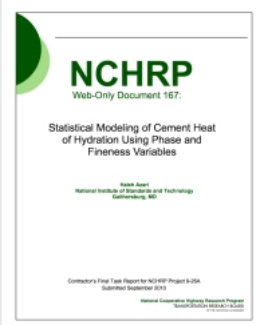

Statistical Modeling of Cement Heat of Hydration Using Phase and Fineness Variables

DETAILS

0 pages | null | PAPERBACK
ISBN 978-0-309-43021-0 | DOI 10.17226/22917

AUTHORS

ACKNOWLEDGMENT

This work was sponsored by the American Association of State Highway and Transportation Officials (AASHTO), in cooperation with the Federal Highway Administration, and was conducted in the National Cooperative Highway Research Program (NCHRP), which is administered by the Transportation Research Board (TRB) of the National Academies.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The heat of hydration of hydraulic cements results from the complex sets of phase dissolution and precipitation activity accompanying the addition of water to a cement. This process generates heat, as well as an increased potential for thermal cracking in some concrete structures. The potential heat of this hydration process is measured in two ways: 1) through an acid dissolution of the raw cement and a hydrated cement after seven days, or 2) isothermal calorimetry. In principal, the heat of hydration should be predictable from knowledge of the cement composition, and perhaps some measure of the cement fineness or total surface area. The improved mineralogical estimates provided by quantitative X-ray powder diffraction, together with improved statistical data exploration techniques that examine nonlinear combinations of candidate constituents, are used to explore alternative predictive models for 7-day heat of hydration (HOH7) based on a set of more complete and more accurate characterizations of portland cements. In the modeling described in this report we make essential use of the groupings, or classes, of potential explanatory variables of phase, fineness, and other physical parameters. An All Possible Alternating Conditional Expectations (APACE) exploratory tool, created by combining All Possible Subsets Regression with the Alternating Conditional Expectation (ACE), is used to determine which variables *within* an explanatory variable class and which subsets of variables *across* explanatory variable classes exhibit the highest potential predictive power for additive nonlinear models for HOH7. While a single, strong candidate model for HOH7 did not emerge from these analyses, some general conclusions did result. Good fitting models include a key structural mineralogical phase (belite preferred), a calcium sulfate phase (bassanite preferred), a total fineness or surface area component (Blaine fineness preferred), and ferrite in conjunction with Fe_2O_3, or aluminate, or cubic aluminate. Surprisingly, TiO2 recurs as a component in good-fitting models.

CHAPTER 1- INTRODUCTION AND RESEARCH APPROACH

1.1 Background

Hydraulic cements react with water through a process called hydration via a series of chemical reactions, ultimately resulting in the precipitation of interlocking hydration products that provide strength to the structure. The hydration process produces heat that in some concrete placements may cause expansion and, potentially, cracking upon cooling to ambient conditions. The temperature rise can also be beneficial in the case of cold-weather concrete placements, where the heat facilitates hydration and keeps the concrete from freezing [1]. ASTM C186 was adopted in 1944 as a standard test method for determining the heat of hydration of hydraulic cements. This test procedure involves measurement of the heat of solution of dry cement specimens that have been hydrated for 7 d and for 28 d. The difference between the heat of solution values between the dry and the partially hydrated cement specimens is taken as the heat of hydration for that time period. The heat values are expressed in SI units[1] of kJ kg^{-1}. This test is time-consuming, involves a hazardous mixture of nitric and hydrofluoric acids, and has low precision with d2s values of 48 kJ kg^{-1} for measurements between different laboratories. Cements in is study range in 7 d heat from the heat values range from 261 kJ $kg^{-1}$to 468 kJ kg^{-1}. Conduction calorimetry provides an alternative with ASTM C1679, "Standard Practice for Measuring Hydration Kinetics of Hydraulic Cementitious Mixtures Using Isothermal Calorimetry." This method has been shown to be useful in the estimation of total heat, in assessing early-age reactions and setting problems, and in measuring the influences of sulfate additions and mineral admixtures on heat evolution [2]. An example is shown in Figure 1-1 for the first 24 h of hydration of a portland cement. This procedure is generally limited to about 3 days as the rate of heat development after that time is difficult to measure.

The rate of hydration of a cement depends upon its mineralogy, the mass of each phase, the particle size distribution, the water-to-cement ratio, and the temperature and relative humidity of curing [3]. Copeland et al. [4] ascribe the total heat of hydration as emanating from two processes: 1) the chemical reactions in the formation of hydration products, thought to be responsible for 80 % of the heat, and 2) the heat of wetting of the subsequent colloidal hydration product accounting for the remaining 20 %. The C186 test utilizes a fixed water/cement ratio but the curing temperature may differ depending upon the rate of heat evolution. The dependence of heat of hydration on cement Type is illustrated by data collected by the Portland Cement Association for a sampling of cements produced between 1992 and 1997 (Figure 1-2), where Type III cements generally have higher heats of hydration and Type IV the lowest [5]. Limits on composition and fineness in ASTM C150 and AASHTO M85 reflect their influences on

[1] Non-SI units (cal/g) may be converted by multiplying by the conversion factor 4.1868.

the heat of hydration. Type II cement with the moderate heat option has restrictions on either the sum of $C_3S+4.75*C_3A \leq 100$ (the heat index equation) or a seven-day heat release at most of 290 kJ kg^{-1} when measured by ASTM C186. Type IV (low heat cement) has limits on either phase mass fraction for C_3S, C_2S, and C_3A of 35 %, 40 %, and 7 %, respectively, or a C186 heat value limit of 250 kJ kg^{-1} at seven days. All these phase estimates are from Bogue-calculated values as described in ASTM C150 and AASHTO M85.

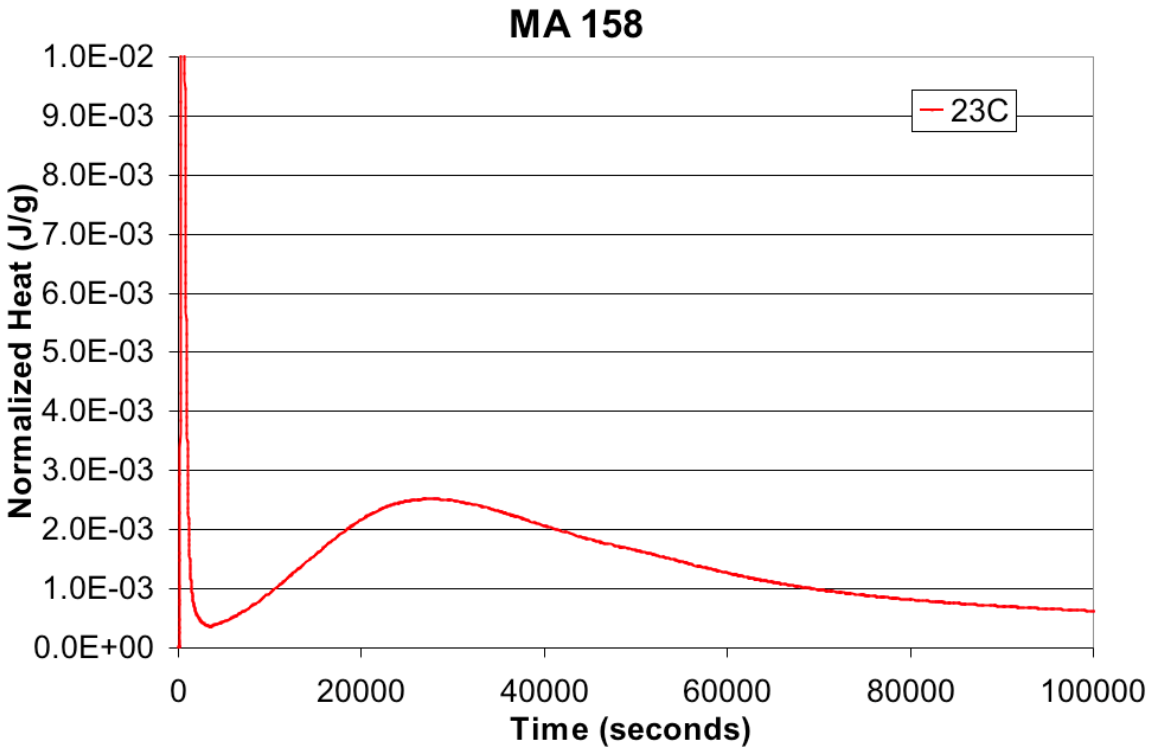

Figure 1-1- Isothermal calorimetry curve of heat evolution based upon a single measurement for a hydrating cement for 24 h shows an initial peak of heat in the first hour, followed by a dormant period and then a gradual rise before tapering off in heat development for the following 23 h

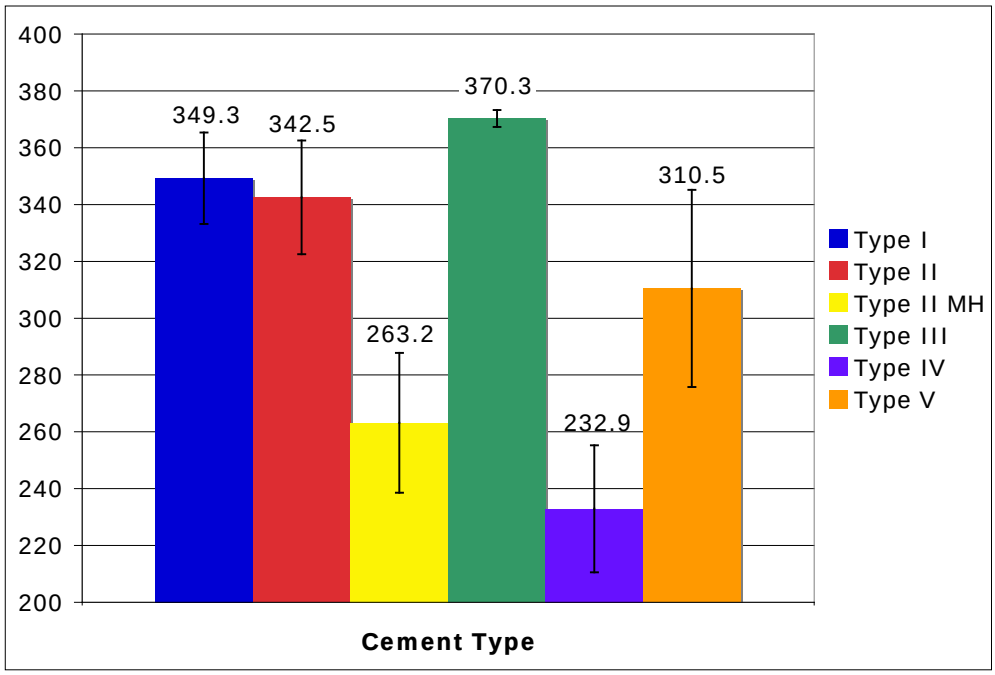

Figure 1-2- The dependence of heat of hydration, with +- 1s uncertainties indicated, on cement Type for a limited sampling of cements produced between 1992 and 1997 [1]

As the specifications recognize the influences of cement phase composition on heat of hydration, a more accurate and more complete characterization of a cement's mineralogical and textural characteristics should provide an improved data set to evaluate their influences. Data exploration techniques beyond multilinear regression should potentially yield improved models for predicting heat of hydration of portland cements.

1.2 Effects of Cement Phase Characteristics

Bulk chemical analyses, typically by X-ray fluorescence, are used to determine cement chemistry with estimates of phase abundance derived using the Bogue formulas. Errors in these estimates arise from the variability of clinker phase chemistry relative to the assumed compositions, from the failure to account for minor constituents, and from inaccuracy in measured analytical values [5,3]. Source materials composition, processing, and kiln operating conditions also affect clinker mineral composition. Subsequent grinding with calcium sulfates and processing additions also affect cement performance attributes.

Early work on developing predictive models for heat of hydration focused on the contributions of the individual clinker phases, the synergistic effects of multi-phase cement hydration, and the heat of precipitation of the resulting hydration products [4,6,7,8]. Lerch [9] concluded that gypsum retards early hydration of cements that have high tricalcium aluminate content, while accelerating hydration of cements with low tricalcium aluminate content, and that the alkali aluminate (the orthorhombic form) is more reactive and requires a larger gypsum addition than a low-alkali aluminate (the

cubic form). More recently, the accelerating effects of potassium oxide (presumably from the alkali sulfates and alkali-substituted tricalcium aluminate) on alite, ferrite, and aluminate have been demonstrated [10,11,13]. Calcium sulfate additions have an accelerating effect on hydration of the silicates and ferrite while retarding the initial set and the reactions of the aluminate phases. Gypsum has also been seen to retard heat development in mixtures of clinker phases. In cases where gypsum may have been partially dehydrated during cement processing forming bassanite (hemihydrate), heat (192 kJ kg^{-1}) would evolve upon rehydration to gypsum [6,7]. Overall, this reflects the complex synergy of the mineral constituents of the portland cement system during the hydration process. Confounding this further is the influence of mineral surface areas exposed upon grinding. A scanning electron microscope (SEM) micrograph of polished cross-section of cement particles (Figure 1-3) illustrates the multi-phase nature and compositional complexity of cements. The complexity of the cement mineral compositions and the interactions between phases during hydration suggest that a more complete and more accurate characterization of cements should yield data that would improve predictive capabilities for cement performance characteristics.

1.3 Previous Work on Heat of Hydration

Taylor [3] summarized multilinear predictive models for heat (H_t) involving least-squares regression using mass percent Bogue potential phase compositions and coefficients accounting for the degree of hydration of each phase at a specific age (Eqn. 1, Table 1-1). Note that the aluminate (C_3A) reaction can result in ettringite (AFt) at early ages and monosulfate (AFm) at later ages, with differing enthalpy of hydration, illustrating the complexity of the cement hydration process. Taylor [3] notes that the enthalpies of formation of clinker phase hydration products would refine estimates of potential heat evolution, but the uncertainties in their values and variability in the reaction stoichiometry of hydration would introduce additional errors into the estimates.

$$H_t \text{ (kJ kg}^{-1}) = a(C_3S) + b(C_2S) + c(C_3A) + d(C_4AF) \qquad \text{Eqn. 1}$$

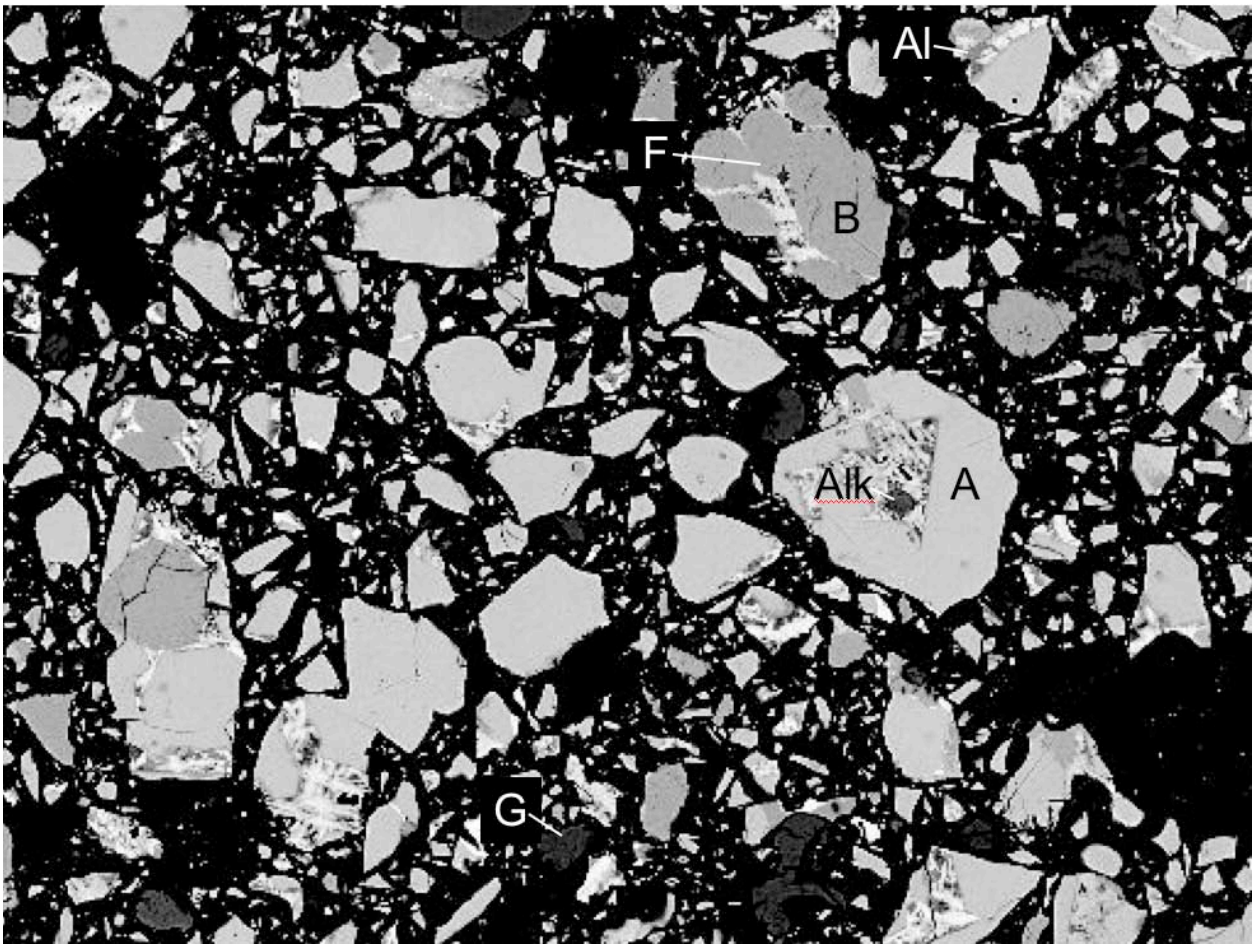

Figure 1-3- SEM backscattered electron micrograph of polished cement grains embedded in an epoxy illustrates the complicated shapes and multiphase particles typical of a portland cement. Phase code is: alite = A, belite = B, aluminate = Al, ferrite = F, alkali sulfate = Alk, and gypsum = G. Field Width = 500 μm

Table 1-1- Heat of hydration values for clinker phases, and coefficients at 7 d and 28 d. From Taylor [3]

	Value of the coefficient (kJ kg $^{-1}$) for age (d)			
Compound	Coefficient	7 d	28 d	Enthalpy of complete hydration (kJ kg $^{-1}$)
C_3S	a	222	126	-517 ± 13
C_2S	b	42	105	-262
C_3A	c	1556	1377	-1144; -1672 (AFm, AFt reactions)
C_4AF	d	494	494	-418

Poole summarized work relating Bogue phase composition to heat of hydration and developed a multilinear regression model (Eqn. 2) that utilized aluminate and alite with an R^2 of 89 % on the data used to develop the model, exhibiting little bias, and a ±21 kJ kg $^{-1}$ 95 % confidence interval on the regression [12]. The cement fineness, as measured by Blaine permeability, was not found to be a significant variable in predicting heat of hydration with simulations showing that a change in fineness from 300 m^2/kg to 400 m^2/kg resulted in an increase in heat of hydration of only 8.1 kJ kg $^{-1}$ in 7 days.

$$HOH_7 \text{ (kJ kg}^{-1}\text{)} = 133.9 + 9.36(C_3A) + 2.13(C_3S) \quad \text{Eqn. 2}$$

Phases not considered in these predictive models, like the alkali and calcium sulfates and the specific form of tricalcium aluminate, can exert significant influences on a cement's hydration characteristics. It is reasonable to ask whether a more complete accounting for cement phases and fineness measures will provide a better set of predictive variables. More generally, would one expect relatively simple multilinear models to work well given the complexity of hydration processes? Phase abundance directly determined by quantitative X-ray powder diffraction analysis (XRD) and some non-phase variables (particle size distribution, fineness) considered to affect heat evolution are considered in this work in modeling 7-day heat of hydration. This more complete accounting of the cement mineral composition and the inclusion of other cement characteristics has the potential to provide a more accurate prediction of a cement's heat performance.

1.4 Problem Statement

The heat of hydration of hydraulic cements results from the complex sets of phase dissolution and precipitation activity accompanying cement hydration. This process generates heat, as well as an increased potential for thermal cracking in some concrete structures. The potential heat of this hydration process is measured in two ways: 1) through an acid dissolution of the raw cement and a hydrated cement after seven days, or 2) isothermal calorimetry.

1.5 Research Objectives

In principal, the heat of hydration should be predictable from knowledge of the cement composition, cement fineness. The improved mineralogical estimates provided by quantitative X-ray powder diffraction, together with improved statistical data exploration techniques that examine nonlinear combinations of candidate constituents, are used to explore alternative predictive models for 7-day heat of hydration (HOH7) based on a set of more complete and more accurate characterizations of portland cements.

1.6 Scope of Study

Quantitative XRD data, bulk chemistry, and ASTM C186 heat of hydration values were included from 31 cements from the CCRL proficiency test program, 18 cements from a NCHRP program 18-05 on cement performance, and 4 cements provided by the US Army Corps of Engineers. The variables of phase content, size distribution and fineness measures, and ancillary measurements on cement characteristics, such as setting time, are organized into logical groupings that will be used in subsequent statistical modeling. Exploratory statistical tools are used to determine which representative variables of the key classes consistently exhibit the most predictive power for HOH7, what ranges of R^2 are achieved for best nonparametrically additive nonlinear fitted models for those choices of variables in various combinations, and the resulting implications for the parametric modeling of HOH7 in terms of phases and ancillary variables.

CHAPTER 2- STATISTICAL MODELING

2.1 The Data

In the modeling we make essential use of groupings, or classes, of potential explanatory variables. Initial data exploration utilized All Possible Subsets Regression (APSR) and Principal Components Regression (PCR) to identify linear combinations of variables that might be useful in predicting 7-day heat of hydration. Nonlinear transformation of key variables identified by APSR should increase their predictive ability, accomplished using Alternating Conditional Expectation (ACE) [13,14]. Our key tool will be All Possible Alternating Conditional Expectations (APACE) created by combining the All Possible Subsets Regression concept with Alternating Conditional Expectation (ACE). Using APACE, we can easily explore which variables *within* an explanatory variable class exhibit the highest potential predictive power for HOH7 based on additive nonlinear models, and which subsets of variables *across* explanatory variable classes exhibit the highest potential predictive power for additive nonlinear models for HOH7. An additive nonlinear model is a weighted sum of nonlinear function summands.

The principal conclusions of this report will be (1) which representative variables of the key classes consistently exhibit the most predictive power for HOH7, (2) what ranges of R^2 are achieved for best nonparametrically additive nonlinear fitted models for those choices of variables in various combinations, and (3) resulting implications for the parametric modeling of HOH7 in terms of phases and ancillary variables.

2.1.1 Phase Measures

The variables (Table 2-1) are organized into logical groupings that will be used in subsequent statistical modeling. The complete data set is found in the Appendix, Quantitative XRD (QXRD) data, bulk chemistry, and ASTM C186 heat of hydration values were included from 31 cements from the CCRL proficiency test program, 18 cements from a NCHRP program 18-05 on cement performance [15], and 4 cements provided by the US Army Corps of Engineers. The QXRD data were the average of three replicates each of a bulk cement and an extraction residue after a salicylic acid / methanol extraction. The chemical extraction serves to remove the calcium silicate phases (C_3S, C_2S), concentrating the interstitial phases (C_3A, C_4AF, periclase, alkali sulfates) and the calcium sulfates (gypsum, bassanite, anhydrite) to facilitate identification and quantitative estimates. The calcium sulfate and arcanite (K_2SO_4) have been found to affect hydration as discussed previously. Bulk oxide and Bogue-calculated values were used for comparative purposes and setting times and 3-day strength were selected as they had been mentioned in studies as being relevant [3]. Measures of fineness were included as it is generally thought that the fineness, or total surface area, should influence the rate of reaction and therefore heat evolution. Two measures of fineness are used: 1) Blaine fineness which is an indirect measure of the total particle surface area and 2) particle size distribution by laser diffraction, which expresses a particle size distribution in terms of a measure of the width of the particle size

distribution. Setting time by Vicat needle (ASTM C191) was included as a higher heat of hydration may result in a more rapid setting time.

Table 2-1- Predictor Variables and Classes used in the exploratory data analysis

Mineral Phase by XRD (mass fraction)	alite, belite, aluminate, cubic aluminate, orthorhombic aluminate, ferrite, periclase, alkali sulfates, gypsum, bassanite, anhydrite
Bulk oxide content	CaO, SiO_2, Fe_2O_3, Al_2O_3, SO_3, MgO, Na_2O, K_2O, TiO_2, P_2O_5, ZnO, Mn_2O_3
Fineness	Blaine; and particle size by laser diffraction: D10, D50, D90, Span
Extras (other physical measurements)	calcite
	Set time (Vicat)
	3d strength

Since knowledge of the cement mineralogy is desired, the direct measurement of phases by XRD provides an alternative means to the Bogue phases for quantitative phase analysis [16]. A standard test method for clinker and cement may be found with ASTM C1365 [17]. XRD has been used in the cement industry since the mid-1920's, and has been more widely used since the development of the commercial powder diffractometers in the late 1940's. It is ideally suited for fine-grained materials (like clinker and cements) for direct phase analysis as each phase produces a unique diffraction pattern independent of the other phases, and the intensity of which is proportional to its concentration. Difficulties encountered in earlier calibration-based analyses included the measurement of intensities and use of suitable reference standards. These difficulties have been largely overcome with the development of the Rietveld method for quantitative analysis [18,19].

The Rietveld method uses crystal structure models for each phase as reference standards, allowing the structural and chemical variables for each phase to be simultaneously iteratively refined using the whole-diffraction pattern, employing a least-squares fitting procedure. The refinement results in a set of best-fit structure models for each phase with scale factors coming from relative phase proportion according to Eqn. 3 [20]:

$$W_p = \frac{S_p (ZMV)_p}{\sum_p [S_p (ZMV)_p]} \qquad [3]$$

where

W_p = the mass fraction of phase p,

S_p = the Rietveld scale factor,

Z = the number of formula units per unit cell,

M = the mass of the formula unit, and

V = the unit cell volume

Results of an international inter-laboratory study produced the within- and between-lab uncertainties of a XRD analysis following ASTM C1365 are shown in Table 2-2, where results of two properly conducted tests by the same operator should not vary by more than 'r' while results of two tests on the same clinker or cement by two different laboratories should not differ from each other by more than 'R', both at 95 % confidence [16].

Table 2-2- Within-laboratory (s-within) and the between-laboratory standard deviations (s-between) and 95 % d2s values expressed as mass percents [16].

	Repeatability Within-lab		*Reproducibility Between-lab*	
	s-within	r (d2s, within)	s-between	R (d2s, between)
alite	0.74	2.04	2.23	6.18
belite	0.64	1.77	1.41	3.91
aluminate	0.47	1.31	0.74	2.05
ferrite	0.49	1.36	0.95	2.63
periclase	0.23	0.63	0.32	0.89
arcanite	0.22	0.60	0.41	1.13
gypsum	0.21	0.59	0.58	1.62
bassanite	0.39	1.08	0.81	2.24
anhydrite	0.27	0.74	0.63	1.75
calcite	0.99	2.73	0.50	1.50

2.1.2 Fineness Measures

All other variables held equal, the more finely ground the cement, the more rapidly it would be expected to react. However, this does not always seem to be the case in actual practice. The practitioner is presented with multiple measures of fineness. The Blaine fineness is an indirect measure of total particle surface area, denominated by volume of material, based on time for unit volume air to flow through a cement powder packed cylinder [21]. It might be expected to be a good predictor of HOH_7 because intuitively the hydration reactions, and resulting heat, would increase with increased surface area. An alternate means of measuring cement fineness is through particle size distribution. Laser diffraction of cements provides size distribution data using measures of D10, D50, and D90, indirect measures of 10th/50th/90th percentiles of particle size distribution. Span, an additional variable is a function of these three, measuring the approximate range of the particle size distribution. The predictive performance of these measures for HOH7 will be examined in this work.

2.1.3 Time of Setting

The Vicat test measures the penetration depth of a standardized needle, where initial and final set times are when the needle penetrates a cement paste less than prescribed limits of 25 mm and zero mm, respectively [22]. Since it is the difference between the final and initial Vicat time, (VicatF - VicatI), which is the true test correlative with rapidity of setting, we can assume that HOH_7 may be expected to correlate with [1/(VicatF - VicatI)]. This simple example illustrates the occasional need to transform raw variables in order to achieve meaningful variable response. A potential confounding factor with this test procedure is the use of normal consistency paste, where the prescribed water content for the Vicat test will vary by cement, which may affect the setting times and may be different from the HOH_7 test conditions.

2.2 Tools of Statistical Modeling

Historically, modeling responses like HOH_7 or compressive strength typically have involved multilinear modeling of raw (as presented by test data) inputs or judiciously transformed inputs, with transformations motivated by established engineering relationships. Excellent examples of this are the multilinear Bogue models relating mineralogical phases to oxide compositions. Typically the existing models have been derived by straightforward multilinear fitting, or by forwards or backwards selection techniques that progress in an automated fashion through choices of subsets of potential predictor variables, comparing explanatory power gained by successive addition or deletion of variables through reductions or decreases in R^2, or residual variance, or F statistic. Over the last forty years, however, sophistication in model selection techniques for both multilinear candidate models and interesting nonlinear extensions of multilinear models has increased tremendously.

In our initial approach to modeling HOH_7 based on the mineralogical, fineness, and particle size distribution data, we employed All Possible Subsets Regression (APSR) and Principal Components Regression (PCR). While neither technique is new, using them separately and in conjunction offers significant improvements over backwards / forwards selection techniques.

Comparing, however, the quality and credibility of models and fits obtained using the linear fits provided by APSR and PCR with the nonlinear-addend multilinear fits obtained by the use of Alternating Conditional Expectation (ACE) and All Possible Subsets Alternating Conditional Expectation (APSACE), we end up emphasizing the results from the superior nonlinear (ACE) tool. We find APSACE linked with the use of automated parametric fitting tools to be an easy-to-use, easy-to-interpret, and insightful approach to meaningful model search that offers an enormous extension of the classical model search employing multilinear functions.

In this work, we adapt the tool to all possible linear combinations of nonlinearly nonparametrically transformed versions of the explanatory variables, thus extending the scope of the approach enormously. Since our algorithm assays combinations by exhaustion, we limit ourselves to combinations of less than or equal to 10 variables. But for the size of the dataset being examined, and the number of explanatory variables being considered, that does not appear to be unduly restrictive. Nonetheless, since a number of the basic techniques used in developing and selecting multilinear models are still applicable, in practice or at least in motivating more modern approaches, we spend a little time discussing APSR and associated statistics.

2.2.1 Prescreening Variables: Scatter Plots

A fundamental principle of modern exploratory data analysis, including model selection, is to prescreen data using graphics. Visualization almost always contributes significantly to understanding. Graphical prescreening enables the modeler to (1) scan for outlying data or obviously anomalous patterns in data, (2) to gauge the potential statistical explanatory power and potential model meaningfulness of each explanatory variable assessed against the response [variable] of interest (HOH_7).

A simple multiplot (Figure 2-1) with associated, or overlaid, table of correlation coefficients (Figure 2-2) is of tremendous value in either motivating selection of variables before the modeling or theorizing why a semi-automated selection technique such as APSACE ends up selecting the best-fit models that it does after the modeling.

A cross-correlation table of the variables serves another purpose. Variables that cross-correlate highly may contain much the same explanatory information for the proposed model. Incorporating both in a model may lead to either an unnecessary degree of redundancy, or overfitting, in the model or to numerical instabilities (e.g., multicollinearities) in the numerical fitting procedure. Alite and belite (variables C_3S and C_2S, respectively) are highly anticorrelated (typically in the -0.80 to -0.90 range). This is

a natural result of their physical co-occurrence as calcium silicates at the expense of one another, depending upon the availability of lime during the clinkering process. The natural modeling solution is to employ one or the other, selecting the variable for the model situation that gives the best goodness-of-fit statistics.

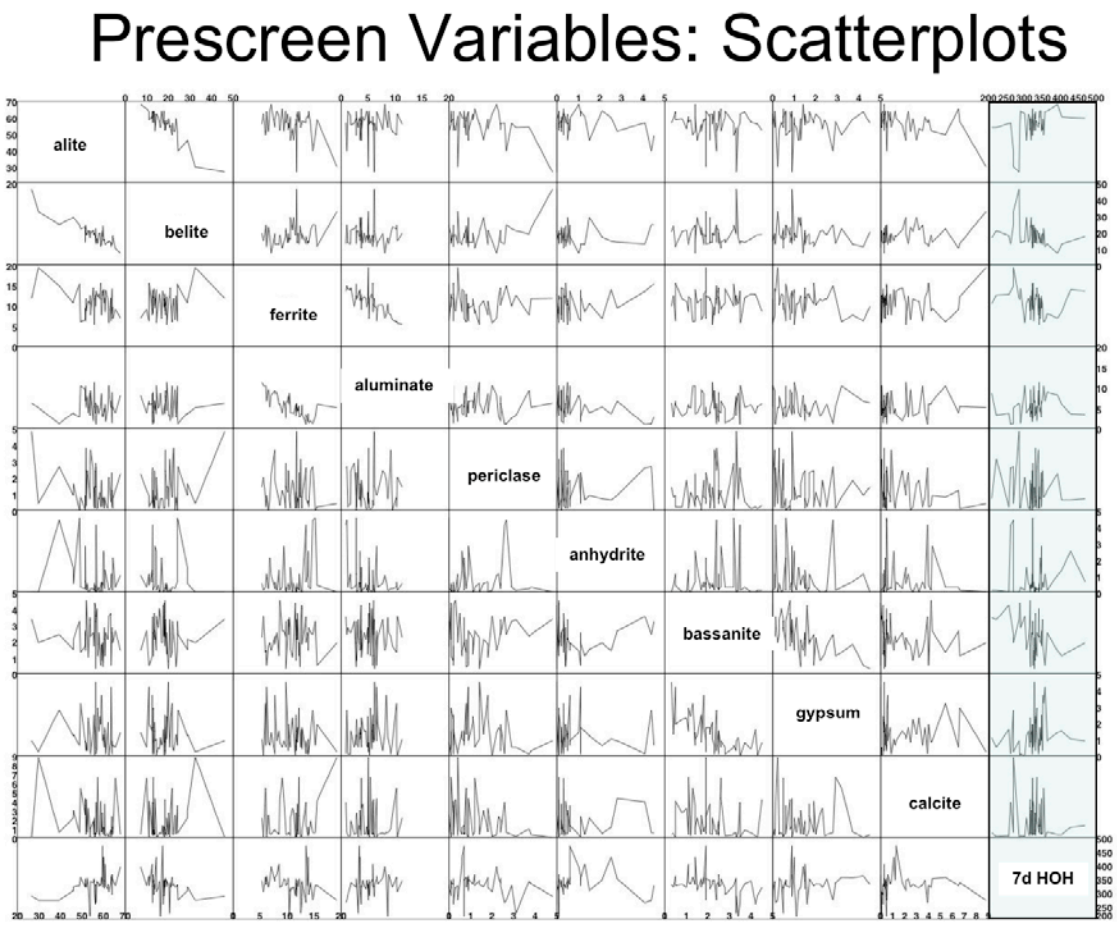

Figure 2-1- Scatter plots are a useful visualization tool to pre-screening tool to look for patterns and anomalous data.

		LIME	C3S	C2S	C3A	C4AF	C4C	C4O	PERI	ANHYD	BASS	GYP	CALC	D10
		2	3	4	5	6	7	8	9	10	11	12	13	14
LIME	2	1.000												
C3S	3	-0.207	1.000											
C2S	4	0.115	-0.867	1.000										
C3A	5	0.016	-0.299	0.201	1.000									
C4AF	6	-0.042	-0.009	-0.052	-0.736	1.000								
C4C	7	-0.106	0.078	-0.104	-0.631	0.846	1.000							
C4O	8	0.131	-0.166	0.115	-0.012	0.034	-0.504	1.000						
PERI	9	-0.046	-0.319	0.308	-0.085	0.000	-0.035	0.073	1.000					
ANHYD	10	-0.015	-0.122	-0.049	0.185	-0.325	-0.150	-0.244	0.054	1.000				
BASS	11	-0.106	-0.129	0.048	0.101	-0.073	-0.147	0.169	0.135	0.037	1.000			
GYP	12	0.143	0.051	-0.102	-0.134	0.084	0.064	0.010	-0.061	-0.052	-0.610	1.000		
CALC	13	0.436	-0.286	0.077	0.135	0.042	-0.009	0.085	-0.198	-0.014	-0.179	0.010	1.000	
D10	14	-0.239	-0.026	-0.089	0.425	-0.303	-0.167	-0.180	-0.118	0.093	-0.083	0.174	0.186	1.000
D50	15	-0.097	-0.038	-0.099	0.265	-0.182	-0.120	-0.070	-0.154	0.250	0.022	0.260	0.119	0.681
D90	16	-0.109	-0.113	-0.068	0.032	-0.005	0.010	-0.023	-0.049	0.413	0.123	0.142	0.142	0.320
SPAN	17	-0.015	-0.117	0.040	-0.388	0.323	0.248	0.063	0.146	0.235	0.201	-0.206	0.043	-0.552
HOH7	18	0.032	0.388	-0.312	-0.232	0.139	0.137	-0.036	-0.142	0.039	-0.431	0.177	-0.034	-0.053

		D50	D90	SPAN	HOH7
		15	16	17	18
D50	15	1.000			
D90	16	0.799	1.000		
SPAN	17	-0.262	0.352	1.000	
HOH7	18	-0.160	-0.130	0.037	1.000

Figure 2-2- The correlation matrix is a useful data screening tool to assess both the correlation between individual variables and 7d heat of hydration, both for correlations and anti-correlations (C3S and C2S) between variables.

2.2.2 Transformations

Often, rather than dealing with raw explanatory and response variables in their original form, it may be necessary to transform the raw variables, or combinations of raw variables, in some manner so as to elicit their best contribution to a model. An example is provided by the 1/[VicatF - VicatI] transformation discussed above, where the subtraction gives a more meaningful time relevant entity for modeling, and inversion is pre-performed because it is anticipated that time to set and HOH7 will be inversely related. We sometimes work with log(Blaine) and log(HOH7) rather than Blaine and HOH7 themselves because logging brings the fineness and heat numbers down to the scale of other contributors to the model, and because it is common to log transform lengths, areas, volumes, or density type numbers to achieve distribution conformity or homogenize variances. The cement engineering literature abounds in rules of thumb, which correspond to potential transformations. For example, the Lime Saturation Factor (LSF) in Eqn. 4 is a ratio.

$$LSF = \frac{CaO}{2.8 \cdot SiO_2 + 1.18 \cdot Fe_2O_3 + 0.8 \cdot Al_2O_3} \qquad Eqn.\,4$$

Alumina Ratio, Silica Ratio, Alkali-Sulfate Ratio all correspond to transformations of basic (oxide) variables which may map more directly to certain structural properties of the cement. Generally speaking, we transform for physical meaningfulness, for model simplicity, and for "statistical niceness." Niceness refers, for example, to transforming to achieve homogeneity or homoscedasticity where the variance or standard uncertainty of an input variable expands, contracts, or varies in a random fashion over the range of the variable being studied.

2.2.3 Log Odds Transform

Another example of a potentially useful transformation, suggested by Tukey and Mosteller [23] is the log odds transformation for mineral phase and oxide composition fractions (Figure 2-3, Figure 2-4). The idea is that log odds transformations of percent compositions may distribute more normally than the raw percent compositions themselves. This is because percent compositions expressed as fractions may follow a Beta distribution and log odd is approximately Gaussian.

The log odds transformation is computed by first expressing percent composition as a decimal fraction (67 % becomes 0.67) and then computing log(comp)/log(1-comp). Computing a correlation coefficient to quantify straightness of normal probability plots for each phase and each oxide quantifies goodness of normal fit. Motivated by the normality enhancing performance of the transform demonstrated in these plots, we periodically throughout this work compare raw phase or raw oxide predictive performance to log odd(phase) or log odd(oxide) predictive performance via All Possible Subsets ACE.

Log odds transforming the phases results in no improvement for alite, ferrite, total aluminate, or bassanite. However, improvement is achieved for belite, cubic aluminate, orthorhombic

aluminate, periclase, anhydrite, gypsum, and calcite as seen in the correlation coefficient at the top of each plot labeled PPCC.

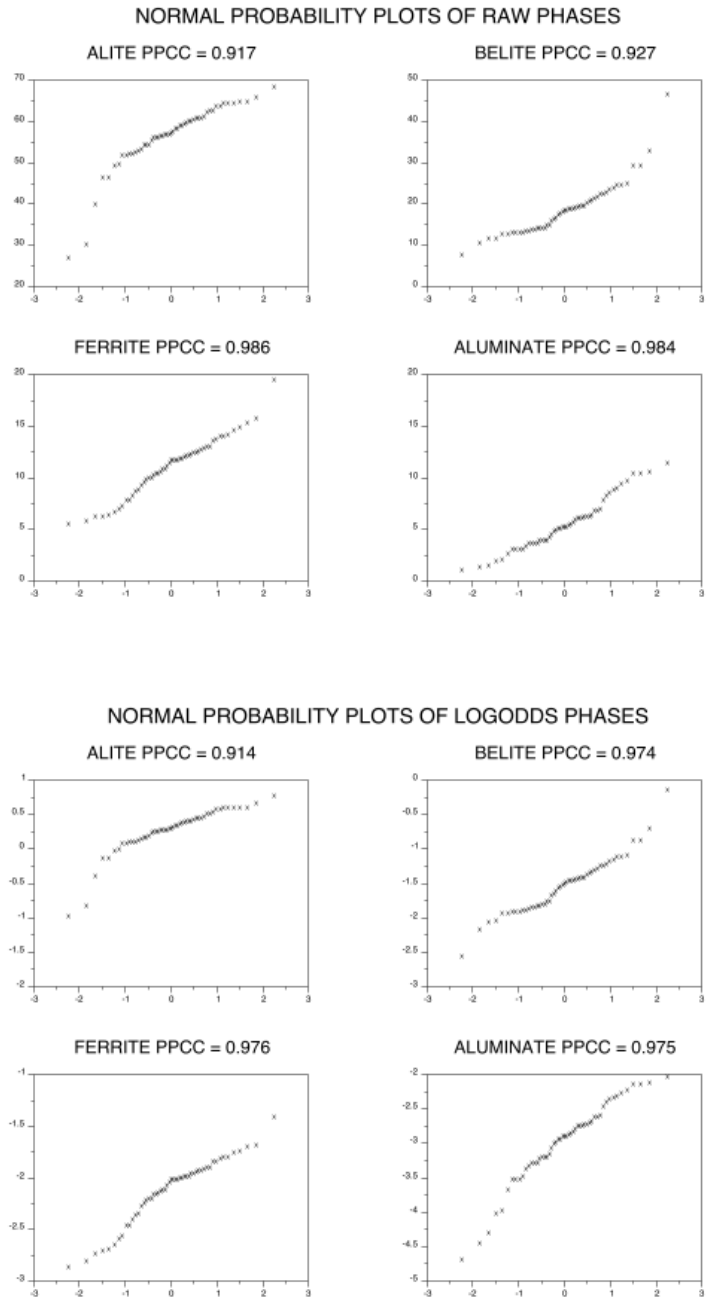

Figure 2-3- Normal probability plots of raw phases and log odds phases for alite, belite, ferrite, and aluminate indicates an improvement only for the belite phase. The Y-axis is sorted sample data and the X-axis is Gaussian median order statistic predictions of sample data

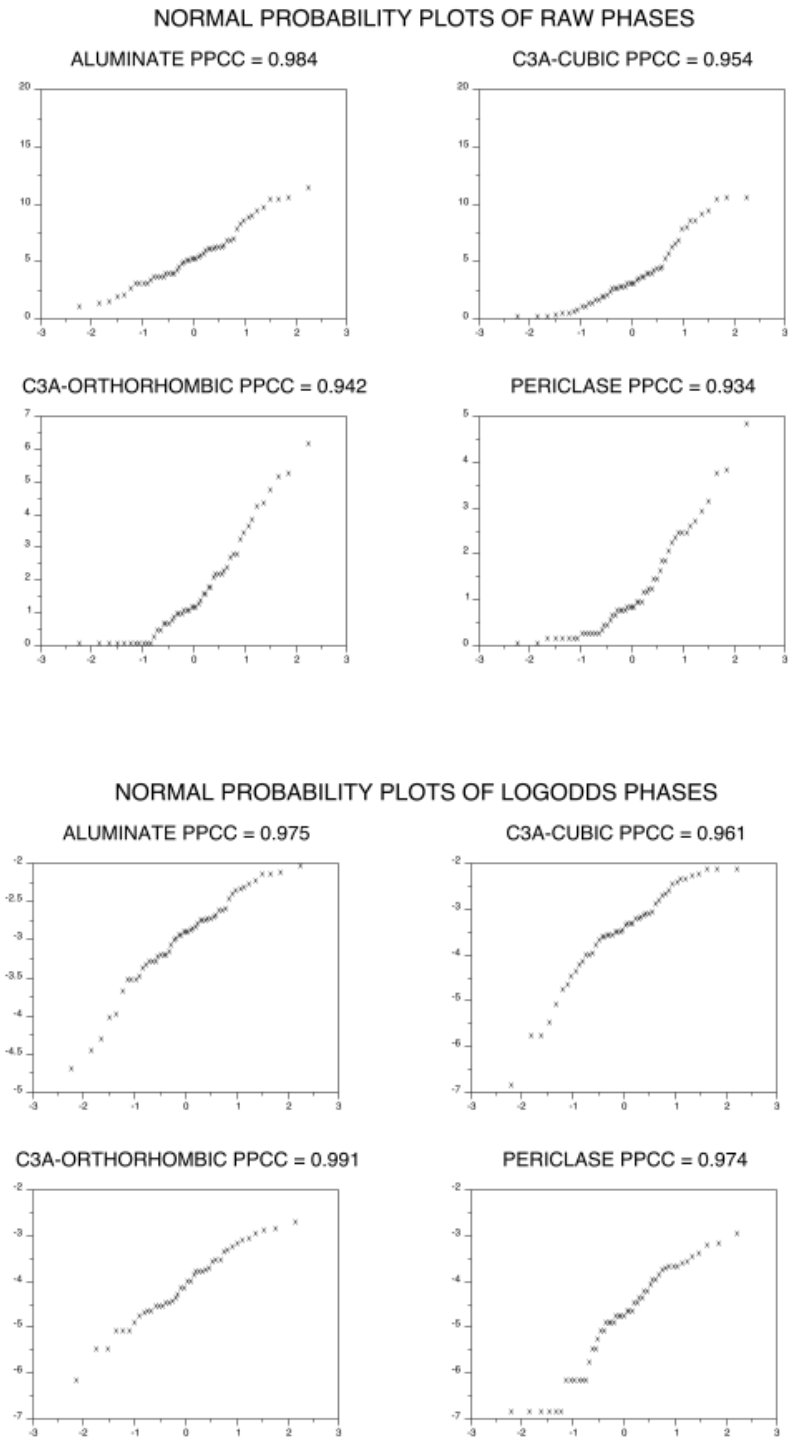

Figure 2-4- Normal probability plots of raw phases and log odds phases for total aluminate, aluminate forms and periclase show improvement for the aluminate forms and for periclase. The Y-axis is sorted sample data and the X-axis is Gaussian median order statistic predictions of sample data

2.3 All Possible Subsets Regression (APSR)

All Possible Subsets Regression regresses a response variable against all possible multilinear combinations of a pre-selected set of explanatory variables, typically screening for the best model fits from among the many fitted using criteria such as R^2 or $R^2_{adjusted}$ and exhibiting only the best models fit as judged by the goodness-of-fit criteria. In standard software[2], the many regressions performed in APSR are multilinear regressions with up to 32 input variables. While it is a linear tool, it is still an excellent screening device, and to be preferred to forward or backwards model subset selection.

So, for example, if there are 10 candidate explanatory variables, APSR software performs

$$\left(2^{10}-1\right) = {}_{10}C_1 + {}_{10}C_2 + ... + {}_{10}C_{10} \qquad Eqn.5$$

or 1,023 regressions, where "${}_{10}C_1$" refers to all possible 1-variable-at-a-time models, "${}_{10}C_2$" refers to all possible 2-variable-at-a-time-combination models, and so forth. Goodness-of-fit is typically assessed by Residual Sum of Squares (RSS). RSS is automatically the chief figure of merit for assessing goodness-of-fit of a model, coming as it does directly from the definition of least squares as searching for those model parameters that minimize RSS:

$$RSS = \sum_{all\,data} \frac{\left[HOH7 - Model\ \mathrm{Pr}ediction\right]^2}{\left(N-P\right)} \qquad Eqn.6$$

where the summation is taken over all the data points, N is the number of data points, P is the number of parameters being fitted (typically either number of explanatory variables or (number of explanatory variables + 1 if an additive constant is being fitted as well). Since an RSS of zero denotes a perfect error-less fit, the closer to zero the RSS from a real data fit the better [24].

The Coefficient of Determination, or R^2,

$$R^2 = \frac{\sum\left(pred\left(HOH7\right) - mean\left(HOH7\right)\right)^2}{\sum\left(HOH7 - mean\left(HOH7\right)\right)^2} \qquad Eqn.7$$

[2] BMDP: http://www.statistical-solutions-software.com/products-page/bmdp-statistical-software/

quantifies the improvement of the model-under-investigation's predictive performance over the naive mean model's (the most primitive model's) prediction. It is often referred to as quantifying the percent of variation in the data explained by the model, where the statistic is multiplied by 100 to express it in percent terms. Expressed in that way, it is clear that the closer the value of the R^2 to 1, or 100, the better. 1 or 100 connotes a perfect fit.

2.3.1 Misspecification: Model Bias

In comparing candidate models' performance, it is not enough to restrict attention to goodness-of-fit statistics. Because, in general, the more variables (parameters) one adds to a model, the better the fit will be. There is also the issue of protecting against model misspecification, or model bias, referring to the possible inclusion of too few or too many predictor variables in the model. For example, on any given pass of an all possible subsets routine (APSR or APSACE), if variables are being included that shouldn't be in the model, the variances (noise levels) of model coefficients and predictions increases and pushes predictions off target. On the other hand, if too few variables are being included in the model, the model will be biased and predictions pushed off target.

Since simply increasing the number of variables incorporated in a model will automatically tend to improve such goodness-of-fit statistics, reference must also be made to an adjustment for bias statistic such as Mallow's C_p, where the "bias" in question refers to the biasing of a model by the incorporation of too few or too many explanatory variables.

2.3.2 Mallow's C_p

The Mallow's C_p statistic is a very useful statistic for assessing misspecification when comparing multilinear models. It assesses the balance between bias due to too few variables and variance due to too many variables by evaluating:

$$C_p = \left\lfloor \frac{\sum\left(VAR\left(predHOH7\right)\right)}{\sigma^2} \right\rfloor + \left\lfloor \frac{\sum\left(BIAS\left(predHOH7\right)\right)}{\sigma^2} \right\rfloor$$

$$= p + \frac{\left(s^2 - \hat{\sigma}^2\right) \cdot (n - p)}{\hat{\sigma}^2} \qquad Eqn.9$$

where p is the number of parameters in the candidate model, s^2 is the residual mean square for the candidate model, and $\hat{\sigma}^2$ is an estimate of the true model variance.

A C_p value of close to p is indicative of least misspecification in a candidate model. The use of C_p in conjunction with a goodness-of-fit metric like R^2 is probably the single most reliable approach to "fitting blind", i.e. searching by statistical trial-and-error for a model where no scientifically derived candidate model exists [25,26,27].

2.3.3 Ensuring Model Validity

To ensure model validity generally, one seeks:

1. physical meaningfulness: use of scientifically meaningful variables in the model,

2. goodness-of-fit: in the RSS or R^2 metric,

3. parsimony: employing as few variables as possible in the model without overly under-fitting and without sacrificing too much goodness-of-fit,

4. avoiding misspecification,

5. cross validation: testing the goodness-of-model achieved on a set of training data by cross-validating against a non-training set (not performed in this study).

Since we will later use APSACE in this work rather than APSR, we do not directly assess (4) by some numerical criterion such as C_p. This in future work would probably be a remediable oversight. Since ACE models consist of (nonparametrically) transformed variables combined additively, it should be possible to define an analogue of the C_p statistic, which would automatically assess for misspecification.

As preparation and motivation for the numerical experiments reported on in this study, we subjected various combinations of the chief untransformed variable classes to APSR analysis. The results of the best for each combination are summarized in Table 2-3 for groupings of: 1) Oxides and Mineral Phases, 2) Mineral Phases and Finenesses, 3) Oxides and Finenesses, 4) Mineral Phases and Extras, 5) Oxides and Extras, 6) calcium sulfate phases, fineness, and extras. The group extras, as described before included variables such as Vicat, 3 day strength, and loss on ignition (LOI). From these combinations, key variables were selected, generating grouping 7 in Table 2-4, noted as the best of best reflecting the comparatively high R^2 and C_p statistic close to the number of variables. It is interesting to note that the combination of alite and aluminate that form the basis for the ASTM and AASHTO heat index equation does not appear in this table, although alite in combination with other phases, fineness measures, and oxides does. While the R^2 values are not strong for these models, they serve to indicate potentially interesting candidate variables for the subsequent step of nonlinear transformation of the data, where the correlations do improve markedly.

In each instance of class combination, the combination with the best statistics is reported. In categories where the mineral variables were included (five of seven), some form of calcium sulfate is selected as a key variable, specifically bassanite and/or anhydrite. Taylor [3] notes that aluminate and ferrite reactions are moderated by the presence of calcium sulfates, and that the form of calcium sulfate, as discussed earlier, affects its availability in solution, with bassanite being the most soluble, followed by gypsum and then anhydrite. Alite and ferrite have high potential enthalpies of hydration. This reflects the potential importance of the calcium sulfates in the reaction process that have not been accounted for in models developed to date.

Table 2-3- APSR regression results according to the untransformed variable class combinations.

Variable Class Combinations	C_p	R^2	$R^2_{adj.}$
1) alite, ferrite, bassanite, Fe_2O_3, TiO_2	5.55	.50	.45
2) alite, bassanite, span	-0.92	.32	.28
3) SiO_2, Fe_2O_3, MgO, SO_3, TiO_2	2.95	.37	.30
4) C_3Ao, anhydrite, bassanite, 3d str	2.52	.42	.37
5) SiO_2, Fe_2O_3, MgO, SO_3, TiO_2, Blaine	3.40	.39	.31
6) bassanite, 3d strength, Blaine	3.19	.35	.31
7) alite, ferrite, anhydrite, bassanite, Blaine, Fe_2O_3, TiO_2 (best of best)	7.44	.55	.48

When the particle size distribution variables (D10, D50, D90 and span) and Blaine fineness are included among the candidate predictor variables, one fineness variable is generally selected. Most often the Blaine fineness is chosen, reflecting the significance of the cement fineness in HOH7. In Group 1, alite, ferrite and bassanite are the mineral phases which, according to Taylor, [3], have a significant contribution to the heat output. The occurrence of bassanite may affect the hydration process through 1) acceleration of alite hydration and 2) suppression of aluminate and ferrite phase reactions. Fe2O3 and TiO2 significance may lie in their occurrence primarily in the ferrite phase, and to a lesser extent co-occurrence with the aluminates and belite [3].

In Group 2, with mineral phases and fineness, alite and bassanite were selected from the phase variables while a fineness variable “span” was selected; the Blaine fineness was not included for selection in this grouping. Group 3 included oxides and fineness as determined by PSD (no Blaine). A repeat of the Fe_2O_3 and TiO_2 occurs. In Group 4, mineral phases and other, we see the first selection of an aluminate phase with orthorhombic (alkali) tricalcium aluminate. The orthorhombic aluminate is the principal aluminate form in cements with low concentrations of aluminate, but does occur in small amounts in high-aluminate cements. The co-occurrence of the calcium sulfates anhydrite and bassanite may also reflect the controlling effect they have on the aluminate reactions. Group 5 examined the oxides, Blaine, Vicat, and 3 day strength. From this set we again see the Fe_2O_3 and TiO_2 and Blaine fineness. In Group 6 which included the

calcium sulfates, fineness, and extras of LOI, Vicat and 3d strength, we see that the calcium sulfate bassanite, strength, and Blaine are selected.

Extracting from each of the first six combinations the variables contributing most frequently to the best performing models yields the overall best performing model in "best of best." It is a 7-variable (8-parameter with the additive constant) model with a $C_p = 7.44$ and an R^2 of 55 %, which is the highest R^2 reported in the Table. The variables selected in the "best of best" are of interest, and will recur in what follows. The presence of one main structural phase (alite) along with sulfate phases anhydrite and bassanite should be noted. In what follows, ferrite and Fe_2O_3 will frequently appear as a selected co-occurring pair. The presence of Blaine signals that fineness may be important for HOH7 prediction, and Blaine may be the best predictor from among the fineness variables considered. Interestingly, TiO_2 recurs frequently in conjunction with Fe_2O_3 and ferrite in good models. We will see this again when we explore nonlinear transformations with APSACE-selected models with $R^2 > 0.90$.

2.4 Principal Components Analysis

The data were also subjected, combining different variable classes as inputs, to Principal Components (PC) Analysis and PC Regression. PC creates a new variable set, called the principal components that are linear combinations of the old variables. The PC's are uncorrelated/orthogonal with/to one another. The first PC establishes the "direction" in the new coordinate system in which the data exhibit maximal variability. The second PC establishes a second direction, orthogonal to the first, in which the data exhibits next (second) most variability. All of this is achieved via a straightforward eigenanalysis of either the covariance or correlation matrix of the original set of variables. We used correlation, as is done in many real data applications, because the normalization to a [-1,+1] scale provided by passage to correlation obviates interpretation problems with variables that have highly disparate magnitudes.

PC analysis is a popular multivariate analysis technique [26,28]. The mathematics and associated inferential (statistical) procedures are valid and interpretation of results is straightforward. However, an additional advantage sought by practitioners, and where the analysis becomes particularly cogent, is if the PC's, can be interpreted in a physically meaningful manner. The first PC is almost always a linear combination of the original raw variables that may have a simple physical interpretation. Other PC's may present signed "contrasts" between subgroups of components: so, a contrast between a longitudinal direction and a transverse direction might represent a kind of aspect ratio. The kinds of examples that are presented in textbooks do have this quality: the PC's are interpretable. Regrettably, in our many attempts to analyze the PC data considered here nothing ever presented itself to us as being simply interpretable. Linear transformations of the input variables ended up being interpretable as just that: linear transformations of the input variables, with no obvious heritable meaning. For this reason, we do not report any of the PC work done in this study.

2.5 Alternating Conditional Expectation

Alternating Conditional Expectation (ACE) is a technique that greatly extends the scope of classical multilinear model selection techniques [13,29]. Given HOH7 data with associated candidate predictor variables X_k, the ACE algorithm finds transformations of the predictor variables and of the HOH7 response variable that maximize the correlation between f(HOH7), the transformed HOH7, and $\sum g_k(X_k)$, the sum of the transformed predictor variables. The development and proof of the validity of the algorithm used to produce these transformations is complicated. ACE theory involves the eigenanalysis of conditional expectations interpreted as projection operators in Hilbert spaces of functions. The transformations are produced nonparametrically in the forms of pictures [31], graphs relating transformed to original variable for each of the variables including HOH7 response. Such pictures can be parametrically modeled either from first principles or by the use of automated software. As the transformation graphs can assume many forms, they need to be evaluated for credibility for incorporation into the predictive model. This evaluation is performed by analyzing each transformation picture for physicality and smoothness. Transforms that would appear to be approximately straight lines, low order polynomials, exponential or logarithmic functions, or circular functions are candidates for incorporation in a model. Transforms with severe inflection points or that have the appearance of multiple distinct behaviors adjoined, for example, might not be considered candidates for incorporation in the model, or might be modeled distinctly for each simple sub-model regime.

For example, Figure 2-5 depicts an ACE transformation example, with original variable on the X-axis and ACE-transformed variable on the Y-axis, C_3S, C_3Ac, C_3Ao, C_4AF all appear reasonably smooth, and possibly even linear. The cusp of the C_4AF graph and the blip on the C_3S graph can be thought of as reflecting possibly one or several aberrant data points in an otherwise clearly shaped transform, so these variables would not necessarily be removed from consideration. The graph of C_3A, on the other hand, presents much more of a challenge in trying to visualize the physicality of the transformed variable represented by the picture. Other graphs (periclase or bassanite) could perhaps be heurized as a linear ramp-up asymptoting or a curving fall-off followed by a second rise: viz, they can be visually smoothed into credible model components, but parameterizing the associated patterns would be considerably more challenging.

In using the ACE algorithm, one notices very quickly that ACE drives R^2 up dramatically. Modeling HOH7, if one feeds 7 or 8 candidate explanatory variables one can easily obtain R^2 values on the order of 0.80 to 0.95, possibly irrespective of how meaningful the incorporated variables are for the "true" prediction of HOH7. Some of the transformations will look smooth and easily parameterizable, but some will not. However, one should not simply drop the transformed variables that look unphysical and proceed with the ACE transforms produced on the remainder. This is for two reasons, both stemming from the same root cause: the set of ACE transforms produced from a set of variables depends critically on the whole ensemble of variables presented to the algorithm. So, running the algorithm and deleting variables from model inclusion after the fact results in (1) decrease in R^2 to uninteresting levels, and more importantly, (2) the remaining transformed variables are incorrect from an ACE procedure point of view, because rerunning ACE on the remaining variables alone may produce new or modified transform pictures, some possibly resembling the corresponding full-set pictures, but others not. One might use a property of persistence of transform shape for a particular variable throughout

changing combinations of variables as a strong argument for the significance of the persistent variable and its persistent transformation for inclusion in the predictive model; however, one cannot count on such persistence.

Since "re-ACE-ing" subsets of variables often gives different transformations from ACE-ing the original full set, it is clear that the appropriate way to proceed for optimal variable selection purposes is to perform an All Possible Subsets ACE (APSACE) of the original full set of variables. That is, match HOH7 response to one variable at a time, then two variables at a time, then three variables at a time, etc., ACE-ing each distinct combination and saving the outputs, the transform pictures and associated R^2 statistics, for each combination. That is what we have done, using a relatively simple nested loop type of S-plus code[3] [30,31]. Doing this guarantees that the ACE transform outputs for each distinct combination of variables is, mathematically meaningful and complete, and that the associated R^2 is correct. The code is sufficiently inefficient and produces sufficiently voluminous output, that we typically run it on at most 8

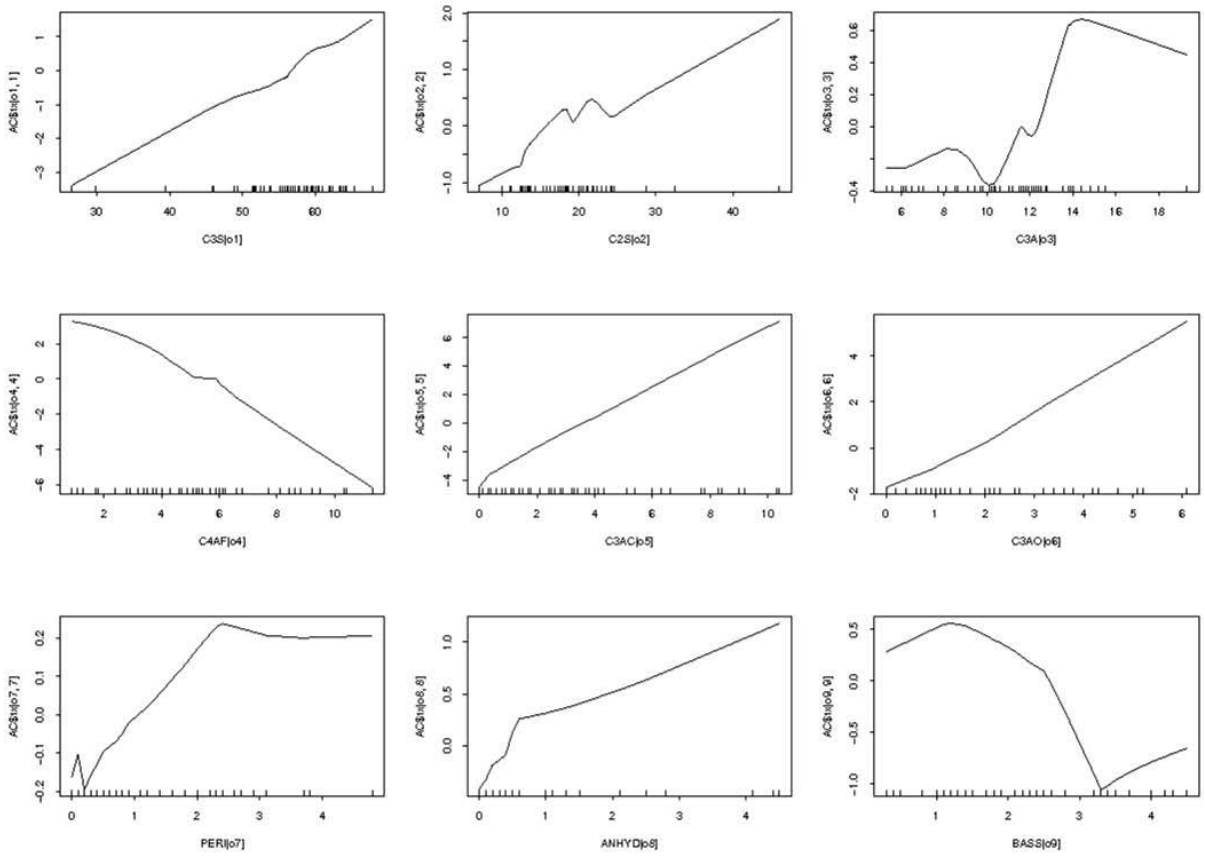

Figure 2-5- ACE Transforms where the x-axes are the original variables and the y-axes are the transformed variables

[3] http://spotfire.tibco.com/products/s-plus/statistical-analysis-software.aspx

variables at a time. For future work, one might consider augmenting the APSACE statistics with a C_p analog and then outputting only the highest R^2 with C_p closest to p combinations. In doing so, however, one might easily ignore interesting subsets of variables that are consistent contributors to good models and that consistently transform cleanly.

2.6 Synergizing Clusters for ACE Analysis

We examine clusters of variables that synergize to give high R^2 and smooth transformations through a three-step process where,

1. we run APSACE on the variable clusters defined earlier on in this document, and take special note of individual variables (e.g., phase or oxide variable) that recurrently contribute to high R^2 explanatory clusters and that "look physical" (smooth, model-able) in their ACE transforms,

2. we then run APSACE on combinations of variables (up to 8 at a time) that cross cluster boundaries: sulfate phases with fineness, again searching for high R^2 model contributors and smooth transforms, and

3. informed by the selections of steps 1 and 2, we seek models that incorporate variables from all the variable classes that yield high R^2 and smooth variable transforms.

2.6.1 Cluster Analysis

2.6.1.1 Phase clusters

“APSACEing” the raw phase cluster of variables by itself does not yield immediate insight into which stand-alone phase (in raw untransformed form) or combination of such yields the best explanatory power for HOH7D. However, certain observations can be made. Transforms involving alite and belite tend frequently to oscillatory behavior, often on a rising, falling, or shifting background line. Almost consistently throughout, C_3Ac behaves better than C_3Ao as an explanatory variable because its ACE transforms are nearly always smoother than those of C_3Ao. This is possibly the result of the low concentrations of the orthorhombic form of aluminate and the uncertainty in estimating its concentrations. This agrees well with observations made in the section devoted to aluminate ACE-ing. Some of the variables, e.g. periclase, can be extremely noisy and unpatterned, especially in regions of low periclase concentration. The lack of any obviously clear phase-based model can be taken as an indication that additional variables are missing. The fault lies in either poor R^2 or high R^2 combinations exhibiting poor ACE curves for phase components (Figure 2-6, Figure 2-7). Data transformation of the mineral constituents may improve these findings and the addition of additional physical variables (fineness, for example) may result in improved models.

Table 2-4- Selected phase clusters exhibit either a high R^2 and poor quality data transforms, or a low R^2 and smooth transforms.

R^2	Phase Cluster
.42	alite
.68	alite belite
.73	alite C_3Ac
.86	alite belite C_3Ac C_3Ao ferrite

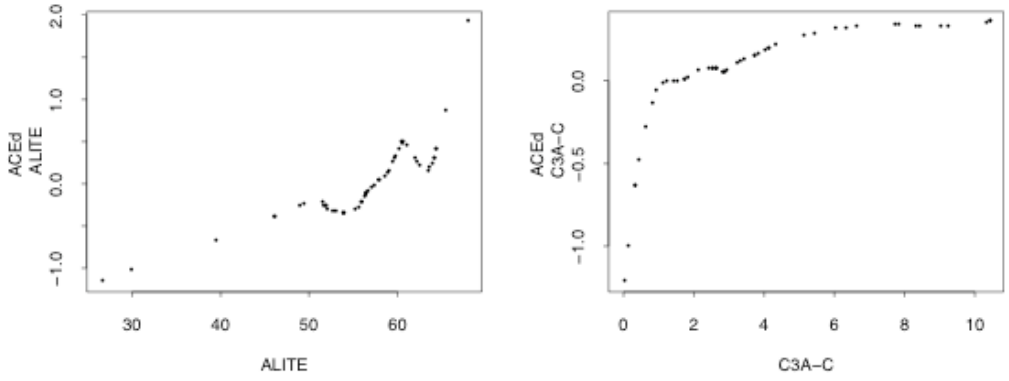

Figure 2-6- Alite and cubic aluminate provide a relatively high R^2 but poor transforms, particularly for alite.

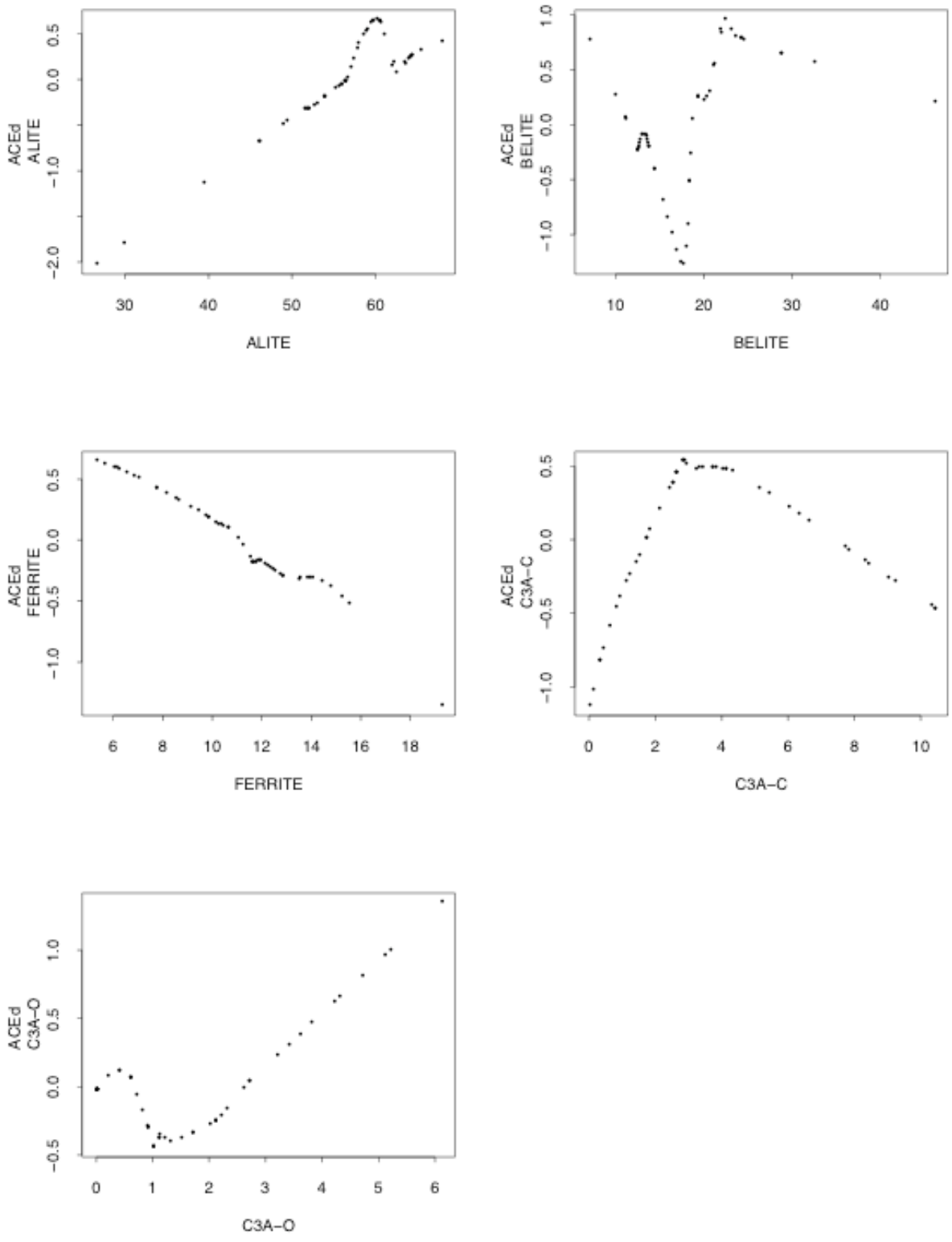

Figure 2-7- Including the primary cement phases results in a high R^2 (0.86), but rough transform curves for several of the constituents.

2.6.1.2 Oxide Clusters

The oxide clusters illustrate the power of ACE to pull out unusual combinations of predictor variables with very high R^2 for HOH7 prediction. Raw (untransformed) oxides were ACE'd to predict raw (untransformed) HOH7. The results are summarized in Table 2-5 with transform plots in Figure 2-8 through Figure 2-11. The easily recognized recurring high contributors were CaO, MgO, TiO_2, and SO_3. The CaO-TiO_2 cluster is combined with each of the other oxides one at a time. The MgO-TiO_2 cluster is combined, one added variable at a time, with each of the other oxides. Multiple features of these results are quite striking. One is the fact that, suitably (ACE-) transformed, one can achieve 93 % explanatory power for HOH7 with the two variables MgO and TiO_2, alone. Equally striking is that with the triples CaO-MgO-TiO_2 and CaO-SO_3-TiO_2 one can achieve 96 % explanatory power.

Going to the corresponding ACE pictures, however, presents a puzzle. The MgO and TiO_2 transformations appear to be relatively flat, hovering around zero, unstructured when paired by themselves. When coupled with another player, as in CaO-TiO_2-MgO, (R^2 = 0.96), the MgO has a nicely inflected broadly unimodal transformation shape, while TiO2 and CaO continue to remain flat, near zero, and uninflected. Is this a numerical artifact of ACE? Or are these pictures and R^2 values telling us something about an underlying relationship between these parsimonious oxide combinations and HOH7 prediction?

It is striking that with the exception of the MgO-TiO_2 doublet, the most powerful explanatory combinations are triplets: CaO-TiO_2-SO_3, CaO-TiO_2-MgO, MgO-TiO_2-SO_3.

Table 2-5- Oxide clusters for combinations of TiO_2 and either CaO or MgO with other oxides.

R^2	Oxides (CaO, TiO_2 plus)
.81	CaO TiO_2
.96	**CaO TiO_2 SO_3**
.78	CaO TiO_2 Al_2O_3
.86	CaO TiO_2 SiO_2
.82	CaO TiO_2 Fe_2O_3
.96	**CaO TiO_2 MgO**

R^2	*Oxides (MgO, TiO_2 plus)*
.93	**MgO TiO_2**
.93	**MgO TiO_2 SO_3**
.92	MgO TiO_2 Al_2O_3
.87	MgO TiO_2 SiO_2
.91	MgO TiO_2 Fe_2O_3
.96	**MgO TiO_2 CaO**

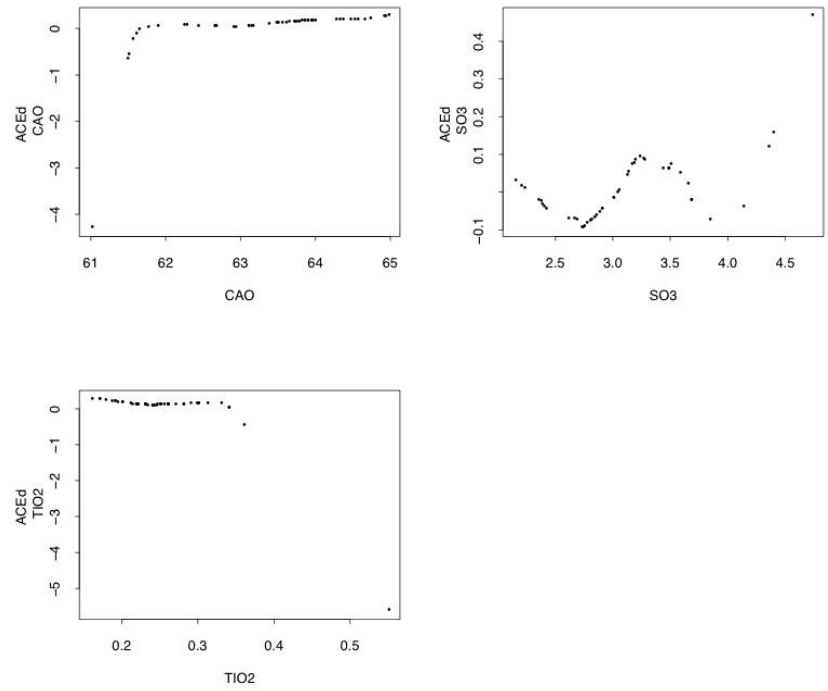

Figure 2-8- Oxide cluster transforms for R^2=0.96 combination of CaO, TiO_2, and SO_3

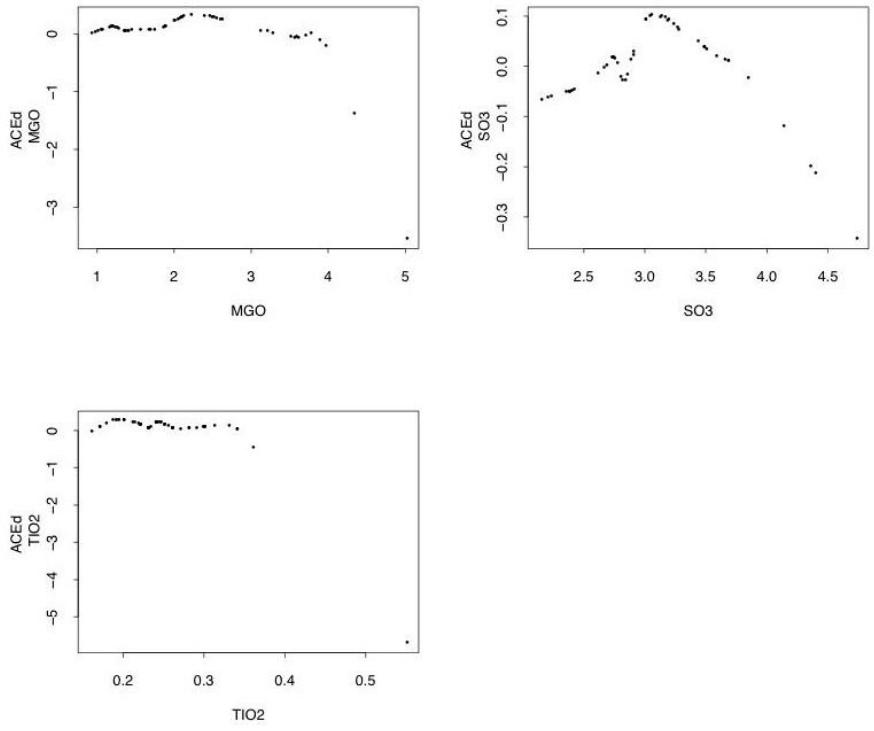

Figure 2-9- Oxide cluster transforms for R^2=0.96 combination of MgO, TiO_2, and SO_3

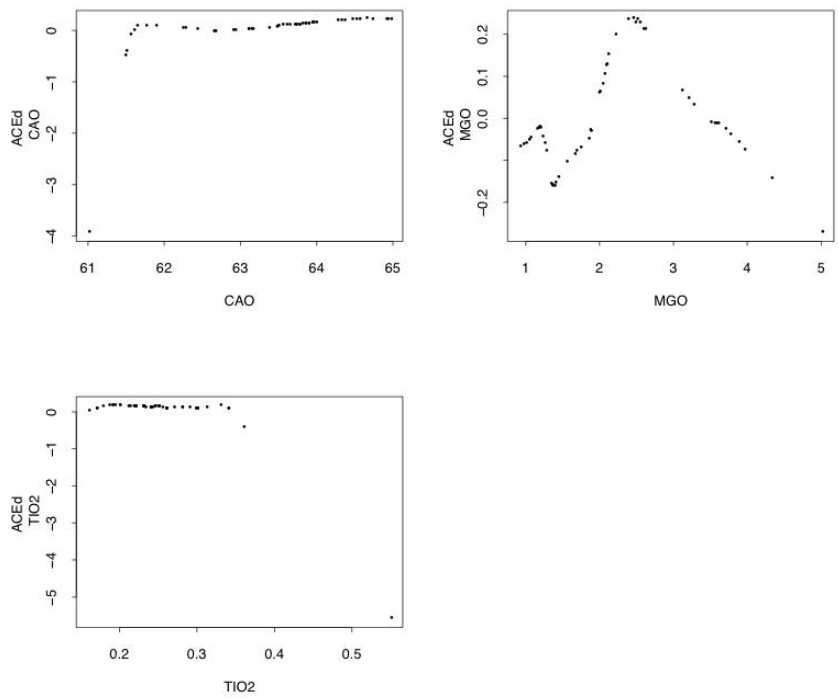

Figure 2-10- Oxide cluster transforms for R^2=0.96 combination of CaO, TiO_2, and MgO

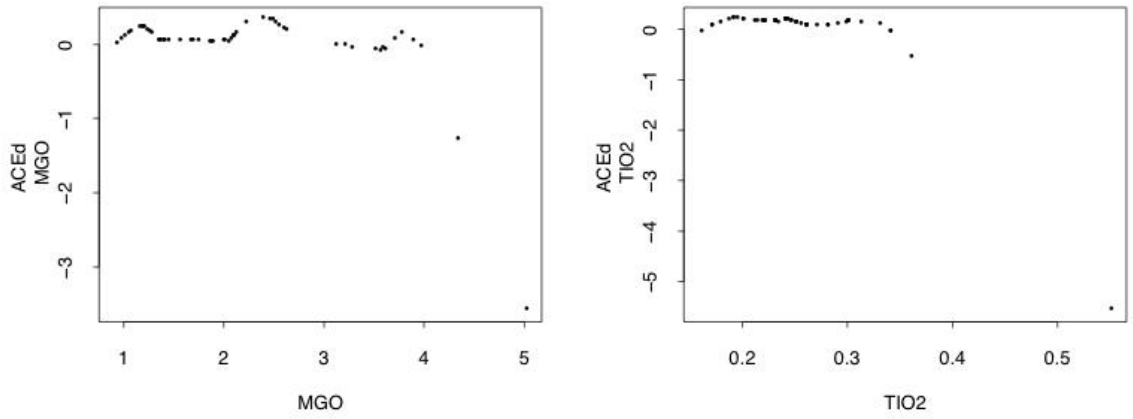

Figure 2-11- Oxide cluster transforms for R^2=0.93 combination of MgO and TiO_2

2.6.1.3 Aluminate Clusters

Aluminate, C3A-c, and C3A-o were deliberately incorporated into the data so we could compare their relative performance in predicting HOH7, as their enthalpy of hydration provided in Table 1-1 is the highest of any of the clinker phases. Furthermore, previous studies found that the aluminate form appeared to significantly affect heat [9]. Aluminate is the sum of the cubic (c) and orthorhombic (o) forms of tricalcium aluminate. They were run in APSACE without, and in combination with, a structural phase, belite, and a sulfate phase, bassanite, to ascertain which component might demonstrate the most modeling power. Results are summarized in Table 2-6. From the Table it is immediately clear that the combination of aluminate with belite and bassanite, with an R^2 of 0.81 and a somewhat smooth ACE transform (Figure 2-12), easily dominates the individual predictive performance of the two component aluminate analytes. Aluminate's nearest competitor, in terms of the two aluminate subcomponents, is C3A-c, which gives a 0.60 R^2 and rough transform, in combination with belite and bassanite. We conclude that there is no modeling advantage in using subconstituents C3A-c or C3A-o over using aluminate itself.

Table 2-6- Total aluminate, cubic and orthorhombic plus structural phase cluster

R^2				
.54	Aluminate			
.55		C_3A-c		
.32			C_3A-o	
.44		C_3A-c	C_3A-o	
.81	**Aluminate**			**Belite, Bassanite**
.60		C_3A-c		Belite, Bassanite
.55			C_3A-o	Belite, Bassanite

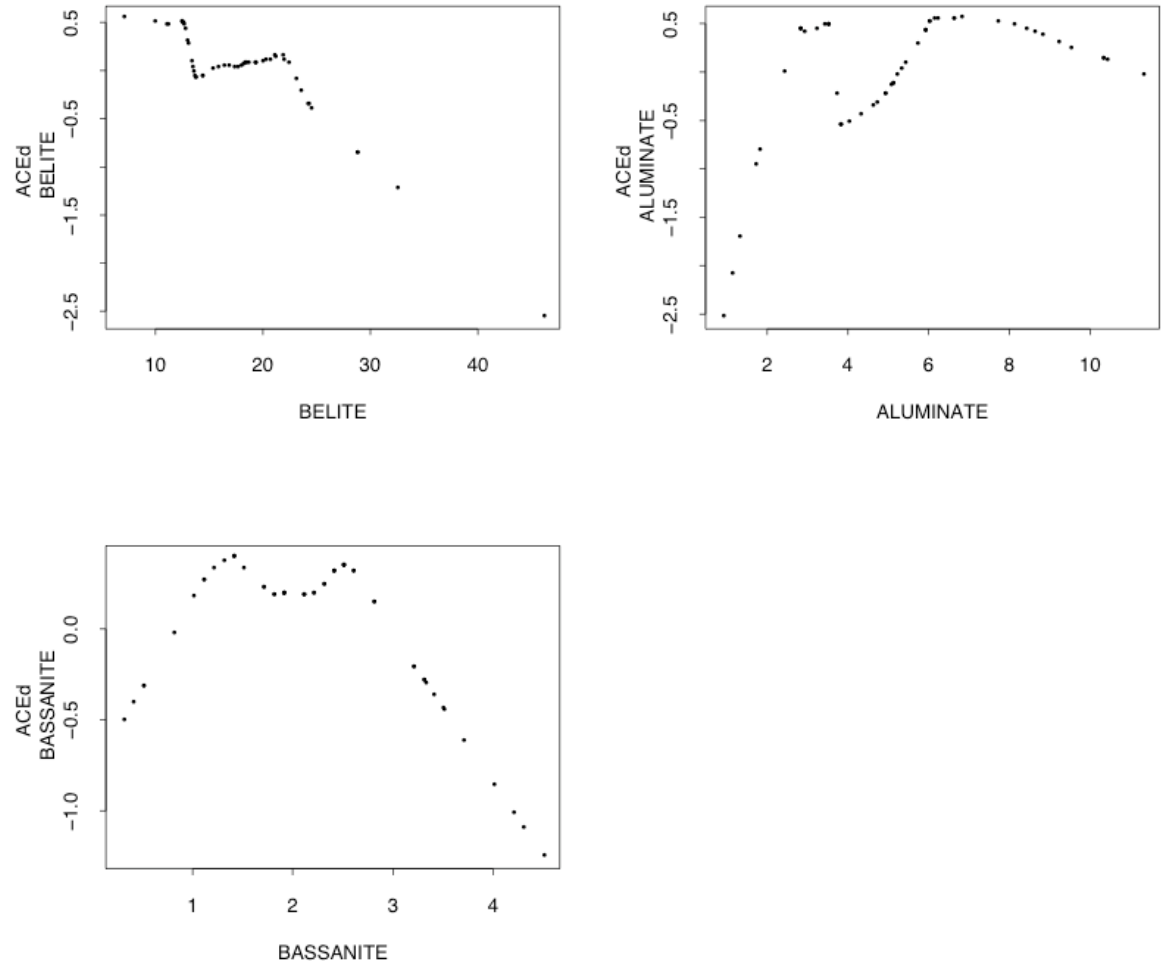

Figure 2-12- Transforms of aluminate phases plus structural phases (belite and bassanite) result in a 0.81 R^2 the total aluminate, belite, and bassanite exhibiting fairly smooth transformed curves.

2.6.1.4 Fineness cluster

Three measures of fineness or particle size distribution were run against a "background" of alite, ferrite, and bassanite for predictive power comparison. Blaine and Span (a function of D10, D50, and D90 reflecting the width of the particle size distribution) were taken in untransformed form as given. D10, D50, D90, however, were also combined into a possibly more meaningful variable in the following way. Mean and standard deviation approximations for the distribution represented by the D percentiles:

$$M = (D10+D50+D90)/3$$

$$S = (D90-D10)/2.56$$

were combined into a coefficient of variation

$$CV = 100*(S/M)$$

expressed as a percent.

APSACE results are shown in Table 2-7, with ACE transforms shown in Figure 2-13 and Figure 2-14. The results are unequivocal: Blaine is the most powerful predictor from among the 3 fineness measures, both in stand-alone mode (R^2 = 0.54 with smooth ACE transform) and in conjunction with alite and bassanite (R^2 = 0.74 with relatively smooth transforms for Blaine, alite, bassanite). This makes good sense, as described earlier in this report, since Blaine is an indirect measure of overall cement particle surface area, which a fortiori must correlate strongly with heat release. From all possible subsets of 6 variables run, ferrite tended to be relatively noncontributory when taken in conjunction here the finenesses. It is interesting to note that the stand-alone Blaine transform strongly resembles pictorially the Blaine transform in conjunction with alite and bassanite. A consistent feature of the Blaine transforms in most of the clusters that include Blaine fineness is a kink at approximately 3700 Blaine, which is roughly in accordance with the upper range of Type I, II, and V cements compared to the finer Type III cements [32]. This suggests that the cement Types should be modeled separately.

Table 2-7- All Possible Subsets ACE for Fineness and Phase

R^2	*fineness*		*phase*
.54	**Blaine**		
.32	CV		
.19		Span	
.74	**Blaine**		**alite, bassanite**
.59	CV		alite, bassanite
.56		Span	alite, bassanite

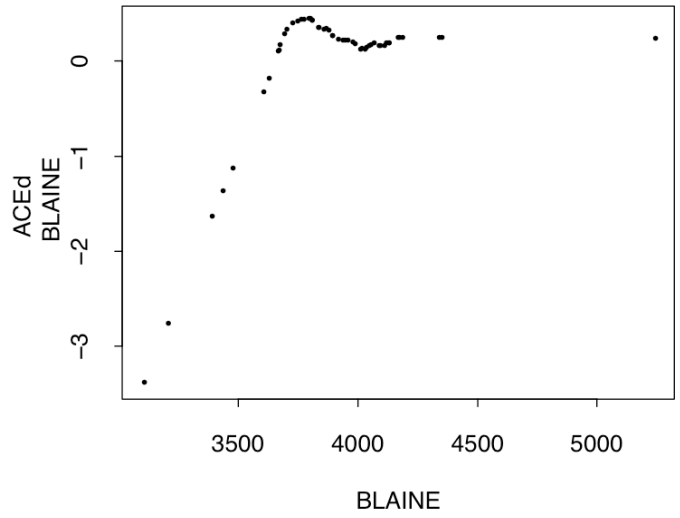

Figure 2-13- ACE Transform for Blaine fineness that yields an R^2 of 0.54

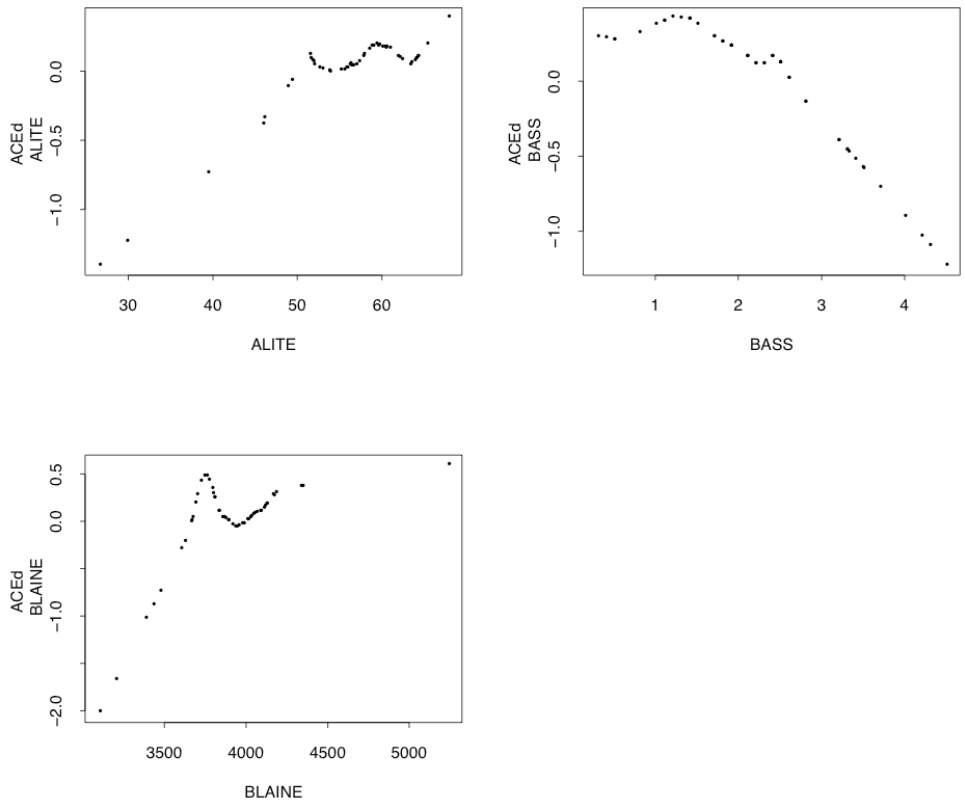

Figure 2-14- ACE Transforms for Blaine fineness, alite, and bassanite that yield an R^2 of 0.74

2.6.1.5 Sulfate Cluster

To examine relative predictive powers of sulfate phase variables, SO_3, anhydrite, bassanite, and gypsum were run in conjunction with "background" variables belite, Blaine, ferrite, Fe_2O_3. Results focusing on (belite + Blaine) combined with one stand-alone sulfate compound at a time are summarized in Table 2-8. While gypsum gives the highest R^2 (0.34) in stand-alone mode, bassanite, gypsum, and SO_3 give comparable R^2's (0.69, 0.73, 0.72) in conjunction with belite + Blaine (Figure 2-15 through Figure 2-18). However, bassanite stands out for the smoothness of its ACE transform. The remarkable smoothness of the bassanite transform can be compared, for example, to the generally rougher anhydrite transform in richer, better-fitting models from this large APSACE set of outputs. Compare Figure 2-19 and Figure 2-20, the fit of (bassanite + gypsum + Fe_2O_3) with (belite + Blaine), giving an R^2 of 0.88, with the fit of (anhydrite + gypsum + SO_3) + (belite + Blaine), also giving an R^2 of 0.88. The smoothness of the bassanite transform seen here is typical of bassanite's transform behavior across many of the bassanite-including models drawn from this experiment, and across cluster-crossing experiments as well.

Table 2-8- Sulfate Cluster APSACE Results

R^2	*Calcium Sulfate Phase*			*Oxides*	*Belite, Blaine*
.30	Anhydrite				
.31		Bassanite			
.34			Gypsum		
.27				SO_3	
.60	Anhydrite				Belite, Blaine
.69		Bassanite			Belite, Blaine
.73			Gypsum		Belite , Blaine
.72				SO_3	Belite, Blaine
.88		**Bassanite**	**Gypsum**		**Belite, Blaine**
.88	**Anhydrite**		**Gypsum**	**Fe_2O_3**	**Belite, Blaine**

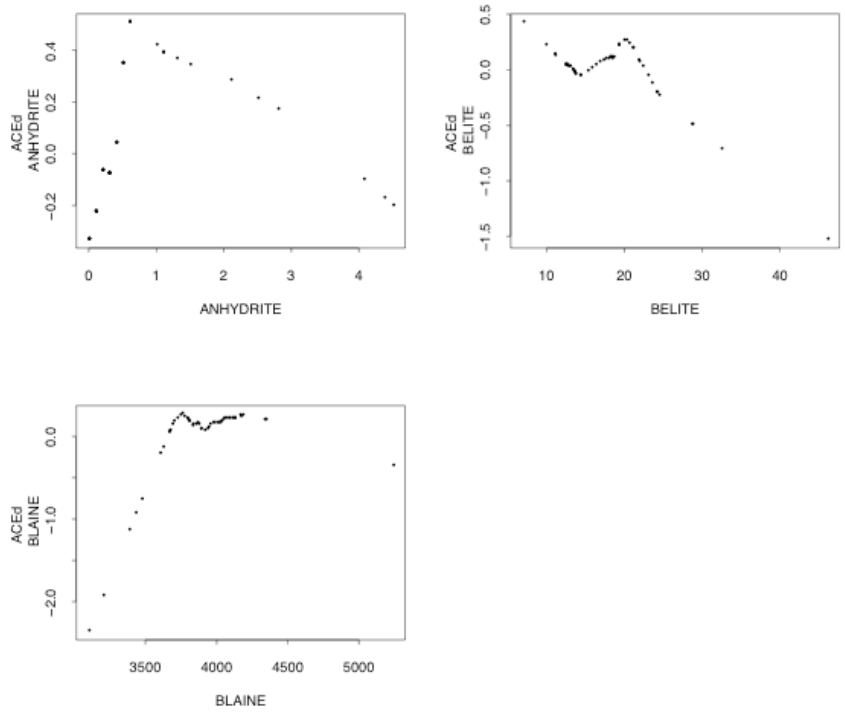

Figure 2-15- ACE transform for anhydrite, belite, and Blaine with an R^2 of 0.60.

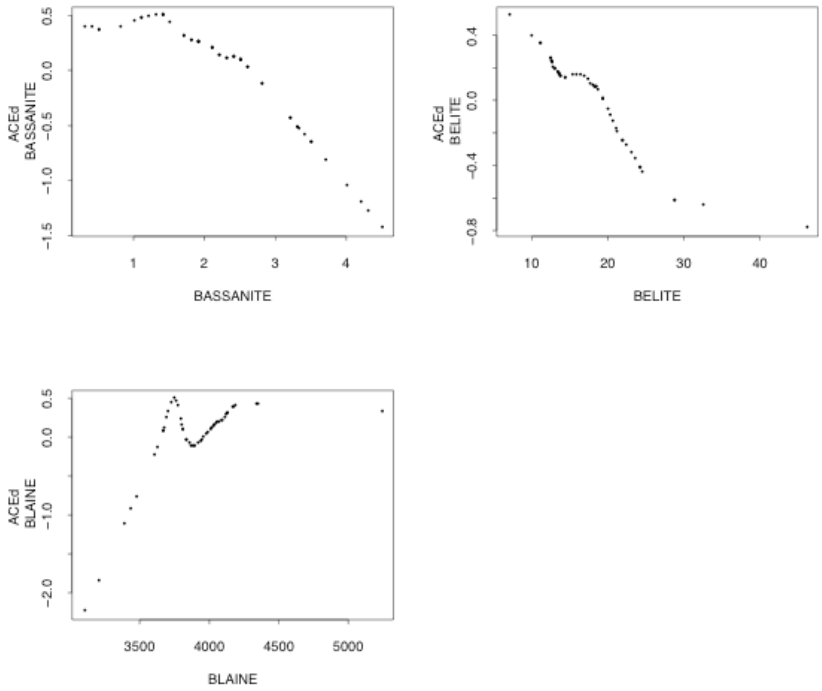

Figure 2-16- ACE transform for bassanite, belite, and Blaine with an R^2 of 0.69.

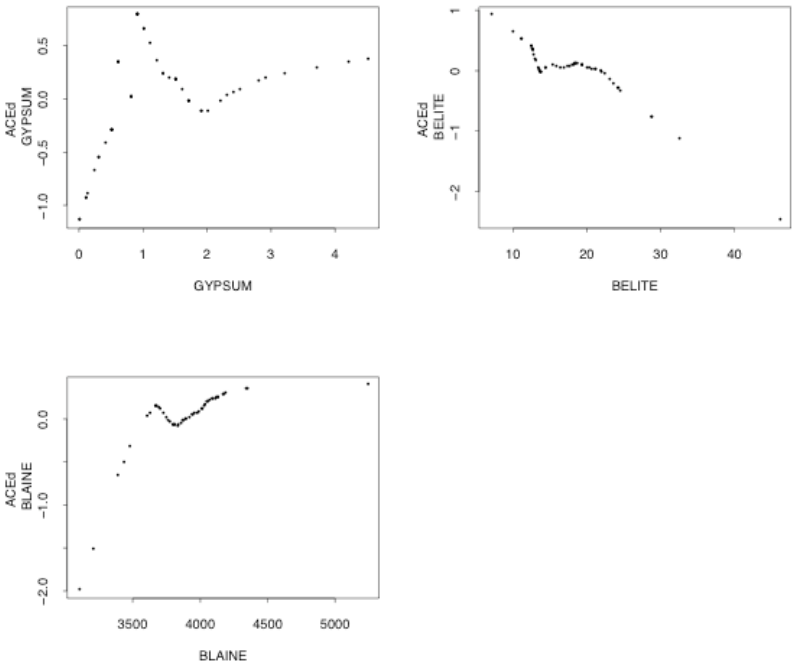

Figure 2-17- ACE Transforms for gypsum, belite, and Blaine with an R^2 of 0.73 are rough.

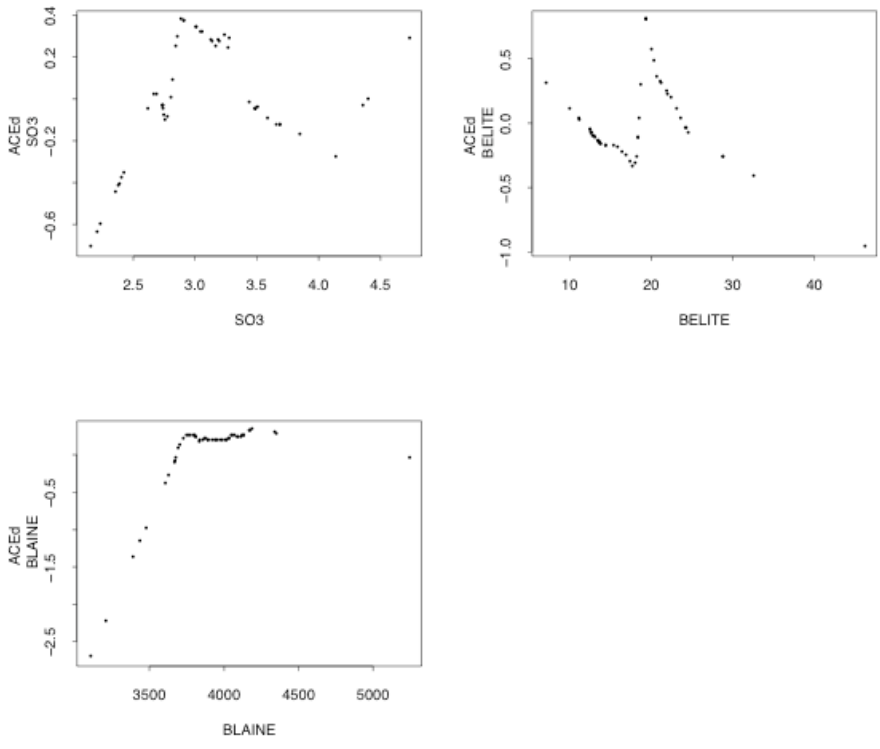

Figure 2-18- ACE Transforms for SO_3, belite, and Blaine with an R^2 of 0.72 are rough.

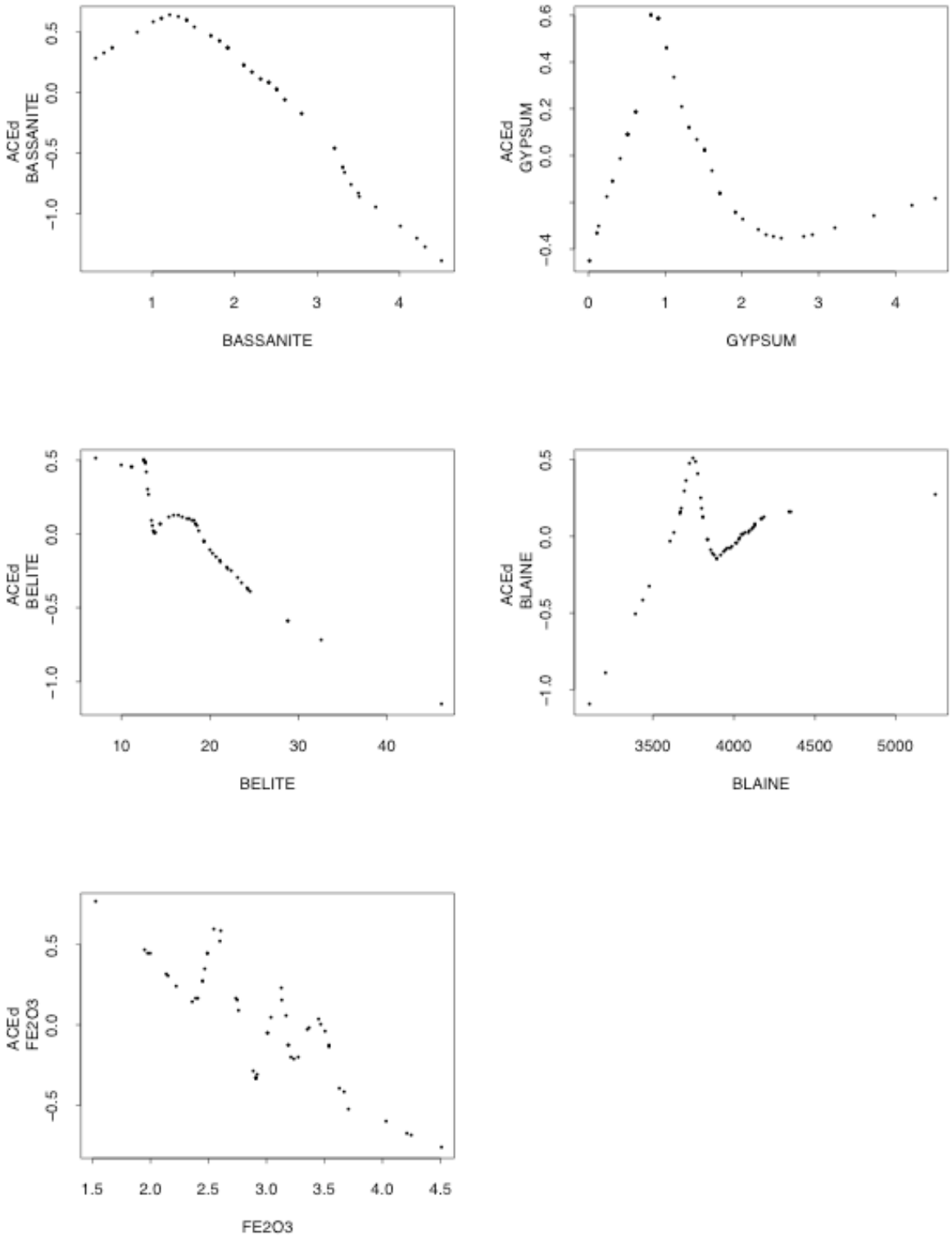

Figure 2-19- (bassanite + gypsum + Fe_2O_3) with (belite + Blaine), giving an R^2 of 0.88, illustrates the smooth bassanite ACE transform.

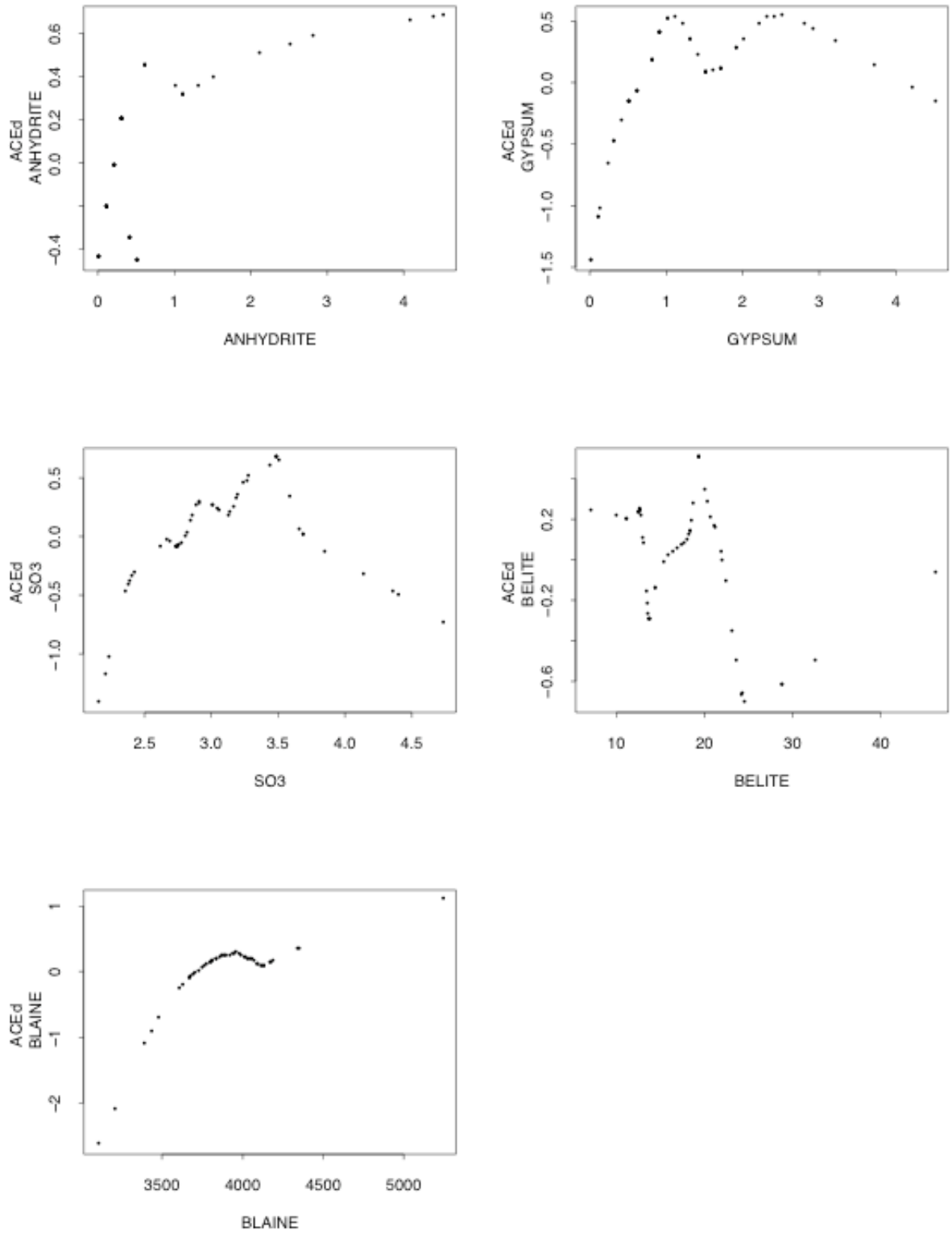

Figure 2-20- (anhydrite + gypsum + SO3) + (belite + Blaine), also giving an R2 of 0.88, but the ACE transforms are much rougher.

2.6.1.6 Extras cluster

To ascertain if the "extra" variables (LOI, 1/Vicat, Calcite, strength) have any HOH7 predictive power, they were run in APSACE mode with belite, Blaine, and bassanite. The results are displayed in Table 2-9. In stand-alone mode, while 1/Vicat by itself gives an R^2 of 0.66, the ACE transform picture is highly nonstructured and uninformative (Figure 2-21). Strength, interestingly, gives an R^2 of 0.58 in stand-alone mode, with a relatively clean and smooth unimodal transform profile.

Combining these variables with (belite + bassanite + Blaine), however, tells a different story. The (belite-Blaine-bassanite) combination by itself gives an R^2 of 0.71, with smooth transforms for each of the three participating variables (Figure 2-22). However, when extra variables are added one at a time, in three of the cases, the R^2 decreases. For calcite however, it increases to 0.78. Note that in the three cases of decrease, either one or two of the participating variables in the 4-variable models exhibit suboptimal, rough or sparse, ACE profiles (Figure 2-23). Only the (calcite+belite+bassanite+Blaine) (R^2 = 0.78) combination appears to have smooth ACE transforms.

Table 2-9- Extra variables with phase and fineness variables belite, bassanite, and Blaine.

R^2	*extras*			*phase, Blaine*		
.66	**1/VICAT**					
.44		Calcite				
.58			Strength			
.71				**Belite**	**Bassanite**	**Blaine**
.57	1/VICAT			Belite	Bassanite	Blaine
.78		**Calcite**		**Belite**	**Bassanite**	**Blaine**
.58			Strength	Belite	Bassanite	Blaine

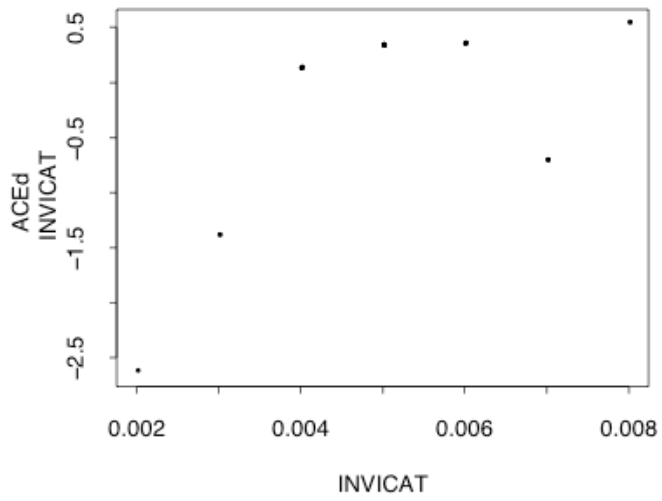

Figure 2-21- The 1/Vicat ACE curve yields a relatively high R^2 but exhibits a rough structure that would be difficult to model, unless the (overlaid) point(s) at 0.007 can be ignored.

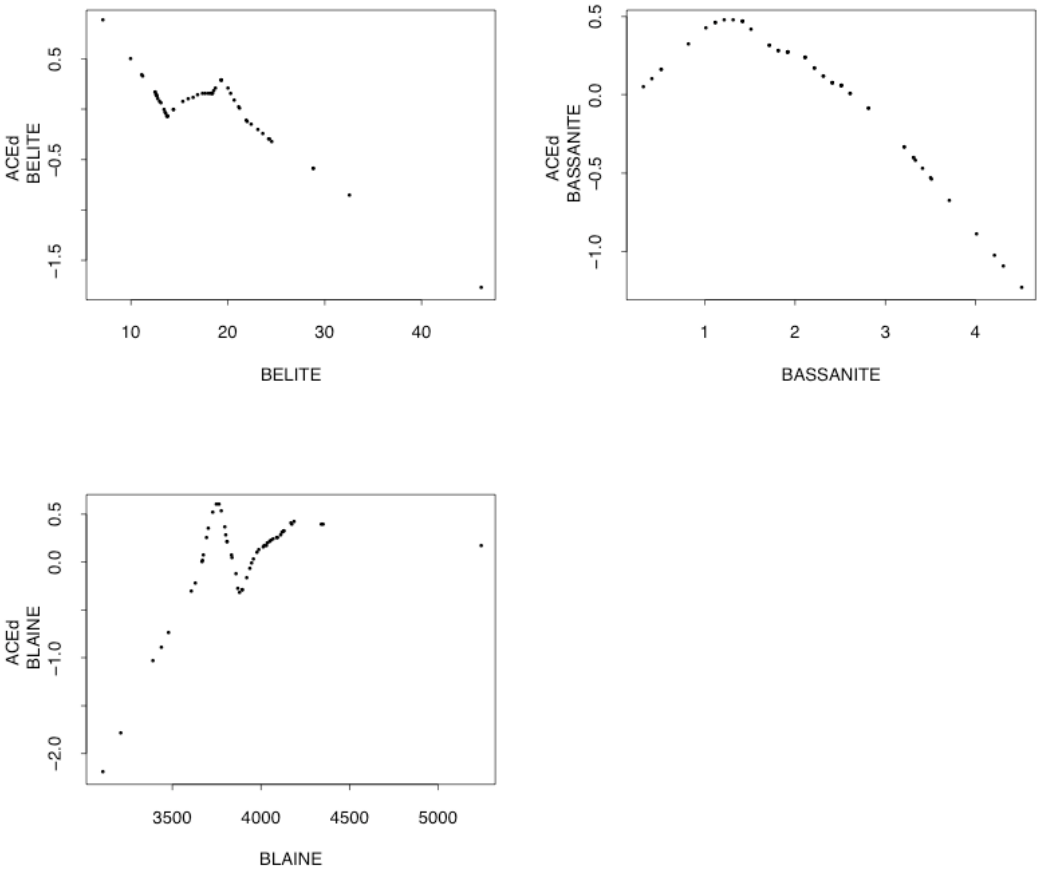

Figure 2-22- The belite, bassanite, Blaine variables yield an R^2 of 0.71 but a rough belite curve and a break in the Blaine curve at around 3800.

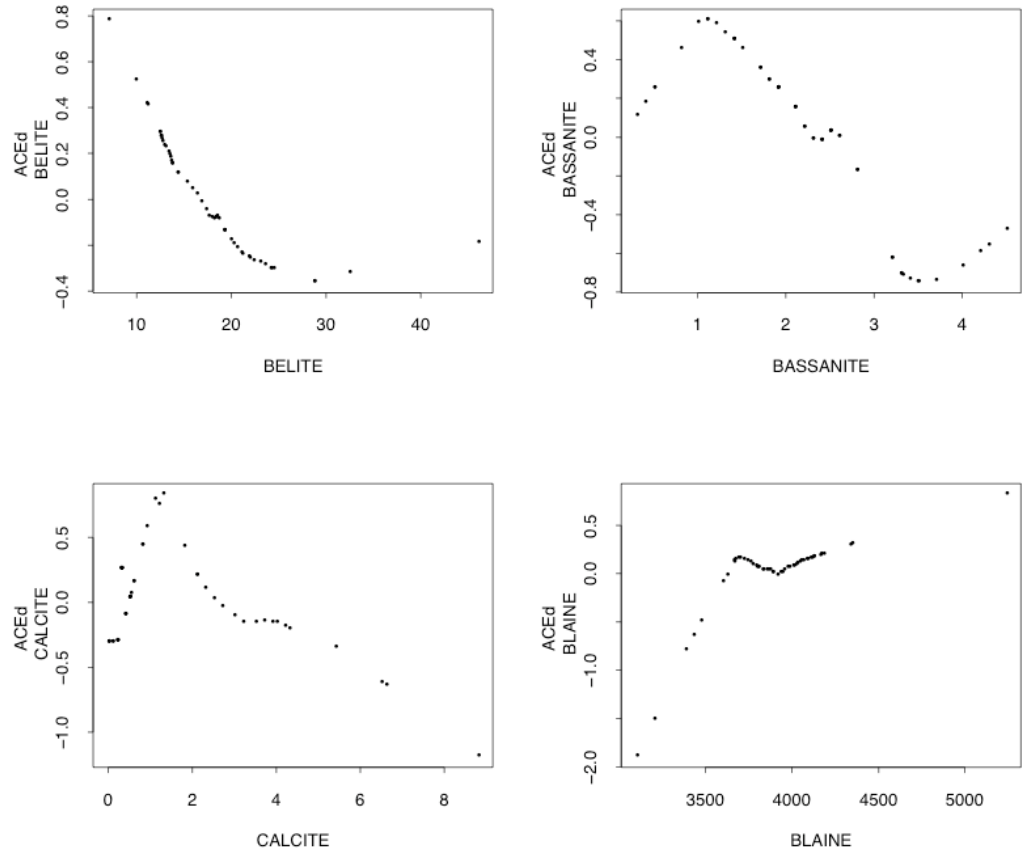

Figure 2-23- Belite, bassanite, Blaine and calcite ACE output gives an R^2 of 0.78 and reasonably smooth curves

2.6.2 Explicit Parameterization of ACE outputs: an example

Running APSACE on a candidate predictor set that includes aluminate, ferrite, bassanite, Blaine, and 1/Vicat yields four different high-R^2 subclusters. In Table 2-10, each subcluster is labeled with its R^2, and the individual variables within the subclusters are labeled with a description of the appearance of the associated ACE transformation. Multiple observations can be made from this Table and Figure 2-24 through Figure 2-27. It is interesting that the ACE transform for 1/Vicat is close to linear for all subclusters until bassanite is added. This is an example of the kind of persistence of inclusion in high R^2 models coupled with persistence of transformation shape that was mentioned above. It is, in contrast, interesting to note how very sensitive the shape and overall appearance of the aluminate transform is to the co-inclusion or -exclusion of the other variables from the parent cluster. Looking at the alternatives, and taking into account the need for parsimony (the fewer the modeling variables the better) and smoothness, the ferrite – Blaine – (1/Vicat) combination seem to be the most attractive for this set of candidate variables.

Why might a model incorporating these variables be good? Blaine fineness is an indirect measure of the total surface area of cement powder particles, which should certainly affect heat

release. Ferrite occurs as medium- to fine-grained crystals that should exhibit greater surface area than some of the other "structural" mineral phases and its enthalpy of complete hydration given in Taylor [3] of -418 $kJkg^{-1}$ ranks just below that of alite. However, at 7 days, the reaction coefficient is almost twice that of alite (Table 1-1). And, as explained above, high HOH7 may also promote a rapid setting time, associated with a low Vicat.

For illustration purposes only, we show how one of the APSACE-suggested models of Table 2-10 can be parametrized by using automated software. The second example from Table 2-10 combines relatively smooth or linear transforms with a relatively high R^2 (0.88), the plots of each transformed variable being shown in Figure 2-28 through Figure 2-30. The code (TableCurve2D) fits up to 3000 parametric models to X Y input, and sorts them according to goodness-of-fit assessed by R^2. Classes of models fit include multilinear, polynomials in x,(1/x), rational functions (quotients of polynomials) in x,(1/x), circular functions and Fourier-type sums, log, exp. and polynomials in log, exp, Chebychev polynomials, and simple classes of nonlinear models. From a large menu of potential parametric fits to any given ACE transform, we select the model that seems to give the best combination of high R^2, visual goodness-of-fit, and is parsimonious in the number of parameters used in the parameterization.

If ACE(HOH7) is model-able either as a simple linear function (which in practice it is often) or more generally as an explicitly invertible function (e.g., log-to-exp or sin-to-arcsin), then we obtain a completely explicitly parametrized model for HOH7 in terms of ferrite, bassanite, and (1/Vicat). The three-variable ACE model developed here, without the explicit parameterization, gives an $R^2 = 0.88$ on the data modeled. The two-variable Poole gives an $R^2 = 0.76$. However, the model considered here for illustrative purposes only, is just one among dozens of high- R^2 ($R^2 > 0.90$) candidate models that APSACE has identified in this work.

Table 2-10- Small clusters of variables that provide high R^2 and smooth transformations. Individual variables and a description of the transform shape are provided.

R^2					
.78	Aluminate			Blaine	1/VICAT
	smooth, with edges			*smooth with cusp*	*almost linear*
.88		**Ferrite**		**Blaine**	**1/VICAT**
		smooth with dropoff		***smooth with asymptote***	***almost linear***
.90	Aluminate	Ferrite		Blaine	1/VICAT
	rough	*line with tent*		*smooth*	*almost linear*
.86	Aluminate	Ferrite	Bassanite	Blaine	1/VICAT
	multi-modal	*rough*	*high-low disconnect*	*rough*	*tent*

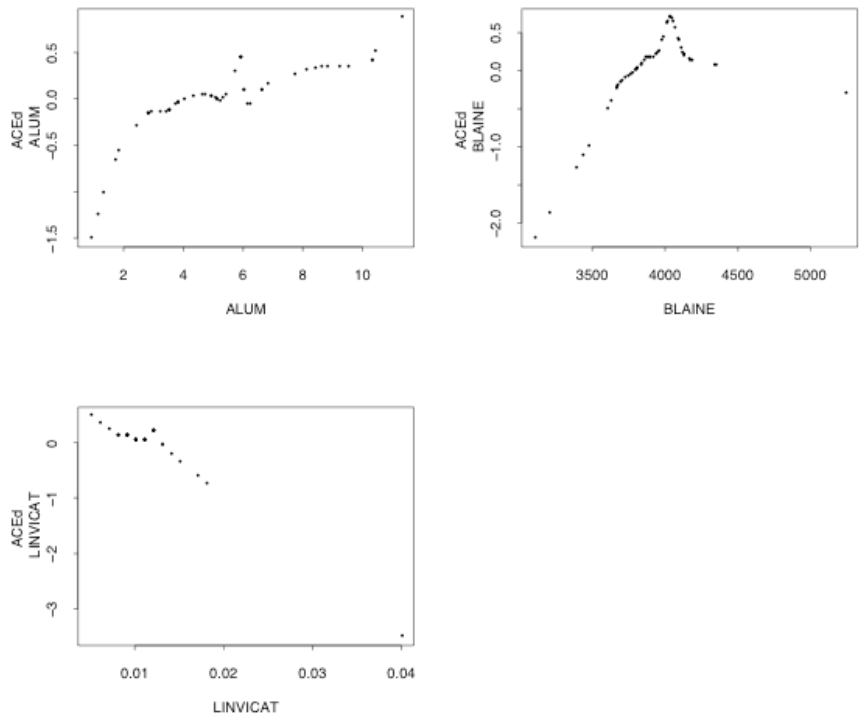

Figure 2-24- ACE transforms for aluminate, Blaine, 1/Vicat with an R^2 of 0.78.

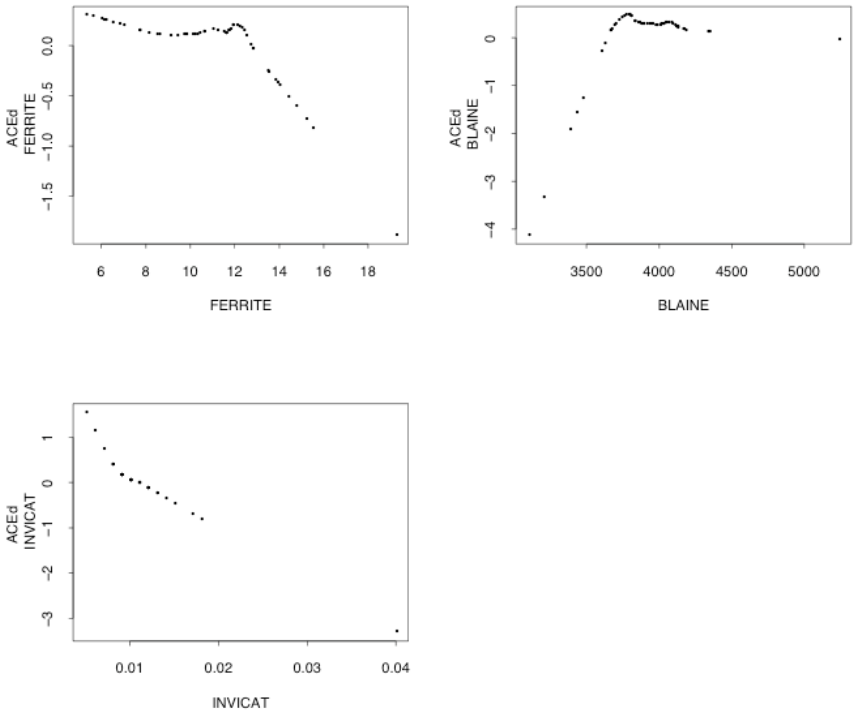

Figure 2-25- ACE transforms for ferrite, Blaine, and 1/Vicat yields a combination of a smooth transform and high R^2 of 0.88. This is the combination chosen to illustrate explicit parameterization.

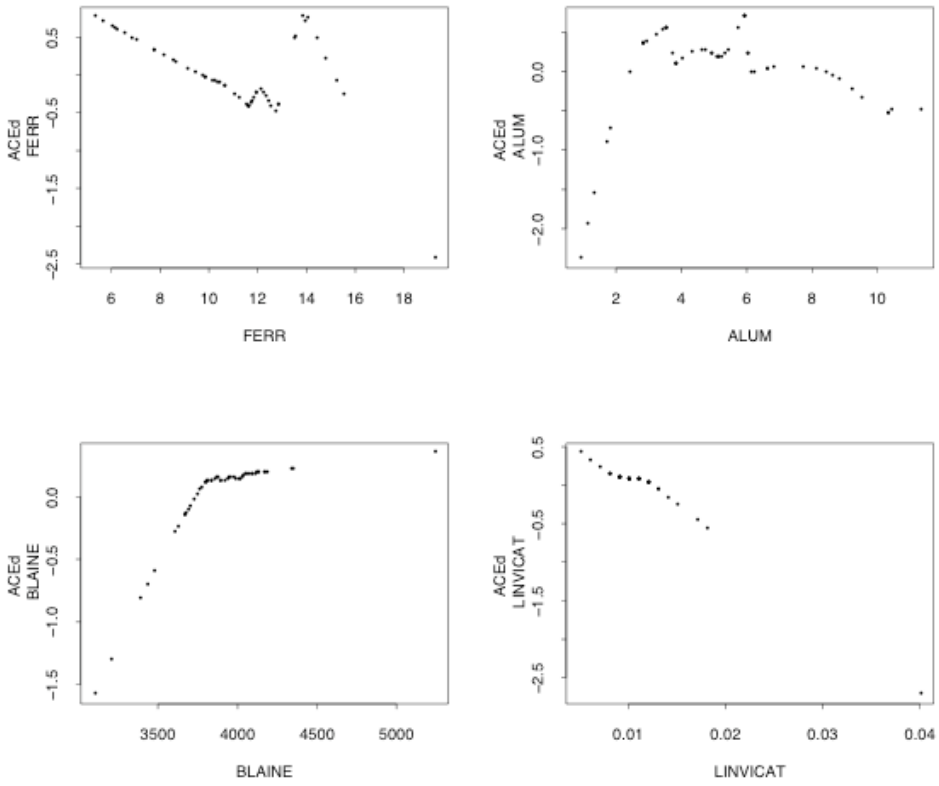

Figure 2-26- ACE transforms for ferrite, aluminate, Blaine, and 1/Vicat yields an R^2 of 0.90 but the curves for ferrite and aluminate appear rough.

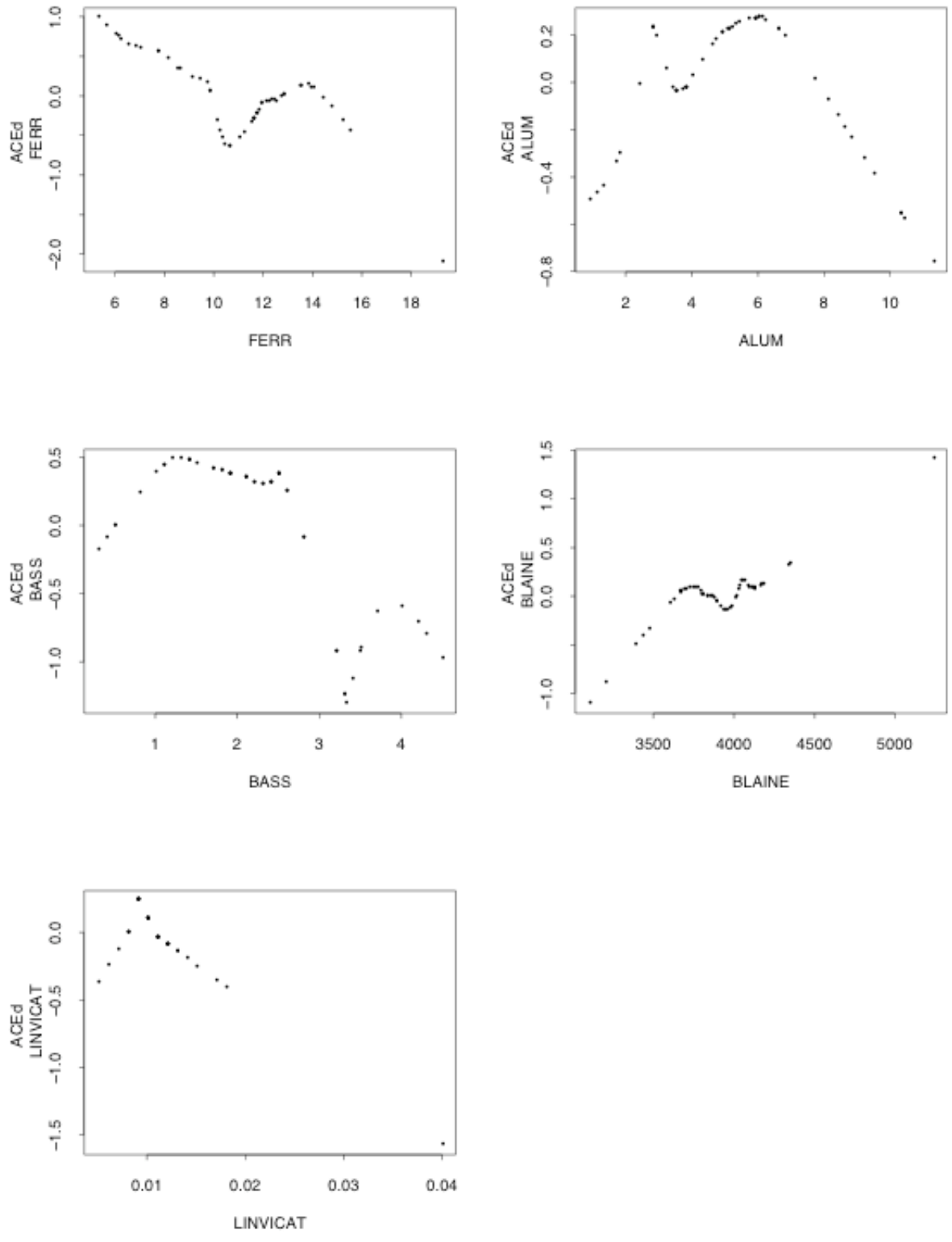

Figure 2-27- ACE transforms for aluminate, ferrite, bassanite, Blaine, and 1/Vicat yields an R^2 of 0.86.

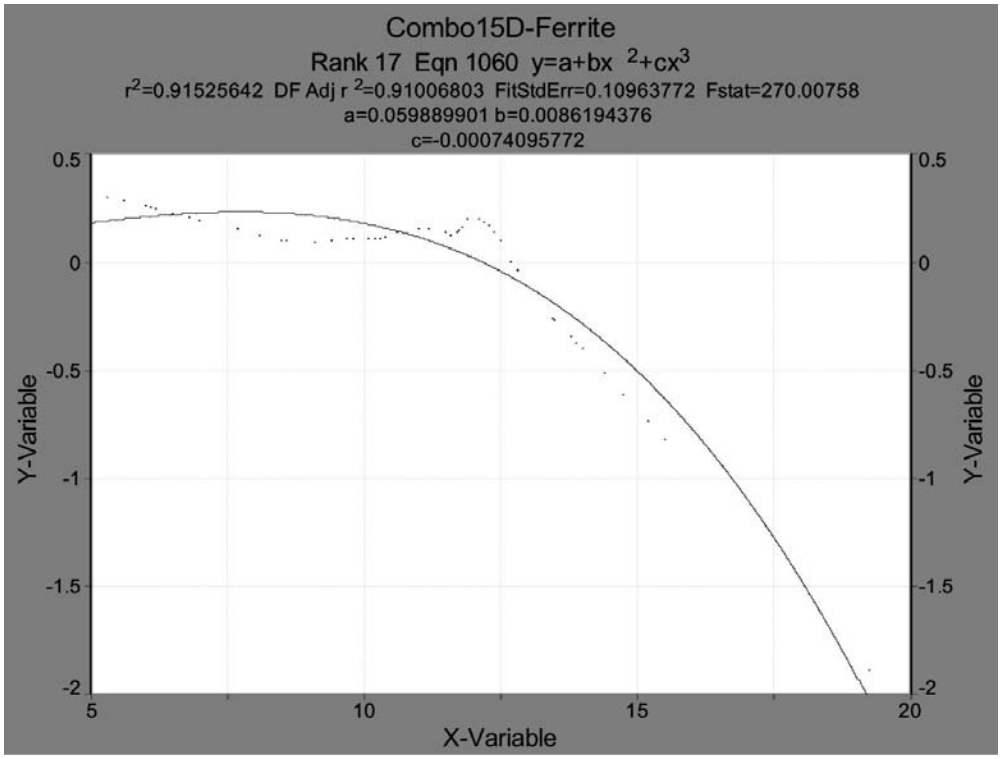

Figure 2-28- Ferrite ACE transform is approximately described by a simplified cubic function.

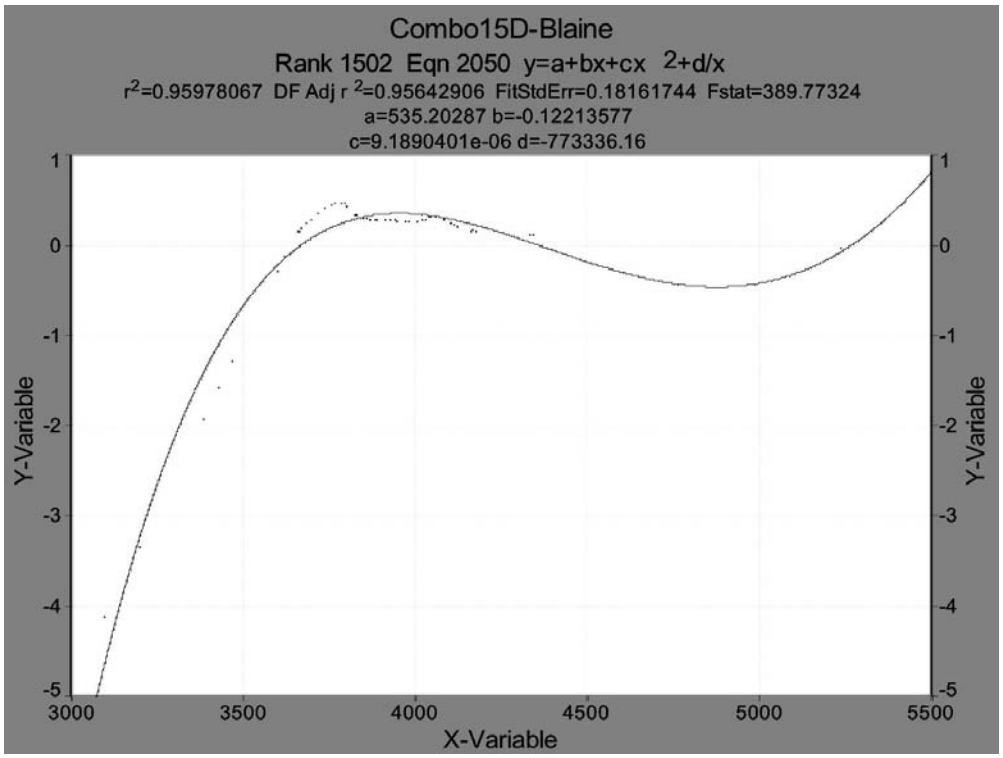

Figure 2-29- Transformed Blaine fineness can be described by a mixed x-(1/x) quadratic.

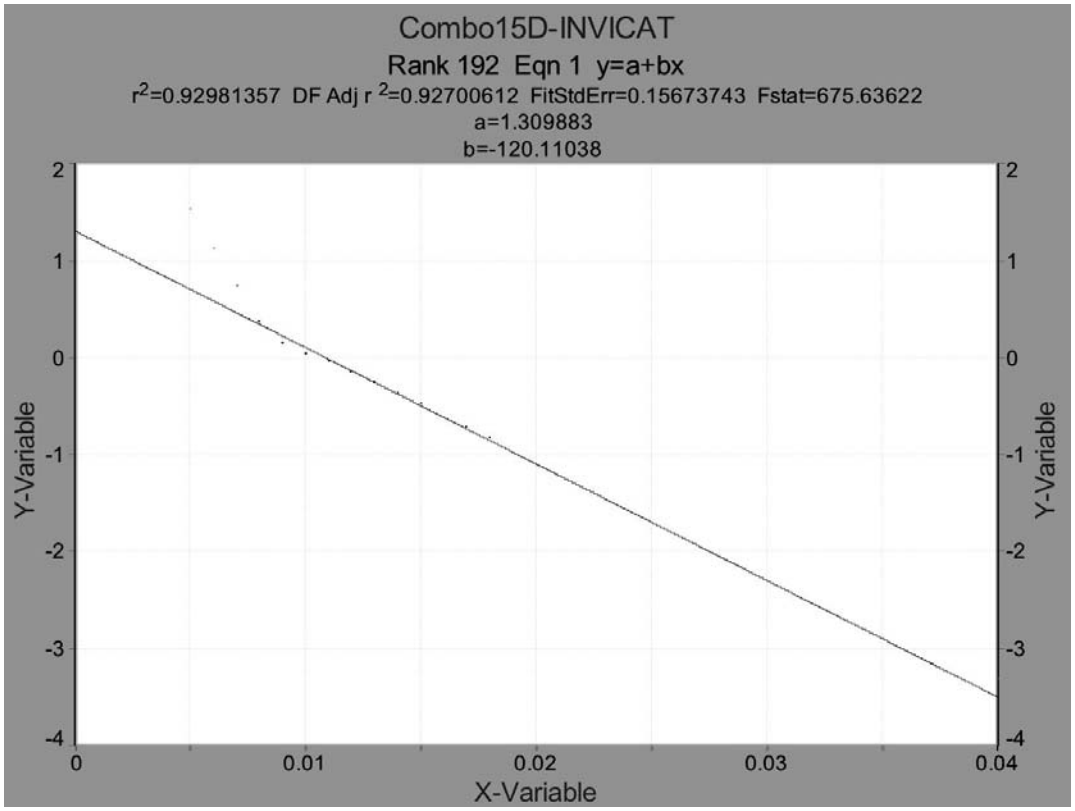

Figure 2-30- Transformed 1/Vicat results in an almost linear structure.

As illustrated in Figure 2-28 through Figure 2-30, a simplified cubic fit captures the main structure in the ferrite transform, a x-(1/x) quadratic captures Blaine structure, and a line captures the descent of the (1/Vicat) transform. It should be clear that approximate parametric fitting of each ACE transformation will contribute to a diminution of the overall model fit's R^2 (0.88), as each approximate parametric fit reduces that variable's contribution to R^2.

Substituting the best-fit parametrizations into the APSACE candidate model gives

$$ACE(HOH7) = ACE(F) + ACE(B) + ACE(V)$$
$$= (0.06 + 0.009F^2 - 0.0007F^3) + \left(535 - 0.12B + \left(9 \cdot 10^{-6}\right)B^2 - \frac{773336}{B}\right) + (-120V + 1.3)$$

where

F = ferrite

V = 1/Vicat

B = Blaine

If ACE(HOH7) were easily parametrically invertible (approximately a straight line), the application of the inverse of ACE(HOH7) to the right hand side of the equation would yield a parametric equation for HOH7 in terms of ferrite, Blaine, and inverse Vicat. In this example, however, the pictorial ACE transform of HOH7 (Figure 2-31) is sufficiently rough that a simply invertible parameterization is unavailable.

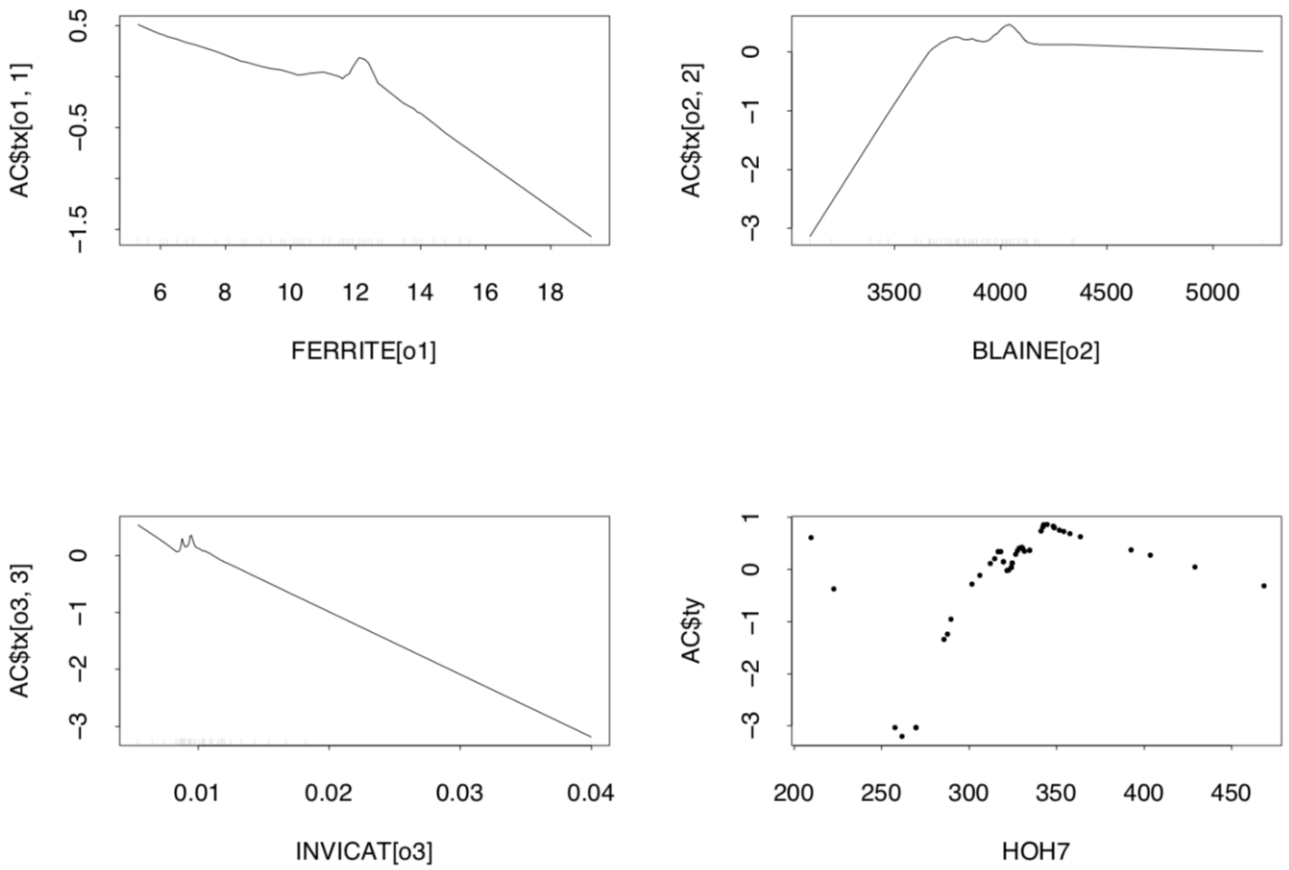

Figure 2-31- The roughness of the ACE(HOH7) transform (upper-right) precludes a simple invertible parametrization

CHAPTER 3- CONCLUSIONS AND FUTURE DIRECTIONS

3.1 Conclusions

This work has not resulted in a simple parametric model for HOH7. Instead, we have found simple conclusions concerning the variables and the data considered that may offer general guidance on the modeling of HOH7. It is to be strongly emphasized that the one explicitly parameterized model presented is meant to be illustrative of the power of the technology only.

Good fitting models for HOH7 often incorporate -

1. A structural mineralogical phase component (belite preferred);
2. A sulfate phase component (bassanite preferred);
3. A total fineness or particle surface area component (Blaine preferred);
4. Ferrite in conjunction with Fe_2O_3, and possibly TiO_2 or aluminate or C_3Ac, but not C_3Ao

Of the extra variables offered here for incorporation in models, good fitting models seem indifferent to the inclusion of -

1. 1/Vicat (generally)
2. compressive strength
3. calcite

The prevalence of noisy and multistructured ACE plots in this study can be attributed to multiple potential causes:

1. The variables are inappropriate: variables are included that have no true HOH modeling content, and variables which should be included - like precipitation products known to influence heat evolution - are not.

2. The inclusion of too many types of cements in the dataset is inappropriate. In the future, attention should be focused for this kind of modeling on one specific Type of cement. It should be easier to capture predictive power for a relatively homogeneous class of cement than for an aggregate of different Types of cements representing possibly different heat evolution mechanisms.

It is interesting in this regard that occasionally sharp transitions, or inflection\ points in the ACE plots could correspond to transition points in the variable being ACE transformed from a range of values characteristic of one Type of cement to a range of values characteristic of another Type.

There are certainly researchers who would argue that the phenomenon under study is much too complex to be captured by a simply parameterized model, regardless of how it is derived. Nonetheless, complex phenomena are often model-able in the large, by coarse-grained parameterizations. Until a commonly accepted, validated fine-grained model emerges, there is always the possibility that macro-modeling of key variables can provide some useful explanatory power for phenomena like HOH7.
In this work, which has not exhaustively explored the various transformation possibilities outlined above, we have found that

1. Certain physically meaningful transformations must be applied if the variable(s) are to have a chance of entering meaningfully into a model: e.g., 1/(VICATF - VICATI).

2. As a transformation inducing (a) change of magnitude to values commensurate with other variables entering into the model, and (b) a degree of variance homogeneity, the log transformation of HOH7 and Blaine in particular have been routinely compared with results achieved by use of the raw variables. Generally, with other variables entering the mix, approximately equivalent R^2 values are achieved and there is not always a noticeable improvement in the smoothness of the ACE transformations.

3. The log odds transformation has not proven particularly useful in cases where it has been tried, but bears further examination if newer, more appropriate, datasets are modeled.

4. Common transformations reported in the literature, but not examined here, should be tested in further work of this kind: Lime Saturation Factor, Alumina Ratio, Silica Ratio, Alkali-Sulfate Ratio, and variables derived from precipitation products.

The data we have used here for demonstration purposes may be inadequate for credible model development. They may cover too broad a range of cements, over too broad a range of HOH7s to be simply and explicitly model-able. The break in the Blaine transform curve that consistently occurs is at a value that roughly corresponds to the differences between Type III (mean Blaine of 556 M^2/kg) and Types I, II, and V (roughly 380 M^2/kg) suggests that these should be modeled separately. The data may be lacking significant, strong variables. Taylor [3] notes that hydration products also have an enthalpy of hydration, but they were not included here. The inclusion may result in higher predictive power for HOH7 than the variables examined so far.

3.2 Future Directions

In future work, we would investigate subsets of the data, concentrating on specific cement Types and variables like belite, bassanite, 1/Vicat, and Blaine have shown promise based on the data utilized in this study. This could be accomplished by:

1. Organizing many of our best candidate several-variable component models for general insight into which variables offer generally the best predictive power for HOH7.

2. Expanding old and new datasets to include wider ranges of variables, for example precipitation products or widely used "derived variables" (ratio's, fraction's, for example), and further exploring the utility of transformations.

3. Developing different classes of models for more narrowly defined classes of cements: Type I's, Type II's, Type III's, Type V's, and C 1157 and C 595 cements.

4. Cross-validating existing best candidate models, when new data appropriate for cross-validation becomes available.

REFERENCES

[1] Portland Cement, Concrete, and Heat of Hydration, Concrete Technology Today, Vol. 18, No. 2, July 1997.
[1] L. Wadsö, "Applications of an eight-channel isothermal conduction calorimeter for cement hydration studies," Cement International, No. 5, 2005, pp. 94-101.
[1] H.F.W. Taylor, Cement Chemistry, Thomas Telford, London, 1997
[1] L.E. Copeland, D.L. Kantro, and G. Verbeck, "Chemistry of Hydration of Portland Cement," Paper IV-3, pp. 429-463, Fourth International Symposium on the Chemistry of Cement, Washington, D.C., 1960
[1] R.H. Bogue, "Calculation of the Compounds in Portland Cement", PCA Fellowship Paper No. 21, October, 1929. Also – Industrial and Engineering Chemistry, Vol. 1, No. 4, P. 192, Oct 15, 1929.
[1] R.H. Bogue and Wm. Lerch, "Hydration of Portland Cement Compounds," Portland Cement Association Fellowship at the National Bureau of Standards Paper No. 27, 30 pp., August 1934.
[1] R.H. Bogue, The Chemistry of Portland Cement, Reinhold Publishing, New York, 1955.
[1] L.J. Parrot and D.C. Killoh, "Prediction of Cement Hydration," in British Ceramic Society Procedings 35, pp. 41-54, 1984.
[1] Wm. Lerch, "The Influence of Gypsum on the Hydration Properties of Portland Cement Pastes," Proceedings of the American Society for Testing Materials, Vol. 46, 1946.
[1] F.J. Tang and E. M. Gartner, "Influence of sulfate source on Portland Cement hydration," Advances in Cement Research, Vol. 1, No. 2, April, 1988.
[1] J. Gebauer and M. Kristmann, "The Influence of the composition of industrial clinker on cement and concrete properties," World Cement Technology, 10(2), March, 1979, 46-51.
[1] T.S. Poole, "Predicting Seven-Day Heat of Hydration of Hydraulic Cement from Standard Test Properties," Journal of ASTM International, Vol. 6, No. 6, 10 pp., June 2009.
[1] L. Breiman and J.H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," Journal of the American Statistical Association, Vol. 80, No. 391, Sept. 1985, pp. 580-598.
[1] D. Wang and M. Murphy, "Estimating Optimal Transformations for Multiple Regression Using the ACE Algorithm," Journal of Data Science 2(2004), 329-346.
[1] D.M. Roy, P. Tilkalsky, B. Sheetz, and J. Rosenberger, "Relationship of Portland Cement Characteristics to Concrete Durability," *in*, Research Results Digest, NCHRP Project 18-05, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC, 2002
[1] P. Stutzman and S. Leigh, "Phase Analysis of Hydraulic Cements by X-Ray Powder Diffraction: Precision, Bias, and Qualification," Journal of ASTM International, Vol. 4, No. 5, JAI Paper ID JAI101085, 11pp., 2007.
[1] ASTM C 1365, "Standard Test Method for Determination of the Proportion of Phases in Portland Cement and Portland-Cement Clinker Using X-ray Powder Diffraction Analysis" Annual Book of ASTM Standards, Vol. 4.01, ASTM International, West Conshohocken, PA.

[1] R.A. Young, 'Introduction to the Rietveld Method', in IUCr Monographs on Crystallography, 5, The Rietveld Method, R.A. Young, ed., pp. 1-37,
[1] L.J. Struble, L.A. Graf, J.I. Bhatty, "X-Ray Diffraction Analysis", *in* Innovations in Portland Cement Manufacturing, Chapter 8.2, J.I. Bhatty, F. M. Miller, and S.H. Kosmatka, eds, PCA SP 400.01, 2004, 1367 pp.
[1] R.J. Hill and C.J. Howard, 'Quantitative Phase Analysis from Neutron Powder Diffraction Data Using the Rietveld Method,' J. Appl. Cryst., 20, 467-474.
[1] "Standard Test Methods for Fineness of Hydraulic Cement by Air-permeability Apparatus," ASTM C 204-07, Annual Book of ASTM Standards, Vol. 4.01
[1] "Standard Test Methods for Time of Setting of Hydraulic Cement by Vicat Needle" ASTM C 191, Annual Book of ASTM Standards, Vol. 4.01
[1] F. Mosteller & J.W. Tukey (1977). *Data analysis and regression : a second course in statistics*. Addison-Wesley.
[1] W.J. Dixon, ed., BMDP Statistical Software manual Volumes 1,2. University of California Press, Berkley, 1992.
[1] R.H. Meyers, Classical and Modern Regression with Applications, Duxbury Press, Boston, 1986.
[1] L.C. Hamilton, Regression with Graphics: A Second Course in Applied Statistics, Brooks/Cole/Wadsworth, Belmont, CA, 1992.
[1] J. Neter, M.H. Kutner, C.J. Nachtsheim, and W. Wasserman, Applied Linear Statistical Models, 4th Edn., WCB/McGraw-Hill, Boston, 1996.
[1] I.T. Jolliffee, Principal Component Analysis, Springer Verlag, New York, 1986.
[1] T.J. Hastie and R.J. Tibshirani, Generalized Additive Models, London: Chapman and Hall, 1990.
[1] M.J. Crawley, Statistical Computing: An Introduction to Data Analysis Using S-Plus, Chichester: John Wiley, 2002.
[1] W.N. Venables and B.D. Ripley, Modern Applied Statistics with S-Plus, 3rd edn. New York: Springer, 1999.
[1] J.I. Bhatty and P.D. Tennis, "U.E. and Canadian Cement Characteristics: 2004, PCA R&D SN2879, Portland Cement Association, 67 pp., 2008.

APPENDIX: CEMENTS DATA

Cement	Source	Alite	Belite	Ferrite	Aluminate	C3Ac	C3Ao	Periclase	Anhydrite	Bassanite	Gypsum	Calcite
140	1	64.2	16.8	6.8	5.2	4.0	1.2	0.7	2.1	1.9	0.6	0.4
141	1	58.9	12.4	7.7	8.4	8.4	0.0	1.8	0.6	1.8	3.7	0.6
142	1	51.8	22.3	11.7	5.1	5.1	0.0	3.8	0.3	2.4	0.3	0.3
143	1	59.6	23.5	12.4	2.8	1.7	1.0	0.8	0.4	0.4	1.3	0.4
144	1	51.4	23.0	6.5	9.5	8.3	1.1	0.0	0.2	1.7	0.8	2.5
145	1	55.7	19.2	5.3	11.3	10.4	0.9	1.4	0.5	2.2	1.0	2.1
146	1	56.3	21.9	9.8	2.4	0.4	2.1	2.9	0.3	3.2	0.5	0.3
147	1	63.8	12.5	12.3	2.8	2.1	0.7	2.3	1.1	1.9	1.7	2.3
148	1	56.9	15.8	12.8	5.7	1.4	4.3	0.1	0.0	4.0	0.4	0.5
149	1	56.3	19.9	9.8	6.2	4.1	2.1	1.4	0.0	0.3	4.5	0.3
150	1	58.4	19.2	11.6	4.9	3.7	1.1	1.1	0.0	1.2	2.4	0.2
151	1	63.3	13.4	11.9	5.1	2.5	2.6	0.5	0.3	3.7	0.8	0.8
152	1	64.0	12.7	8.1	8.8	5.4	3.4	1.2	0.1	2.5	0.9	0.6
153	1	59.3	20.2	11.6	2.9	2.4	0.4	0.2	0.0	0.8	2.2	2.7
154	1	65.3	9.9	9.1	3.8	0.1	3.8	1.2	0.3	2.6	0.5	6.5
155	1	64.2	17.9	9.7	4.7	4.3	0.4	0.1	0.1	2.3	0.6	1.2
156	1	51.7	24.1	6.1	10.3	10.3	0.0	0.2	0.6	1.4	0.6	3.5
157	1	52.9	21.1	11.0	3.8	0.6	3.2	1.8	0.0	1.0	2.3	3.2
158	1	60.4	13.0	11.9	4.0	0.3	3.6	0.8	0.2	2.1	1.7	3.7
159	1	49.3	21.8	6.0	10.3	9.0	1.3	0.8	0.3	1.3	3.2	5.4
160	1	58.7	11.1	15.5	5.9	4.1	1.7	0.2	0.4	0.5	1.9	4.0
161	1	57.7	18.4	12.8	3.5	2.9	0.6	0.8	0.3	1.8	1.7	0.8
162	1	55.1	21.0	14.4	1.8	1.8	0.0	2.4	0.2	2.5	1.5	1.8
163	1	62.0	13.5	10.1	6.8	2.6	4.2	1.1	0.3	2.2	2.0	0.3
164	1	48.8	24.2	15.2	2.8	2.8	0.0	0.0	4.5	3.2	0.6	0.5
165	1	46.0	28.7	11.2	3.5	2.6	1.0	1.6	0.5	2.1	1.3	3.0
166	1	51.5	13.7	9.4	6.6	6.0	0.6	0.9	2.8	2.6	0.9	4.3
167	1	52.5	18.6	10.2	7.7	7.7	0.0	2.4	0.5	3.3	0.3	0.5
168	1	53.8	18.6	7.7	9.2	9.2	0.0	3.7	0.2	2.3	0.1	0.9
169	1	45.9	28.7	10.6	3.8	1.5	2.3	0.9	1.5	1.5	1.2	2.1
170	1	51.9	18.3	8.6	6.0	0.9	5.1	0.3	0.1	4.5	0.8	4.2

Cement	Source	Alite	Belite	Ferrite	Aluminate	C3Ac	C3Ao	Periclase	Anhydrite	Bassanite	Gypsum	Calcite
1007	2	62.3	14.3	5.6	10.4	10.4	0.0	2.0	0.1	3.5	0.0	0.1
1015	2	60.5	12.4	8.5	5.9	1.1	4.7	0.6	1.3	1.1	1.5	0.3
1016	2	53.7	16.3	10.4	8.6	7.8	0.8	0.7	0.1	3.4	1.1	0.5
1017	2	57.8	19.2	12.1	4.3	3.4	0.9	0.7	0.2	2.5	1.5	0.3
1020	2	60.3	15.3	10.6	4.9	3.4	1.5	0.1	0.4	2.8	0.5	0.2
1024	2	57.2	12.9	12.2	5.4	3.8	1.5	0.1	0.1	1.1	2.9	6.6
1027	2	59.5	17.3	13.5	3.4	2.8	0.6	0.7	0.6	1.9	0.9	1.3
1028	2	60.0	14.3	13.9	3.5	3.3	0.2	0.6	2.5	1.4	1.0	1.1
1039	2	55.8	18.1	11.5	5.3	2.6	2.7	0.1	0.1	4.3	0.0	0.0
1043	2	67.8	7.0	7.0	8.1	6.3	1.7	2.2	1.0	1.4	1.4	0.3
1051	2	53.7	20.6	12.5	3.2	1.2	2.0	3.1	0.1	3.3	0.5	0.1
1059	2	63.4	11.0	6.2	6.6	6.6	0.0	0.9	1.1	0.5	4.2	0.0
1060	2	55.5	18.3	11.6	5.9	3.2	2.7	0.2	0.1	1.7	2.5	0.3
1104	2	61.8	13.7	11.8	4.6	2.5	2.1	0.1	0.5	2.1	1.5	0.4
2312	2	64.0	13.3	14.0	1.3	0.3	1.0	0.2	0.3	2.8	0.1	0.2
2322	2	60.9	12.6	13.8	1.7	1.7	0.0	0.4	1.1	2.5	0.1	0.2
2334	2	56.2	13.6	10.3	6.0	0.8	5.2	2.4	0.4	2.4	1.9	0.5
2403	2	56.1	17.6	12.7	3.7	3.7	0.0	0.2	0.0	4.2	1.6	0.2
ERDC 18-01	3	29.82	32.47	19.25	5.05	2.86	2.2	0.4	0	1.9	0.23	8.8
ERDC 26-05	3	56.5	12.54	13.47	0.89	0	0.89	2.57	4.07	3.49	0.12	3.9
ERDC 70070	3	39.37	24.45	14.75	1.11	0	1.11	2.65	4.37	2.4	2.79	0.53
ERDC 24-08	3	26.57	46.05	11.73	6.11	0	6.11	4.8	0	3.32	0.9	0

Cement	D10	D50	D90	Span	Blaine	LOI	VicatI	VicatF	Str3d	SiO_2	Al_2O_3	Fe_2O_3
140	1.98	12.39	35.16	2.68	4023	0.84	136	256	3253	21.490	5.082	2.729
141	2.40	12.61	37.13	2.75	3971	1.88	138	253	4063	18.940	5.680	2.458
142	2.50	10.46	31.15	2.74	4166	0.20	128	236	3904	20.233	4.669	2.739
143	1.56	10.21	30.84	2.87	3982	1.51	120	217	3871	21.143	4.500	3.531
144	1.63	11.26	39.54	3.37	4123	2.15	90	173	3611	20.327	5.324	2.354
145	1.76	12.52	39.66	3.03	3669	1.21	126	229	3846	19.758	5.658	2.126
146	1.51	9.57	30.73	3.05	4335	1.07	136	248	3724	22.186	3.363	3.121
147	1.53	10.97	32.76	2.85	4114	1.61	98	204	4068	19.638	4.327	3.264
148	1.66	12.10	36.11	2.85	3756	1.64	140	245	3775	20.231	4.788	3.162
149	2.33	12.90	40.22	2.94	3828	1.98	120	222	3539	20.180	4.990	3.030
150	1.66	12.10	36.11	2.49	3862	1.67	135	231	3485	20.720	4.660	3.180
151	1.78	11.95	37.20	2.96	3788	1.42	98	190	3982	20.030	4.910	3.460
152	1.64	11.11	35.27	3.03	4086	1.35	133	232	3955	19.960	5.680	2.400
153	2.45	13.12	37.40	2.66	3913	2.01	157	269	3414	20.870	5.030	3.440
154	2.25	13.52	42.84	3.00	3887	1.81	138	241	3243	22.120	3.470	2.900
155	1.46	11.57	33.28	2.75	4042	1.56	89	184	3867	20.450	5.000	3.343
156	1.40	11.09	39.58	3.44	3941	2.48	118	209	3807	20.170	5.720	2.213
157	1.93	11.13	31.98	2.70	3888	2.91	152	272	3253	21.110	3.730	3.360
158	1.69	10.73	30.51	2.69	4006	2.71	129	243	3954	20.830	4.000	3.000
159	1.55	12.79	40.72	3.06	3662	2.65	124	236	3783	20.030	5.100	1.990
160	1.58	12.75	36.73	2.76	3795	2.48	115	226	3539	20.510	5.120	3.620
161	2.17	12.95	38.09	2.77	3686	1.67	121	229	3653	20.360	5.180	3.660
162	2.61	14.26	40.14	2.63	3740	2.00	139	253	3554	20.340	4.750	3.530
163	1.81	13.94	40.88	2.80	3698	1.44	165	272	3624	20.590	4.930	2.750
164	2.10	13.18	42.19	3.04	3720	1.09	129	235	3598	20.200	5.130	4.240
165	1.98	13.20	39.45	2.84	3803	2.14	123	232	3557	20.630	4.490	2.900
166	1.87	12.04	36.67	2.89	4061	2.40	118	223	4252	19.040	5.260	2.380
167	1.06	8.17	28.29	3.33	5237	0.84	142	257	5109	19.290	5.920	2.440
168	1.26	10.88	36.21	3.21	4083	0.86	128	241	4189	19.910	5.110	2.140
169	1.32	9.77	26.61	2.59	4161	2.30	148	264	2976	22.070	3.170	3.700
170	1.31	11.93	42.96	3.49	4124	2.78	89	185	3064	21.710	3.770	2.480

Cement	D10	D50	D90	Span	Blaine	LOI	VicatI	VicatF	Str3d	SiO_2	Al_2O_3	Fe_2O_3
1007	2.08	12.86	35.37	2.59	3663	0.73	115	190	4801	19.720	5.270	1.970
1015	2.05	11.26	34.85	2.91	4345	1.44	60	145	4152	21.120	4.420	2.480
1016	2.30	13.49	40.24	2.81	3831	1.02	115	205	3303	20.450	5.610	2.900
1017	2.57	14.20	38.71	2.55	3384	0.84	143	298	3900	21.250	4.290	2.440
1020	1.87	12.97	41.25	3.04	3803	0.70	90	145	4589	20.290	5.540	2.900
1024	3.54	14.42	37.59	2.36	4009	2.11	110	170	3860	20.080	5.470	2.900
1027	2.42	12.35	37.02	2.80	4179	1.74	85	170	4549	20.200	4.580	2.590
1028	2.62	13.46	39.06	2.71	4029	1.71	90	185	3760	19.610	5.250	2.600
1039	2.27	12.92	38.30	2.79	3871	1.41	100	165	3528	20.490	5.140	2.540
1043	2.03	13.21	43.05	3.11	3871	1.23	135	205	3422	20.560	5.020	1.940
1051	2.07	13.07	37.29	2.69	3950	1.12	150	235	3674	20.540	4.560	3.180
1059	1.87	12.66	37.82	2.84	3850	1.46	92	175	3767	20.770	4.550	1.520
1060	2.27	13.50	37.13	2.58	3600	1.46	84	170	4404	21.080	4.700	3.120
1104	2.22	12.64	36.83	2.74	3768	1.52	83	170	4921	21.330	4.040	2.880
2312	2.13	12.63	35.95	2.68	3622	0.59	200	285	3289	21.450	4.010	4.020
2322	2.02	11.99	35.03	2.75	3930	0.70	90	165	4218	20.680	4.840	3.230
2334	1.89	11.39	34.02	2.82	4104	1.36	145	225	3740	20.400	4.630	2.910
2403	2.17	12.64	35.33	2.62	4048	1.52	125	150	4271	20.260	4.450	3.200
ERDC 18-01	2.856	13.106	36.935	2.6	3470	1.2	186	304	1830	21.800	5.300	4.500
ERDC 26-05	2.48	14.111	46.964	3.15	3200	1.05	247	382	1545	21.600	2.800	4.200
ERDC 70070	2.188	15.558	51.896	3.2	3100	0.88	398	582	2030	21.100	5.000	3.000
ERDC 24-08	1.735	12.308	36.938	2.86	3430	1.7	337	448	1590	23.000	3.200	3.500

Cement	CaO	MgO	SO_3	Na_2O	K_2O	7dHOH
140	64.559	1.355	2.837	0.093	0.186	344
141	61.488	2.612	4.728	0.405	1.042	351
142	61.555	4.330	3.500	0.317	1.086	330
143	63.730	1.680	2.735	0.121	0.719	317
144	64.513	0.966	3.037	0.111	0.757	317
145	62.670	2.504	3.651	0.349	1.165	341
146	63.160	3.576	2.417	0.281	0.394	324
147	63.850	3.500	2.610	0.132	0.692	347
148	64.270	1.360	2.901	0.132	0.823	340
149	62.240	2.450	3.480	0.320	0.826	327
150	63.990	2.210	2.370	0.068	0.695	330
151	64.730	1.340	3.120	0.269	0.320	330
152	64.460	2.060	2.660	0.191	0.500	363
153	63.730	1.180	2.770	0.110	0.420	329
154	63.550	2.380	2.740	0.110	0.690	323
155	64.910	1.000	2.850	0.090	0.360	342
156	63.600	1.190	3.270	0.140	0.870	326
157	63.490	2.540	2.230	0.213	0.470	316
158	64.930	1.660	2.350	0.114	0.310	342
159	64.360	1.270	3.680	0.120	0.510	353
160	62.930	0.920	3.160	0.065	0.750	324
161	63.890	1.170	2.730	0.094	0.630	331
162	61.890	3.550	3.000	0.127	0.530	334
163	63.940	1.870	2.880	0.196	0.682	328
164	63.640	1.060	3.580	0.058	0.383	323
165	62.650	2.590	3.230	0.160	0.726	319
166	63.490	2.080	3.680	0.156	1.175	348
167	61.640	3.110	4.390	0.355	1.068	350
168	62.280	3.880	3.480	0.226	1.227	339
169	63.470	2.000	2.150	0.201	0.429	305
170	64.320	1.440	2.750	0.092	0.628	316

Cement	CaO	MgO	SO_3	Na_2O	K_2O	7dHOH
1007	63.140	3.270	3.430	0.320	1.150	301
1015	63.890	1.990	2.810	0.470	0.670	403
1016	63.940	1.880	3.000	0.179	0.117	209
1017	63.370	2.040	3.190	0.390	0.710	321
1020	64.970	1.040	2.730	0.112	0.830	314
1024	63.770	1.150	3.130	0.121	0.790	334
1027	62.650	2.090	3.260	0.200	0.930	468
1028	62.430	2.110	4.350	0.200	1.090	428
1039	63.100	1.740	3.840	0.180	1.000	287
1043	63.770	3.770	2.730	0.090	0.560	392
1051	61.010	5.010	3.480	0.250	0.820	222
1059	63.710	2.480	3.050	0.056	0.017	357
1060	63.960	1.220	2.680	0.096	0.610	348
1104	63.850	1.250	2.680	0.133	0.360	311
2312	64.650	1.390	2.380	0.199	0.490	289
2322	64.650	1.850	2.900	0.038	0.540	322
2334	61.760	3.960	3.180	0.320	0.570	322
2403	63.810	1.550	4.130	0.156	0.460	257
ERDC 18-01	61.600	1.400	2.400	0.030	0.600	269
ERDC 26-05	63.100	3.600	2.800	0.120	0.230	261
ERDC 70070	62.900	3.200	2.800	0.240	0.090	269
ERDC 24-08	61.500	3.700	2.200	0.490	0.300	285