0563	0.1960
27	0.5112	0.0972	0.0795	0.0333	0.0122	0.0936	0.1731
28	0.5412	0.1016	0.0938	0.0496	0.0212	0.0000	0.1926
29	0.5348	0.0817	0.0765	0.0784	0.0301	0.0345	0.1640
30	0.5542	0.0673	0.1130	0.0000	0.0000	0.0631	0.2025
31	0.5099	0.1002	0.0776	0.0346	0.0109	0.0368	0.2300
32	0.5264	0.1130	0.0943	0.0442	0.0282	0.0541	0.1400
33	0.5116	0.1234	0.0877	0.0372	0.0181	0.0430	0.1790
Min	0.4550	0.0010	0.0410	0.0000	0.0000	0.0000	0.1340
Max	0.6050	0.1570	0.1540	0.0980	0.0340	0.1040	0.2730
μ	0.5350	0.0750	0.0930	0.0380	0.0130	0.0500	0.1970

4.1 Analysis with Inclusion of pH Term

We have two subsets of variables, $\boldsymbol{x}$ and $\boldsymbol{y}$, where $\boldsymbol{x}$ has seven durability measures [x_1, ... ,x_7] , namely (Si, B, Na, Ca, Al, Fe, pH) that had the correlation matrix R_x, where, R_x indicates that sodium, boron, and silicon release are highly correlated, i.e. leaching occurs by network dissolution. Low correlation with Fe, Ca, and Al is because of large fraction of samples containing zero for these elements in the balanced design, and possible precipitation of these elements. On the other hand $\boldsymbol{y}$ has a measure of seven oxide mole fraction measures [y_1, ... ,y_7], namely (SiO_2, Al_2O_3, B2O3, Fe_2O_3, FeO, CaO, Na_2O) that had the correlation matrix R_y, where R_y shows that the sample space is balanced, with the exception of the Fe_2O_3 and FeO which is caused by the natural equilibrium with air at melt temperature of these species. Further more the cross-correlation matrix R_{xy} for these sets of variables provides cross-correlation information between dependent variable set $\boldsymbol{x}$ and independent variable set $\boldsymbol{y}$.

R_x =

	Si	B	Na	Ca	Al	Fe	pH
Si	1	0.806	0.956	−0.057	−0.361	−0.14	0.397
B	0.806	1	0.914	−0.034	−0.384	−0.148	0.311
Na	0.956	0.914	1	−0.044	−0.396	−0.145	0.438
Ca	−0.057	−0.034	−0.044	1	−0.22	−0.204	−0.154
Al	−0.361	−0.384	−0.396	−0.22	1	0.115	−0.032
Fe	−0.14	−0.148	−0.145	−0.204	0.115	1	0.182
pH	0.397	0.311	0.438	−0.154	−0.032	0.182	1

R_y =

	SiO_2	Al_2O_3	B_2O_3	Fe_2O_3	FeO	CaO	Na_2O
SiO_2	1	−0.288	−0.208	−0.284	−0.216	−0.187	−0.107
Al_2O_3	−0.288	1	−0.07	−0.219	0.007	−0.153	−0.335
B_2O_3	−0.208	−0.07	1	−0.257	−0.25	−0.117	−0.222
Fe_2O_3	−0.284	−0.219	−0.257	1	0.85	−0.221	−0.064
FeO	−0.216	0.007	−0.25	0.85	1	−0.186	−0.322
CaO	−0.187	−0.153	−0.117	−0.221	−0.186	1	−0.246
Na_2O	−0.107	−0.335	−0.222	−0.064	−0.322	−0.246	1

R_{xy} =

	SiO_2	Al_2O_3	B_2O_3	Fe_2O_3	FeO	CaO	Na_2O
Si	0.303	−0.529	0.075	−0.335	−0.343	0.075	0.435
B	0.087	−0.532	0.321	−0.329	−0.36	0.048	0.47
Na	0.192	−0.576	0.136	−0.342	−0.372	0.135	0.501
Ca	0.162	−0.108	0.164	−0.37	−0.352	0.451	−0.234
Al	−0.193	0.791	−0.312	−0.076	0.058	−0.33	0.025
Fe	−0.004	−0.108	−0.299	0.502	0.425	−0.49	0.32
pH	−0.072	−0.399	−0.618	0.098	−0.014	0.306	0.664

If $\boldsymbol{a}$ is a nonnull vector, then the determinant of $R_x^{-1}R_{xy}R_y^{-1}R'_{xy} - \lambda I$ must vanish since the columns of the matrix must be linearly dependent to meet the conditions of the characteristic equation. Substituting the appropriate matrix from our example into $|R_x^{-1}R_{xy}R_y^{-1}R'_{xy} - \lambda I|$, we find the characteristic roots λ_i = [0.969, 0.847, 0.659, 0.504, 0.263, 0.085,0] . The square root of these characteristic roots yields the canonical correlations, namely $\sqrt{\lambda_i}$.

The next step is to find the characteristic vectors $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ associated with the largest canonical correlation namely $\sqrt{\lambda_i}$. The vectors $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ of interest are found by solving the homogeneous equations

$R_x^{-1}R_{xy}R_x^{-1}R'_{xy}a_i - \lambda_i a_i = 0$ and $R_y^{-1}R'_{xy}R_x^{-1}R_{xy}b_i - \lambda_i b_i$ respectively.

The eigen vectors of $\boldsymbol{x}$ and $\boldsymbol{y}$ corresponding to characteristic roots mentioned above are summarized by $\boldsymbol{A}$ and $\boldsymbol{B}$ matrix below. The columns of $\boldsymbol{A}$ and $\boldsymbol{B}$ matrix are the eigen vectors $\boldsymbol{a}_i$, $\boldsymbol{b}_i$ corresponding to the characteristic root λ_i For interpretive convenience, we can rescale the vectors by multiplying all the elements by the unique scalar that will set the element with the largest absolute value to one. For this example, the eigen vectors of $\boldsymbol{x}$ and $\boldsymbol{y}$ obtained has been rescaled based on the following largest absolute values (0.663, 0.672, -0.616, 0.772, 0.801, 0.727, -0.818) and (0.445, 0.443, 0.445, -0.444, -0.445, -0.444, -0.444) respectively, and summarized in $\boldsymbol{A}$ and $\boldsymbol{B}$. The other elements or weights in the vector may then be easily compared relative to the most important variability in the canonical variate.

$$\mathbf{A} = \begin{pmatrix} -0.601 & 0.149 & -0.735 & 0.035 & 1 & -0.074 & -0.499 \\ -0.284 & -0.395 & -0.791 & 0.471 & -0.203 & 0.02 & -0.476 \\ 1 & -0.305 & 1 & 0.578 & -0.719 & 0.208 & 1 \\ 0.229 & -0.461 & 0.322 & 0.244 & 0.04 & 1 & -0.031 \\ -0.414 & 0.704 & 0.203 & 1 & -0.012 & 0.261 & 0.014 \\ -0.072 & 0.485 & -0.513 & -0.158 & -0.004 & 0.866 & 0.039 \\ 0.778 & 1 & 0.25 & -0.186 & -0.003 & -0.162 & -0.121 \end{pmatrix}$$

$$
E = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0.986 & 0.987 & 0.971 & 0.981 & 0.973 & 0.975 & 0.977 \\
0.858 & 0.843 & 0.857 & 0.859 & 0.852 & 0.856 & 0.854 \\
0.767 & 0.765 & 0.758 & 0.734 & 0.758 & 0.761 & 0.778 \\
0.248 & 0.262 & 0.259 & 0.269 & 0.254 & 0.255 & 0.232 \\
0.878 & 0.889 & 0.885 & 0.889 & 0.888 & 0.89 & 0.887 \\
0.963 & 0.987 & 0.976 & 0.988 & 0.971 & 0.974 & 0.972
\end{pmatrix}
$$

Based on rescaled vector the first canonical variate for the first set of dependent variables, $\boldsymbol{x}$, from $\boldsymbol{A}$ is :

$\boldsymbol{a}'_1\boldsymbol{x}$ = -0.601 Si -0.284 B +Na + 0.229Ca - 0.414 Al - 0.072 Fe +0.778 pH

Similarly based on rescaled vectors the first canonical variate for the first set of independent variables, $\boldsymbol{y}$, from $\boldsymbol{B}$ is:

$\boldsymbol{b}'_1\boldsymbol{y}$ = SiO_2 + 0.986 Al_2O_3 + 0.858 B_2O_3 + 0.767 Fe_2O_3 + 0.248 FeO + 0.878 CaO + 0.963 Na_2O

We have found a set of weights for both sets of variables resulting in two composites that have the maximum correlation among all possible pairs of composites. The first canonical variable, which is a composite of the durability variables is primarily defined by pH, Na, Ca and with a relatively large negative weight for Si and relatively small negative weights for Al, B and Fe. The second canonical variable primarily defined by all. These first linear composites $\boldsymbol{a}'_1\boldsymbol{x}$ and $\boldsymbol{b}'_1\boldsymbol{y}$ are plotted in Figure 1.

The overall significance test based on Wilks' Λ=0.001, see (1), indicates strong relationship between dependent and independent sets with 8 degrees of freedom as chi-square approximation or F-ratio approximation. Also based on Barlett's chi-square approximation or Rao's F-ratio approximation we have observed significant (linear) relationship between the two sets of variables after the effects of the first, second, third, fourth, canonical-variate pairs have cumulatively been removed. Therefore we may conclude the first four canonical correlations are significant.

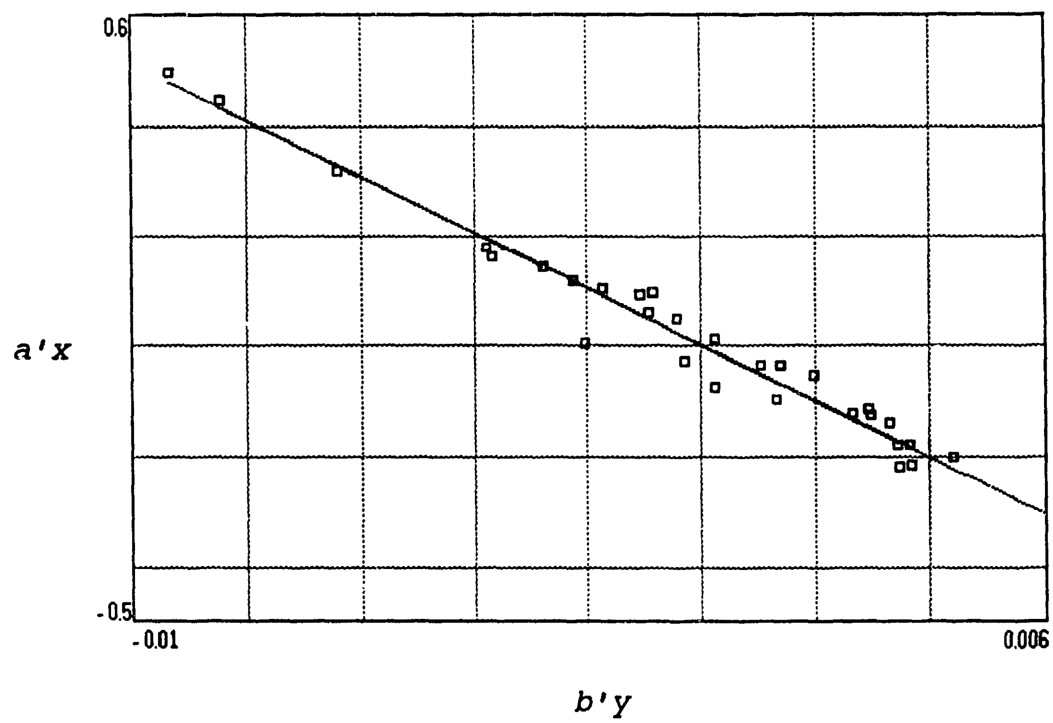


Figure 1. Dependent variable set **x** versus independent variable set **y** at canonical correlation -0.984, and (a'_1x, b'_1y) regressed (including pH)

4.2 Analysis with Exclusion of pH Term

We have two subsets of variables, x and y, where x has six durability measures [x_1, ... ,x_6] , namely (Si, B, Na, Ca, Al, Fe) that had the correlation matrix R_x, where, R_x indicates that sodium, boron, and silicon release are highly correlated, i.e. leaching occurs by network dissolution. Low correlation with Fe, Ca, and Al is because of large fraction of samples containing zero for these elements in the balanced design, and possible precipitation of these elements. On the other hand y has a measure of seven oxide mole fraction measures [y_1, ... ,y_7], namely (SiO_2, Al_2O_3, B2O3, Fe_2O_3, FeO, CaO, Na_2O) that had the correlation matrix R_y, where R_y shows that the sample space is balanced, with the exception of the Fe_2O_3 and FeO which is caused by the natural equilibrium with air at melt temperature of these species. Further more the cross-correlation matrix R_{xy} for these sets of variables provides cross-correlation information between dependent variable set x and independent variable set y.

If $\boldsymbol{a}$ is a nonnull vector, then the determinant of $R_x^{-1}R_{xy}R_y^{-1}R'_{xy} - \lambda I$ must vanish since the columns of the matrix must be linearly dependent to meet the conditions of the characteristic equation. Substituting the appropriate matrix from our example into $|R_x^{-1}R_{xy}R_y^{-1}R'_{xy} - \lambda I|$, we find the characteristic roots λ_i = [0.9, 0.691, 0.519, 0.306, 0.263, 0.081,0] . The square root of these characteristic roots yields the canonical correlations, namely $\sqrt{\lambda_i}$.

The next step is to find the characteristic vectors a_i and b_i associated with the largest canonical correlation namely $\sqrt{\lambda_i}$. The vectors a_i and b_i of interest are found by solving the homogeneous equations

$R_x^{-1}R_{xy}R_x^{-1}R_{xy}'a_i - \lambda_i a_i = 0$ and $R_y^{-1}R_{xy}'R_x^{-1}R_{xy}b_i - \lambda_i b_i$ respectively.

The eigen vectors of x and y corresponding to characteristic roots mentioned above are summarized by A and B matrix below. The columns of A and B matrix are the eigen vectors a_i, b_i corresponding to the characteristic root λ_i For interpretive convenience, we can rescale the vectors by multiplying all the elements by the unique scalar that will set the element with the largest absolute value to one. For this example, the eigen vectors of x and y obtained has been rescaled based on the following largest absolute values (-0.763, 0.625, 0.806, -0.795, 0.779, -0.642) and (0.445, 0.445, -0.485, -0.445, 0.445, 0.444, -0.444) respectively, as summarized in A and B. The other elements or weights in the vector may then be easily compared relative to the most important variability in the canonical variate.

$$\mathbf{A} = \begin{pmatrix} -0.528 & 0.646 & -0.365 & -0.513 & 1 & -0.528 \\ -0.072 & 0.742 & -0.138 & -0.561 & -0.127 & -0.567 \\ 1 & -0.504 & 1 & 1 & -0.794 & 1 \\ 0.313 & -0.565 & 0.173 & 0.013 & 0.035 & -0.708 \\ -0.552 & -0.135 & 0.588 & -0.012 & -0.011 & -0.165 \\ -0.181 & 1 & -0.111 & 0.051 & -0.010 & -0.544 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 0.890 & 1 & 1 & 1 & 1 \\ 0.989 & 0.971 & 0.931 & 0.975 & 0.973 & 0.975 & 0.977 \\ 0.850 & 0.855 & 0.754 & 0.860 & 0.852 & 0.856 & 0.854 \\ 0.768 & 0.757 & 0.484 & 0.757 & 0.758 & 0.761 & 0.777 \\ 0.253 & 0.261 & 0.333 & 0.256 & 0.254 & 0.255 & 0.233 \\ 0.883 & 0.886 & 0.826 & 0.886 & 0.888 & 0.891 & 0.887 \\ 0.974 & 0.979 & 1 & 0.972 & 0.971 & 0.974 & 0.972 \end{pmatrix}$$

Based on rescaled vector the first canonical variate for the first set of dependent variables, x, from A is :

a'_1x = -0.528 Si -0.072 B +Na + 0.313Ca - 0.552 Al - 0.181 Fe

Similarly based on rescaled vectors the first canonical variate for the first set of independent variables, y, from B is:

b'_1y = SiO_2 + 0.989 Al_2O_3 + 0.85 B_2O_3 + 0.768 Fe_2O_3 + 0.253 FeO + 0.883 CaO + 0.974 Na_2O

We have found a set of weights for both sets of variables resulting in two composites that have the maximum correlation among all possible pairs of composites. The first canonical variable, which is a composite of the durability variables is primarily defined by Na, Ca and with a relatively large negative weight for Si and relatively small negative weights for Al, B and Fe. The second canonical variable primarily defined by all. These first linear composites a'_1x and b'_1y are plotted in Figure 2.

The overall significance test based on Wilks' Λ=0.007, see (1), indicates strong relationship between dependent and independent sets with 7 degrees of freedom as chi-square approximation or F-ratio approximation. Also based on Barlett's chi-square approximation or Rao's F-ratio approximation we have observed significant (linear) relationship between the two sets of variables after the effects of the first, second canonical-variate pairs have cumulatively been removed. Therefore we may conclude the first two canonical correlations are significant.

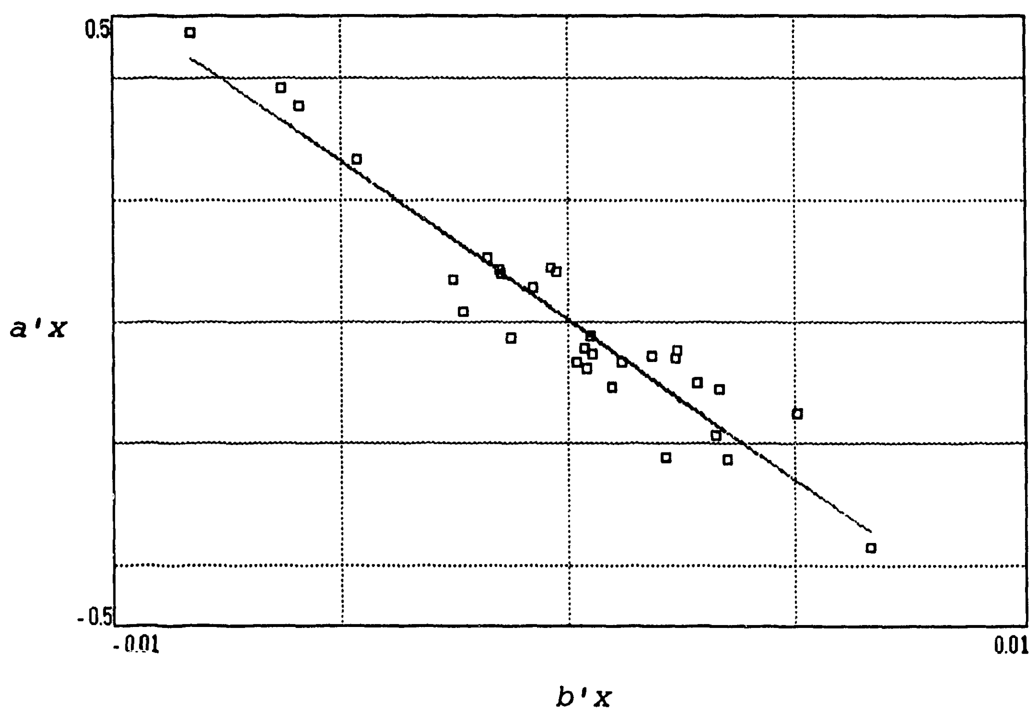


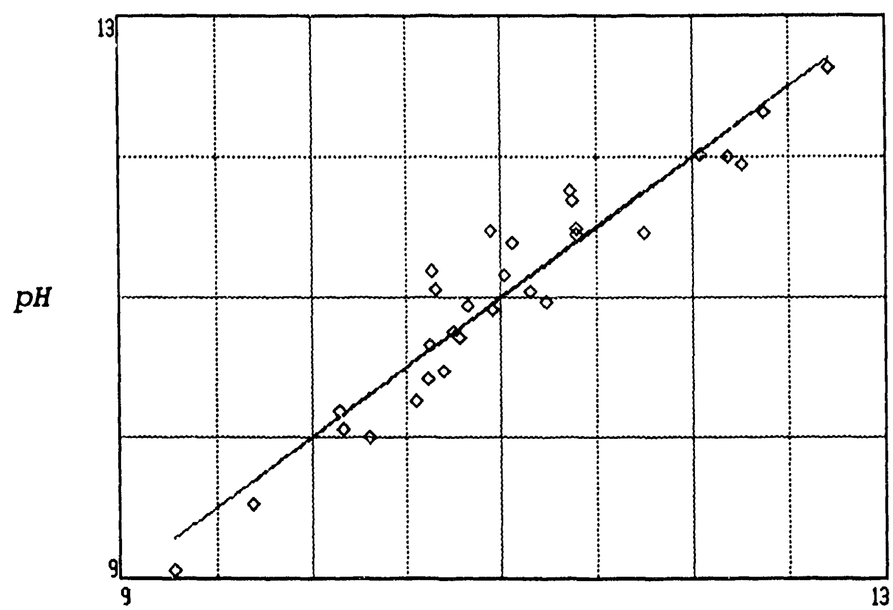
Figure 2. Dependent variable set **x** versus independent variable set **y** at canonical correlation -0.948, and (a'_1x, b'_1y) regressed (excluding pH)

4.3 Multivariate regression of pH against Glass Composition

We also studied the relation of pH to glass composition mole fractions and obtained the following relation for the estimate of pH, correlation of correlation of R = 0.952:

$$\hat{pH} = -144.585 + 153.054 SiO_2 + 150.188 Al_2O_3 + 145.518 B_2O_3 + 146.907 Fe_2O_3 + 185.031 FeO + 162.315 CaO + 167.477 Na_2O$$

This relation of pH against the estimate of pH is plotted in Figure 3.



$$\hat{\text{pH}} = \beta_0 + \sum \beta_i \cdot Oxide_i$$

Figure 3. pH regressed against $\hat{pH}$

ACKNOWLEDGMENT

The research for this project was supported by Grant No. DOE-WSRC-AA464150; we are grateful to this agency for their continuing support

REFERENCES

[1] Dunteman, G. H. (1984). Introduction to Multivariate Analysis, *Sage*, California
[2] Harris R. J. (1975). A primer of Multivariate Statistics, *Academic Press*, New York.
[3] Hotelling, H. (1935). The most predictable criterion, *Journal of Educational Psychology*, 26, 139-142.
[4] Jantzen, C. M. and Bibler, N.E. (1990) " Nuclear Waste Glass Product Consistency test (PCT) Version 3.0", WSRC-TR-90-539, *Westinghouse Savannah River Co.*
[5] Matchcad 3.1, (1993), *MathSoft Inc.*,
[6] Ramsey, W. G. (1993), Ceramics Engineering Ph.D. Thesis, *Clemson University*.
[7] SYSTAT W5.1, (1993), *SYSTAT Inc.*
[8] Tatsuoka, M. M. (1971). Multivariate Analysis, Techniques for Education and Psychological Research, *Wiley*, New York.

DATE FILMED

7/29/93

END