DISTRIBUTION SHEET

To	From	Page 1 of 5
Distribution	FR Reich	Date 11/15/94
Project Title/Work Order		EDT No. 609082
CPAC Optical Moisture Monitoring: Characterization of Composition and Physical Effects on Moisture Determination Task 2A Report WHC-SD-WM-ER-397		ECN No.

Name	MSIN	Text With All Attach.	Text Only	Attach./ Appendix Only	EDT/ECN Only
U.S. Department of Energy, Richland Field Office					
DA Brown	K8-50	X			
RF Christensen	S7-54	X			
JM Clark	S7-54	X			
RE Gerton	S7-54	X			
RG Harwood	S7-54	X			
WF Hendrickson	S7-54	X			
DE Trader	K8-50	X			
Westinghouse Hanford Company					
Central Files (2)	L8-04	X			
Document Processing and Distribution	L8-15	X			
EDMC	H6-08	X			
OSTI(2)	L8-07	X			
Tank Farm Information Center	R1-20	X			
H Babad	S7-30	X			
DC Board	S1-57	X			
GL Borsheim	H5-27	X			
RJ Cash	S7-15	X			
C DeFigh-Price	X3-71	X			
DA Dodd	T6-50	X			
JG Douglas	L5-55	X			
GT Dukelow	S7-15	X			
SJ Eberlein	S7-31	X			
KD Fowler	R2-11	X			
JM Grigsby	H4-62	X			
CS Homi	R2-12	X			
MN Islam	R3-08	X			
DW Jeppson	L5-31	X			
M Kummerer	H4-62	X			
LL Lockrem	S3-90	X			
T Lopez	L5-55	X			
JE Meacham	S7-15	X			
SJ Mech	L5-55	X			
MA Payne	S7-14	X			
RS Popielarczyk	R1-30	X			
AK Postma	H4-62	X			
TV Rebagay	T6-30	X			

FR Reich (12)	L5-55	X
CP Schroeder	L7-06	X
JP Sederburg	R2-11	X
BC Simpson	R2-12	X
H Toffer	H0-38	X
DA Turner	S7-15	X
WT Watson	H0-38	X
WD Winkelman	L5-55	X

Mactech

V FitzPatrick	K8-50	X

Pacific Northwest Laboratory

JS Hartman	K5-25	X
RL Hockey	K5-25	X
GF Schiefelbein	P8-38	X
TM Sloane	K5-25	X
MJ Quadrel	K9-69	X

OFFSITE

1 Department of Energy,
Morgantown Energy Technology Center
PO Box 880, 3610 Collins Ferry Road
Morgantown, WV 26507-0880

Curtis Nakaishi

1 Department of Energy -Savannah
River Operations Office
PO Box A
Aiken, SC 29808

Thomas Temple

1 Department of Energy
Science Applications International Corporation
12850 Middlebrook Road
Trevion I, Suite 300
Germantown, MD 20874

Ray Daniels

2 Department of Energy
Science Applications International Corporation
2030 Century Blvd, Suite 200B
Germantown, MD 20874

Herb Sutter
P Szerszen

6 Tank Advisory Panel Members

Charles S. Abrams
1987 Virginia
Idaho Falls, ID 83404

David O. Campbell
102 Windham Road
Oak Ridge, TN 37830

Fred N. Carlson
6965 North 5th West
Idaho Falls, ID 83401

Donald T. Oakley
409 12th Street SW, Suite 310
Washington, DC 20024-2188

William R. Prindle
1556 Crestline Drive
Santa Barbara, CA 93105

Alfred Schneider
5005 Hidden Branches Drive
Dunwoody, GA 30338

1 Brookhaven National Laboratory
Upton, NY 11973

Kamal Bandyopadhyay

1 Fauske and Associates, Inc.
16W070 W. 83rd Street
Burr Ridge, IL 60521

Hans Fauske

1 Florida State University
Department of Chemistry B-164
Tallahassee, FL 32306

Thomas Vickers

1 Harvard University
295 Upland Avenue
Newton Highlands, MA 02161

Melvin First

2 Lawrence Livermore National Laboratory
PO Box 808, L-221
Livermore, CA 94550

Billy Hudson
F Milanovich

5 Los Alamos National Laboratory
PO Box 1663
Los Alamos, NM 87545

Steve Agnew
RJ Donohoe — INC-14, MS C345
Steve W. Eisenhawer
Thomas Larson — MS-P915, DX-DO
L. Harold Sullivan

3 Oak Ridge National Laboratory
PO Box 2008
Oak Ridge, TN 37831

Emory Collins — 7930, MS-6385
Charles Forsberg — MS-6495
Thomas Kress — 9108, MS-8088

1 Rice University
5211 Paisley
Houston, TX 77096

Andrew Veletsos

2 Sandia National Laboratory
PO Box 5800
Albuquerque, NM 87815

Dana Powers — 6404, MS-0744
Scott Slezak — 6415, MS-0741

1 University of Idaho
Chemistry Department
Moscow, ID 83844-2343

P. Griffiths

1 University of South Carolina
Department of Chemistry and Biochemistry
Columbia, SC 29208

SM Angel

1 University of South Carolina
Department of Electrical and Computer Engineering
Swearingen Engineering Center
Columbia, SC 29208

Joseph Byrd

1 University of Washington
Department of Chemistry, BG-10
131 Chemistry Library Bldg.
Seattle, WA 98195

BR Kowalski

1 Vanderbilt University
PO Box 1596, Station B
Nashville, TN 37235

Frank Parker

2 Westinghouse Savannah River Laboratory
PO Box 616
Aiken, SC 29802

Ronald Livingston
Patrick O'Rourke

DEC 22 1994 38

ENGINEERING DATA TRANSMITTAL

Page 1 of 1

1. EDT 609082

2. To: (Receiving Organization) Distribution	3. From: (Originating Organization)	4. Related EDT No.: N/A
5. Proj./Prog./Dept./Div.: WM-820	6. Cog. Engr.: F. R. Reich	7. Purchase Order No.: N/A
8. Originator Remarks: CPAC Optical Moisture Monitoring: Characterization of Composition and Physical Effects on Moisture Determination Task 2A Report		9. Equip./Component No.: N/A
		10. System/Bldg./Facility: N/A
11. Receiver Remarks:		12. Major Assm. Dwg. No.: N/A
		13. Permit/Permit Application No.: N/A
		14. Required Response Date: 12/21/94

15. DATA TRANSMITTED

(A) Item No.	(B) Document/Drawing No.	(C) Sheet No.	(D) Rev. No.	(E) Title or Description of Data Transmitted	(F) Approval Designator	(G) Reason for Transmittal	(H) Originator Disposition	(I) Receiver Disposition
1	WHC-SD-WM-ER-397		0	CPAC Optical Moisture Monitoring: Characterization of Composition & Physical Effects on Moisture Determination Task 2A Report	~~NA~~ 4	2	1	

16. KEY

Approval Designator (F)	Reason for Transmittal (G)	Disposition (H) & (I)
E, S, Q, D or N/A (see WHC-CM-3-5, Sec.12.7)	1. Approval 4. Review 2. Release 5. Post-Review 3. Information 6. Dist. (Receipt Acknow. Required)	1. Approved 4. Reviewed no/comment 2. Approved w/comment 5. Reviewed w/comment 3. Disapproved w/comment 6. Receipt acknowledged

17. SIGNATURE/DISTRIBUTION (See Approval Designator for required signatures)

(G) Reason	(H) Disp.	(J) Name	(K) Signature	(L) Date	(M) MSIN	(J) Name	(K) Signature	(L) Date	(M) MSIN	(G) Reason	(H) Disp.
2	1	Cog.Eng. F. R. Reich	F R Reich		L5-55						
2	1	Cog. Mgr. W. D. Winkelman	W Winkelman		L5-55						
		QA									
		Safety									
		Env.									

18. F. R. Reich 11/16/94
F. R. Reich — Signature of EDT Originator — Date

19. _____ — Authorized Representative for Receiving Organization — Date

20. W. D. Winkelman 11-17-94
W. D. Winkelman — Cognizant Manager — Date

21. DOE APPROVAL (if required) Ctrl. No.
[ ] Approved
[ ] Approved w/comments
[ ] Disapproved w/comments

BD-7400-172-2 (04/94) GEF097

BD-7400-172-1

INSTRUCTIONS FOR COMPLETION OF THE ENGINEERING DATA TRANSMITTAL

(USE BLACK INK OR TYPE)

BLOCK	TITLE	
(1)*	EDT	● Pre-assigned EDT number.
(2)	To: (Receiving Organization)	● Enter the individual's name, title of the organization, or entity (e.g., Distribution) that the EDT is being transmitted to.
(3)	From: (Originating Organization)	● Enter the title of the organization originating and transmitting the EDT.
(4)	Related EDT No.	● Enter EDT numbers which relate to the data being transmitted.
(5)*	Proj./Prog./Dept./Div.	● Enter the Project/Program/Department/Division title or Project/Program acronym or Project Number, Work Order Number or Organization Code.
(6)*	Cognizant Engineer	● Enter the name of the individual identified as being responsible for coordinating disposition of the EDT.
(7)	Purchase Order No.	● Enter related Purchase Order (P.O.) Number, if available.
(8)*	Originator Remarks	● Enter special or additional comments concerning transmittal, or "Key" retrieval words may be entered.
(9)	Equipment/Component No.	● Enter equipment/component number of affected item, if appropriate.
(10)	System/Bldg./Facility	● Enter applicable system, building or facility number, if appropriate.
(11)	Receiver Remarks	● Enter special or additional comments concerning transmittal.
(12)	Major Assm. Dwg. No.	● Enter applicable drawing number of major assembly, if appropriate.
(13)	Permit/Permit Application No.	● Enter applicable permit or permit application number, if appropriate.
(14)	Required Response Date	● Enter the date a response is required from individuals identified in Block 17 (Signature/Distribution).
(15)*	Data Transmitted	
	(A)* Item Number	● Enter sequential number, beginning with 1, of the information listed on EDT.
	(B)* Document/Drawing No.	● Enter the unique identification number assigned to the document or drawing being transmitted.
	(C)* Sheet No.	● Enter the sheet number of the information being transmitted. If no sheet number, leave blank.
	(D)* Rev. No.	● Enter the revision number of the information being transmitted. If no revision number, leave blank.
	(E) Title or Description of Data Transmitted	● Enter the title of the document or drawing or a brief description of the subject if no title is identified.
	(F)* Approval Designator	● Enter the appropriate Approval Designator (Block 15). Also, indicate the appropriate approvals for each item listed, i.e., SQ, ESQ, etc.
	(G) Reason for Transmittal	● Enter the appropriate code to identify the purpose of the data transmittal (see Block 16).
	(H) Originator Disposition	● Enter the appropriate disposition code (see Block 16).
	(I) Receiver Disposition	● Enter the appropriate disposition code (see Block 16).
(16)	Key	● Number codes used in completion of Blocks 15 (G), (H), and (I), and 17 (G), (H) (Signature/Distribution).
(17)	Signature/Distribution	
	(G) Reason	● Enter the code of the reason for transmittal (Block 16).
	(H) Disposition	● Enter the code for the disposition (Block 16).
	(J) Name	● Enter the signature of the individual completing the Disposition 17 (H) and the Transmittal.
	(K)* Signature	● Obtain appropriate signature(s).
	(L)* Date	● Enter date signature is obtained.
	(M)* MSIN	● Enter MSIN. Note: If Distribution Sheet is used, show entire distribution (including that indicated on Page 1 of the EDT) on the Distribution Sheet.
(18)	Signature of EDT Originator	● Enter the signature and date of the individual originating the EDT (entered prior to transmittal to Receiving Organization). If the EDT originator is the cognizant engineer, sign both Blocks 17 and 18.
(19)	Authorized Representative for Receiving Organization	● Enter the signature and date of the individual identified by the Receiving Organization as authorized to approve disposition of the EDT and acceptance of the data transmitted, as applicable.
(20)*	Cognizant Manager	● Enter the signature and date of the cognizant manager. (This signature is authorization for release.)
(21)*	DOE Approval	● Enter DOE approval (if required) by signature or control number that tracks the approval to a signature, and indicate DOE action.

*Asterisk denote the required minimum items check by Configuration Documentation prior to release; these are the minimum release requirements.

SUPPORTING DOCUMENT

1. Total Pages 153

2. Title	3. Number	4. Rev No.
CPAC Optical Moisture Monitoring: Characterization of Composition and Physical Effects on Moisture Determination Task 2A Report	WHC-SD-WM-ER-397	0

5. Key Words	6. Author
Near Infrared (NIR), Moisture sensing, NIR moisture sensing, tank simulant, particle size interferences, composition interferences.	Name: Dr. David J. Veltkamp Signature Organization/Charge Code 8A820/H2CAA

7. Abstract

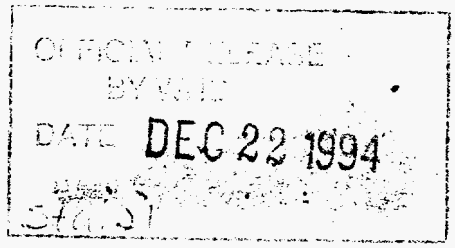
The impacts of particle size and chemical composition variations on the determination of tank simulant moisture from near infrared (NIR) optical spectra are presented. This work shows that particle size and chemical variations will impact moisture predictions from NIR spectra. However, the prediction errors can be minimized if calibration models are built with samples containing these variations as interferents. Building calibration models with interferents produces a more robust prediction performance.

Prior work showed that the NIR spectral region (1100 to 2500 nm) could be used to predict moisture content of BY-104 tank simulant with a standard error less of approximately 0.5 wt%. Particle size will increase moisture prediction error if calibration models do not include the same particle size ranges as unknown samples. A combined particle size model with 0-420 $x10^{-6}$m, 420-841 $x10^{-6}$m, and 841 $x10^{-6}$m-2 mm diameter particles predicted 0.59, 0.34 nd 0.23 wt% errors respectively for samples containing only these size ranges and 0.80 wt% error for a samples with all particle size ranges.

Chemical composition would also increase moisture prediction error if calibration model samples chemically differ from unknown samples. For a BY-104 simulant, increases in NaOH, $NaAlO_2$, Na_2SiO_3, and Na_3PO_4 produced moisture predictions that were lower than the actual moisture levels while increases in $Fe(NO_3)_3$, $Ca(NO_3)_2$, and $Mg(NO_3)_2$ resulted in a higher than actual moisture prediction. Root causes of these changes were not investigated, but systematic changes in the NIR spectra could be observed for these families of materials. When all of the composition variations were included in a single model, the model had a moisture prediction error of 1.41 wt% as compared to a 2.96 wt% error when changes were not in the calibration model.

This work demonstrates that a calibration model based on a single set of tightly controlled experimental conditions will tend to have a somewhat larger prediction errors when applied to samples collected with variations outside of these conditions.

8. RELEASE STAMP



RELEASE AUTHORIZATION

Document Number:	WHC-SD-WM-ER-397, Revision 0
Document Title:	CPAC Optical Moisture Monitoring: Characterization of Composition and Physical Effects on Moisture Determination Task 2A Report
Release Date:	November 29, 1994

This document was reviewed following the procedures described in WHC-CM-3-4 and is:

APPROVED FOR PUBLIC RELEASE

WHC Information Release Administration Specialist:

M. Boston — 11/29/94

M.N. Boston

WHC-SD-WM-ER-397
Revision 0

CPAC Optical Moisture Monitoring: Characterization of Composition and Physical Effects on Moisture Determination Task 2A Report

Prepared for the U.S. Department of Energy
Office of Environmental Restoration and
Waste Management

Westinghouse
Hanford Company Richland, Washington

Hanford Operations and Engineering Contractor for the
U.S. Department of Energy under Contract DE-AC06-87RL10930

Task 2A Report

Characterization of Composition and Physical Effects on Moisture Determination

CPAC Optical Moisture Monitoring of Hanford Waste Tanks Project

Westinghouse Hanford Company
Contract MRL-SVV-562951

Prepared by
Dr. David J. Veltkamp
(206) 543-6364

Center for Process Analytical Chemistry (CPAC)
University of Washington, Seattle, WA 98195

Table of Contents

List of Tables

List of Figures

1. Executive Summary

This report represents fulfillment of Task 2A of the CPAC Optical Moisture Monitoring of Hanford Waste Tanks project (Westinghouse Hanford Company Contract MRL-SVV-562951). Task 2A was to extend the initial feasibility study, completed under Task 1, to more fully characterize the effects of particle size and chemical composition on the optical determination of moisture. The initial Phase 1 studies were based on using the BY-104 salt cake surrogate over a water concentration range from zero to approximately twenty weight percent (wt%) water. Multivariate calibration methods were used to relate the changes in spectral response to the moisture concentration in the samples. Results from the previous study indicated that moisture could be determined with a standard error of 1.4 wt% using the visible spectral region (400 to 1100 nm), less than 0.5 wt% using the near infrared (NIR) spectral region (1100 to 2500 nm), and about 0.70 wt% using the mid infrared (IR) spectral region (400 to 4000 cm^{-1}). All the Phase 1 results were obtained using a single BY-104 salt cake simulant composition under a single set of experimental conditions.

While the Phase 1 results were encouraging, additional work was needed to evaluate the feasibility of the optical moisture determination under a wider range of conditions. Three additional parameters were investigated as part of this Task 2A. First, we wanted to study the effects of different sample particle sizes on the moisture determination. Since the actual samples from the waste tanks are expected to range in consistency from a rocky salt cake to a liquefied slurry, it is important to characterize any effect the sample particle size has on the spectral measurements, and ultimately on the moisture prediction.

Secondly, we wanted to investigate the experimental procedure used to prepare the calibration set samples. In all the Phase 1 studies the calibration samples were prepared by successively adding water to fully dried BY-104 simulant. In this Task we investigated the effect of preparing the samples at different moisture levels by drying "wet" BY-104 simulant for different periods of time. The main emphasis in this study was to determine if the sample preparation method effected the calibration model and the resulting moisture predictions.

The third parameter investigated was the effect of sample composition changes on the moisture determination. Since all the previous work was limited to a single composition, there was no information available about the sensitivity of the moisture determination to changes in the waste composition. It is known that the waste in the tanks will have different compositions depending on which tank, and where in the tank, the sample came from. In order to have any confidence in the optical moisture determinations, it is important to characterize the sensitivity of the method to changes in composition. The important experimental conditions, results, and conclusions of each of the three studies which comprise the Task 2A work scope are summarized below.

Particle size study

In this study three different particle sizes were investigated. The dried BY-104 simulant was size sorted using U.S. Series stacked sieves with 10, 20, and 40 mesh sieves. The sieving operation resulted in three particle size ranges; small particles that passed through the 40 mesh sieve (less than 420 μm), medium particles that passed through the 20 mesh sieve but not the 40 mesh sieve (420 μm to 841 μm), and large particles which passed through the 10 mesh sieve but not the 20 mesh sieve (841 μm to 2 mm). Calibration data sets were prepared for each particle size by successively adding water to obtain a range of moisture levels and measuring the VIS

and NIR spectra of each sample. The resulting data was analyzed for each particle size individually and as a combined data set containing all three particle size ranges. The results of these calibration models, and their performance in predicting the weight percent (wt%) water for the other particle sizes, are summarized in Table 1-1 and Table 1-2 below for the VIS and NIR spectral regions, respectively.

The boldface entries in the tables represent the individual model's root mean square error (RMSE) when used to predict with the same particle size data used to build the model. The first column in the tables is the number of PLS factors included in the model (determined by cross validation). The second row and column labels correspond to the different patricle size ranges as follows: Small = < 420 μm, Medium = 420 to 841 μm, Large = 0.841 to 2.0 mm, and Combined = < 420 μm to 2 mm. From Table 1-1 we can see that the one factor PLS model derived from the VIS spectra measured on the small particle size samples (< 420 μm) has a RMSE of 0.59 wt% when predicting moisture from small particle size spectra. The same small particle size model has RMSE errors of 5.17, 4.98, and 4.15 wt% moisture when used to predict moisture from the medium (420 to 841 μm), large (841 μm to 2 mm), and combined (<420 μm to 2 mm) particle size spectra, respectively.

# of Factors	Calibration Model Set	Prediction Set			
		Small	Medium	Large	Combined
1	Small	***0.59***	5.17	4.98	4.15
3	Medium	2.23	***0.34***	1.26	1.49
3	Large	3.78	0.95	***0.23***	2.26
3	Combined	0.79	0.83	0.77	***0.80***

Table 1-1. Root Mean Square Error (RMSE) for the prediction of wt% moisture from the VIS spectra using PLS.

# of Factors	Calibration Model Set	Prediction Set			
		Small	Medium	Large	Combined
2	Small	***0.61***	1.95	2.00	1.65
2	Medium	2.38	***0.47***	1.27	1.58
3	Large	4.85	1.86	***0.24***	3.00
4	Combined	1.03	0.65	0.52	***0.76***

Table 1-2. Root Mean Square Error (RMSE) for the prediction of wt% moisture from the NIR spectra using PLS.

The results of the studies indicated that there was a significant effect of the particle size on the moisture determination. This can be seen in the tables above where larger RMSE values were obtained when an individual particle size model was used to predict wt% moisture from the spectra of a different particle size. This effect was more pronounced for the VIS study results and less pronounced for the NIR results. However, the results also show that including a range of particle sizes in the calibration set effectively minimizes the influence of particle size on the overall calibration model. The combined particle size model had similar RMSE values for both the VIS and NIR data sets. In both cases, the RMSE for the combined particle size model was slightly larger than for any of the individual particle size models. It should be noted that even though there was an observed particle size effect on the moisture prediction accuracy, even the worst case was very close to the ±5 wt% moisture performace target stated by WHC (in the 10-25 wt% moisture range).

Composition study

The main purpose of this study was to evaluate the effect of variations in the chemical composition of the waste simulant on the optical moisture determination method developed at CPAC. In this study, the BY-104 salt cake simulant obtained from WHC was again used as the base chemical composition. Experiments were

performed where the BY-104 chemical composition was changed by adding amounts of the simulant's individual pure components to the base composition. For each change in the pure component level, the NIR spectral response was measured at different moisture levels. The results of changing the chemical composition were evaluated by observing the effect of the composition change on the moisture prediction from a calibration model derived from the base composition at different moisture levels. Further studies were conducted to develop estimates of the effect of the composition changes on the moisture determination. The experimental design used in this study allowed for the moisture level and composition effects to be studied independently and together. In all cases, the experimental data was analyzed using Partial Least Squares (PLS) regression to generate appropriate calibration models.

This study was the first conducted at CPAC to use a direct insertion fiber optic probe for the collection of the experimental spectra. The fiber optic probe consists of a fiber bundle to deliver the monochromatic illumination light from the spectrometer, a sampling window, and a collection fiber bundle to transmit the reflected light back to the detector. When considering the data without composition changes, a two factor model was able to predict moisture with a root mean square error (RMSE) of 0.734 wt% moisture. This prediction error estimate was slightly higher than the estimate from the previous Phase 1 study, 0.42 wt% moisture, but most of the difference was attributed to variability between the replicate experiments in the experimental design. The use of the fiber optic probe did not seem to adversely effect the moisture predictions and has several advantages over the previous method of using sample cups for measuring the spectra. Unfortunately, the experimental setup restricted the spectral region to the NIR region (1100 nm to 2500 nm) with the use of the fiber probe.

In analyzing the composition variation data, it was observed that the effect of changing the BY-104 component compositions did lead to increased prediction errors for the moisture calibration models. The nature of the composition effect was dependent both on the simulant component which was changing and the moisture level of the sample. There were two general effects that were observed. For the NaOH, $NaAlO_2$, Na_2SiO_3, and Na_3PO_4 components, the effect was a bias towards lower moisture predictions. This bias effect was larger at the higher moisture levels and was most pronounced for the NaOH component. The source of this bias was related to the increased spectral background due to the component spectra around the 1424 nm first OH overtone and 1925 nm OH combination band. The second main effect was observed for the $Fe(NO_3)_3$, $Ca(NO_3)_2$, and $Mg(NO_3)_2$ components which exhibited increased predicted moisture values. This effect was strongest for the $Fe(NO_3)_3$ composition changes and was mainly observed at the low to intermediate moisture levels. The main cause for this observed sensitivity was a decrease in the background adsorption in the 1400 nm to 1900 nm region due to the pure component spectra.

The changes in component compositions were effectively unknown interferents to the moisture calibration model. To illustrate the multivariate model's ability to correct for these interferents, a model was developed which included both moisture variation and NaOH concentration variation effects in the spectra. This model was able to correct for the NaOH composition change interferent and predict the moisture level without the bias observed for the pure BY-104 calibration model. The RMSE for the corrected model was 0.860 wt% moisture versus a RMSE of 4.80 wt% moisture for the model without the correction (the errors for the high moisture samples were even larger due to the strong bias of the uncorrected model). The corrected model regression coefficient vector indicated that accounting for the interference caused by the NaOH concentration variations resulted in an increased emphasis on the 1425 nm spectral region (OH first overtone band) and a decreased emphasis on the 1925

nm spectral region (OH combination band), relative to the single composition BY-104 moisture model. When all the composition variations were included in the model, the resulting model had a moisture RMSE of 1.41 wt% moisture compared to a RMSE of 0.734 wt% moisture for the moisture only prediction and a RMSE of 2.96 wt% moisture for all the compositions data predicted with the single BY-104 composition model. Therefore, by including the spectral variation due to the composition changes along with the moisture variations, the resulting model was able to correct for the interferences due to composition changes and reduce the prediction error by more than a factor of two. The price paid for the increased model complexity was that the moisture prediction errors were twice as large as the results obtained from the pure BY-104 moisture model with no composition variation. However, the resulting model is now much more robust to composition variations. The sensitivities for most of the BY-104 components (shown in Table 1-3 below) were below the statistically significant level over the composition variation ranges studied when the corrected model was used. The numbers in the table correspond to the expected moisture prediction error (in wt% moisture) to a one percent increase in the wt% concentration of the component.

The implication for future development and deployment of the spectroscopic moisture monitoring system is quite clear. It will be important to develop the moisture calibration models with a range of compositions expected to be encountered in the routine analysis of waste from the waste tanks. The exact sensitivity of the moisture prediction model will depend on the magnitude of the composition variation within the waste tanks, the waste components which are varying, and the desired moisture range of the calibration model. These studies indicate that by including composition variation in the calibration data, the resulting model can be made relatively insensitive to composition variation.

		Moisture Level	
Component	RMSE	15%	25%
$NaNO_3$	2.31	-0.45	-0.08
$NaAlO_2$	1.58	-0.04	0.32
NaOH	1.26	-0.06	0.16
Na_2SiO_3	1.55	0.03	0.29
$Fe(NO_3)_3$	1.69	-0.12	0.03
Na_3PO_4	0.77	0.00	-0.02
$Ca(NO_3)_2$	1.54	-0.01	0.08

Table 1-3. Sensitivity of moisture prediction to changes in component concentration for the overall model.

Finally, it appears that the same spectra used for the wt% moisture prediction can be used to predict the wt% NaOH in the same sample. The NaOH calibration model was seen to use mainly the first overtone of the OH band (at 1425 nm) for the wt% NaOH prediction. The RMSE of the NaOH prediction was 0.68 wt% and already included corrections for the water level interferent. Of course, it should be remembered that the NaOH calibration models only considered NaOH concentrations above the original BY-104 concentration and did not include corrections for the other component composition changes. Even so, the resulting NaOH prediction errors were quite good and suggests that the potential exists to monitor both the wt% moisture and wt% NaOH in a waste sample using a single spectroscopic measurement.

Sample drying study

In this study we investigated the effect of two different sample treatments in preparing the calibration data set for the moisture determination. The first sample treatment involved the complete drying of the BY-104 simulant followed by successive additions of water to obtain samples with different moisture levels. This was the

method used to obtain the calibration samples for the Phase 1 studies. The second sample treatment involved drying the wet BY-104 simulant sample for increasing periods of time. As the drying time was increased, the sample contained progressively less moisture. This second sample treatment, which we will call successive drying, was investigated to determine if there is a quantitative difference in the moisture model calibration results related to the two different methods of preparing the calibration samples. This study explored several variations of the sample drying experiment.

One of the issues was whether sample drying introduced changes in the spectral response that would effect the calibration for moisture. For the NIR results, we clearly saw that drying the sample below approximately two weight percent water induced a change in the spectral response. Based on our analysis we attributed this change to a surface enrichment in sodium nitrate and a depletion of sodium hydroxide as the sample became fully dehydrated. For the VIS spectral region, we were unable to satisfactorily answer this question since the recorded spectra appeared different from our earlier observations, even without the oven drying.

The second issue addressed by this study was the relative performance of the calibration models obtained with the different sample treatments. For the individual calibration models, there does not seem to be any advantage to either sample preparation technique. For the NIR spectra, both sample treatments gave very similar results in terms of the model fit to the calibration data. The drying models were able to predict moisture from the water addition spectra than the better reverse case (using the water addition model to predict moisture from the drying data). However, much of that difference can be attributed to the poor predictions for the low moisture samples from the drying experiments. Since those samples exhibited unique spectral features, related to the nitrate migration, not present in the water addition

data, it is difficult to say one method is better than the other. The cross prediction results in the VIS spectral region were considerably worse than in the NIR region.

Conclusions

The recurrent theme from these studies is that a calibration model based on a single set of tightly controlled experimental conditions will tend to have somewhat larger prediction errors when applied to samples collected with variations outside those conditions. However, the moisture calibration models are inherently quite good and can tolerate a range of conditions and still give prediction results within the ±5 wt% moisture tolerance specified by WHC for the tank moisture analysis. It was demonstrated, from both the particle size and compositions studies, that a model built from a range of sample particle sizes or compositions is much more robust when predicting new samples which are slightly different than the "reference" samples. When the more general models are built by including the expected sources of variation, the model fit statistics are not as favorable, i.e., a slight increase in the model RMSE, but the resulting models can safely be applied to a wider range of new unknown samples.

What this means to the optical moisture determination project at WHC is that, while more work should be done to fully characterize the model response to the anticipated range of sample conditions in the waste tanks, the NIR and VIS models are able to provide reasonable prediction accuracy over a range of sample conditions. These studies indicate that including more sources of variation in the calibration model results in more robust prediction models at the expense of increased overall prediction error. The larger errors observed when more sources of variation were included in the model are, in part, a function of the relatively small number of samples considered in the studies. The effort to fully characterize the composition

effect could take many more samples but the increased information from such a large data set could lower the overall prediction errors. These studies indicate that while moisture prediction errors of 1-2 wt% are certainly attainable under laboratory conditions, that level of accuracy may not be possible over a larger range of compositions and sample conditions found in the waste tanks. However, since the acceptable accuracy of the moisture determination has been stated as ±5 wt% moisture, these studies indicate that the potential for success with the optical detection method is fairly high. Even with the very large composition variations considered in this study, the errors of the models are still about half of the 5.0 wt% requirement.

One of the limitations in the waste tank analysis program is the limited amount of hard data on the actual composition variations within the tanks. This implies that the type of composition variation experiments presented here should be conducted with simulants to obtain estimates of the sensitivity of the calibration models to waste component variations. More emphasis and resources should be devoted to the composition variation characterization experiments than to other parameters such as particle size and sample treatment, which have much smaller effects.

2. Particle Size Study

Results from the previous Phase 1 study indicated that moisture could be determined with a standard error of 1.4 wt% using the visible (VIS) spectral region (400 to 1100 nm), less than 0.50 wt% using the near-infrared (NIR) spectral region (1100 to 2500 nm), and about 0.70 wt% using the mid infrared (IR) spectral region (400 to 4000 cm^{-1}). The purpose of this study was to determine the effect, if any, sample particle size variation has on the prediction of moisture from the spectral data. In this study three different particle sizes were investigated. The dried BY-104 simulant was size sorted using U.S. Series stacked sieves with 10, 20, and 40 mesh sieves. The sieving operation resulted in three particle size ranges; small particles that passed through the 40 mesh sieve (less than 420 μm), medium particles that passed through the 20 mesh sieve but not the 40 mesh sieve (420 μm to 841 μm), and large particles which passed through the 10 mesh sieve but not the 20 mesh sieve (841 μm to 2 mm). A calibration data set was prepared for each particle size by adding different amounts of water to the dried samples and measuring the resulting sample's spectra in the visible and near-infrared spectral regions. Calibration models were developed for each of the particle size calibration sets as well as for a combined set containing all three particle size ranges. The results of these models, and their performance in predicting the weight percent (wt%) water from the spectra of other particle sizes, indicate that particle size does impact the moisture determination. However, the results also show that including a range of particle sizes in the calibration set effectively minimizes the influence of particle size on the overall calibration model.

Experimental

The BY-104 simulant material supplied by WHC was dried in an oven at 160° C for approximately forty-eight hours and then cooled in a desiccator. The dried starting material was ground using a ceramic mortar and pestle to break up the aggregate "clumps" that form during the drying process. The dried and ground starting material was then separated into three different particle size ranges using U.S. Series standard mesh sieves. The first sieve (U.S Sieve #10) had 2 mm openings, the second (U.S Sieve #20) had 841 μm openings, and the third (U.S Sieve #40) had 420 μm openings. The sieving operation resulted in base materials with three particle size ranges, small (<420μm), medium (420-840 μm), large (840μm-2mm), which were then used in the moisture determination studies. The size sorting was accomplished by stacking the three sieves in increasing mesh number (decreasing opening size) and then shaking the dried BY-104 through the stack. This results in an increasingly smaller particle size range being trapped at each succesive sieve layer. [The larger particles which were trapped by the first sieve (>2mm, #10 mesh) were not used in this study.]

To determine the amount of water added to the sample at each step, the samples were weighed before and after the water additions. For each particle size, the first two samples (first four spectra) contained zero weight percent water and the remaining eight samples contained increasing percentages of water. The water was added using a spray bottle and the sample was thoroughly mixed using a putty knife following the water additions. Although the exact amount of water needed for a given weight percent of water was not calculated, the approximate target moisture range was zero to twenty percent water. An arbitrary amount of water was added and the weight difference measured. From this information the exact weight percent water was calculated.

The experimental spectra were collected on the same spectrometer used in the previous Phase 1 studies. Each optical measurement of the sample consisted of thirty-two reflectance scans co-averaged to produce a single spectrum, Spectra from both the visible (400 nm to 1100 nm) and NIR (1100 nm to 2500 nm) wavelength regions were measured during each scan. The spectral reference used in these experiments was the internal ceramic reference built into the spectrometer. For each sample two spectra were recorded. The sample cup was rotated between the two spectral scans in order to reduce the differences due to sample cup packing. Since there were three sets of ten samples (ten moisture levels at each of the three particle sizes), and each sample was scanned twice, a total of sixty spectra were taken during the experiment.

All data were analyzed on a computer workstation using custom chemometrics routines developed at CPAC. The spectra for the sixty samples was split into the visible and NIR spectral regions and analyzed separately. Calibration models were developed for the two sets of spectra using partial least squares (PLS) regression to relate the spectral responses to the wt% moisture values. The data for the sixty samples were then broken into sets of twenty, based on the different particle sizes, and run through the same series of calculations. In all cases, the results obtained from the PLS models were calculated from the second derivative spectral data. As discussed in the Phase 1 report, the second derivative of the spectra was used to eliminate baseline offsets in the reflectance spectra due to scatter effects.

Results and Discussion

In analyzing the data from the particle size experiments we will first consider the visible (400 nm to 100 nm) and then the near-infrared region (1100 nm to 2500 nm). For each spectral region we are interested in discovering if any changes in the measured spectra occur which are related to the particle size of the sample. We are

most interested in determining if those spectral variations due to different particle sizes have an effect on the performance of the moisture prediction models.

Visible Results

The visible spectra were first analyzed as a combined set including all three sieve sizes. Cross validation indicated that a three factor PLS model gave the best prediction results without overfitting the data. The amount of variance described by each factor in the spectral response and moisture concentration blocks is given in Table 2-1 below.

	Spectral Response		Moisture Concentration	
Factor	Individual	Total	Individual	Total
1	90.06	90.06	75.60	75.60
2	7.99	98.04	12.31	87.91
3	1.08	99.13	10.32	98.23

Table 2-1 Percent variance described by each factor of the PLS model for the combined particle size VIS data set.

The first factor scores for the spectral block are shown in Figure 2-1. In this plot one can clearly see three different groups, or clusters, of score values corresponding to the three particle sizes. The first four samples within each group (i.e., samples 1-4, 21-24, and 41-44) correspond to the zero water added samples. Notice that these zero water samples do not have the same scores on this factor. While the medium and large particle size samples have roughly equivalent scores, the score values for the small particle size samples are different. Nominally, the only difference among these samples is that they have been particle size sorted. In the absence of a particle size effect, one would expect that all twelve samples of the dry base material would have similar scores. Also note that the score values are generally increasing within each

particle size group corresponding to the increasing water content of the samples. It is interesting that the small particle size scores are more linearly dispersed over the range of samples while the medium and large particle size sample scores are more nonlinear at the higher water concentration samples. Therefore, this factor can be interpreted as describing the main moisture concentration response of the spectra over all the particle sizes. This description of the response is much more linear over the range of water concentrations for the small particle size samples than for the two larger particle sizes.

The corresponding loading vector for the first factor is shown in Figure 2-2. The loadings are dominated by the peaks at 450 nm, 550 nm, and 700 nm. This is the characteristic plot of the moisture response of the BY-104 material in the visible region of the spectra.

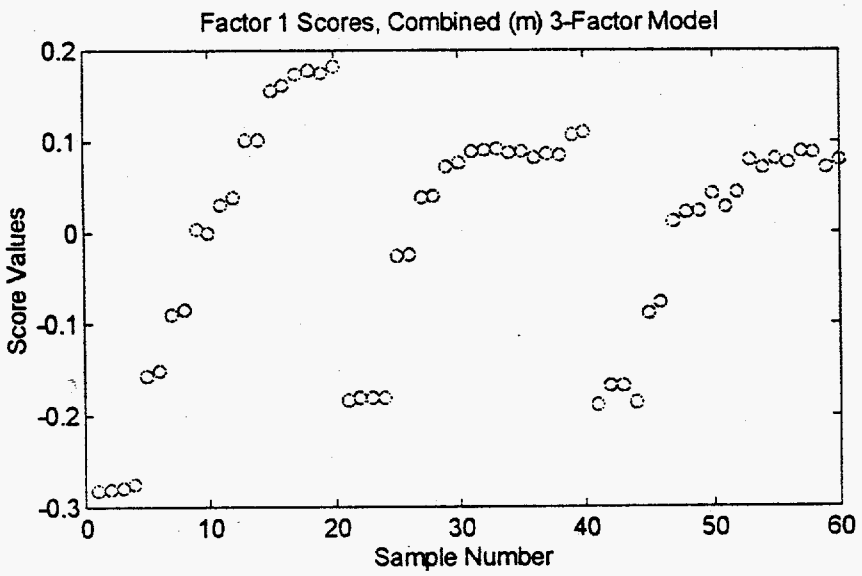


Figure 2-1. Sample scores on the first factor of the combined data set VIS model.

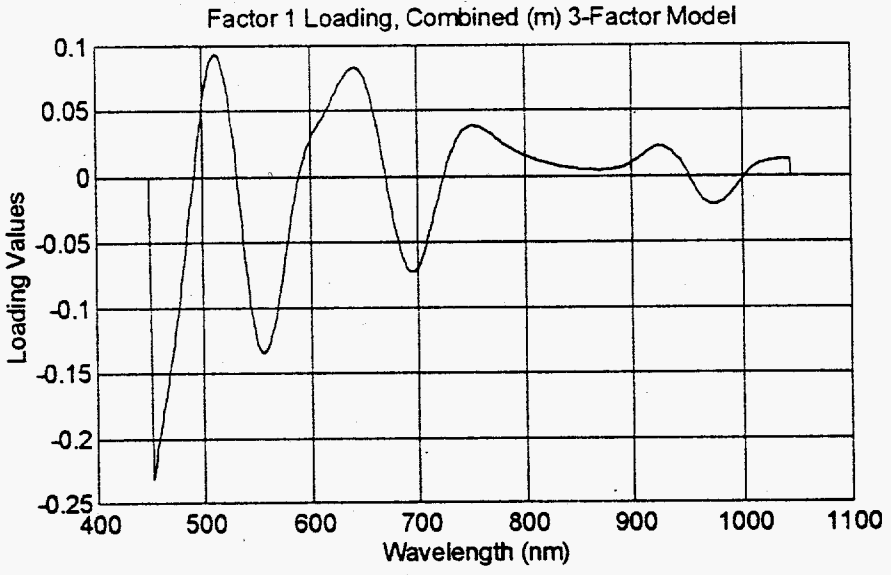


Figure 2-2. Loading vector for the first factor of the combined particle size VIS model.

The sample scores on the second factor of the model are shown in Figure 2-3. This factor is mainly describing the difference between the small particle size samples and the medium and large particle size samples in their spectral response to changes in the water concentration. Besides the differences between the small and the two larger particle sizes, this factor is also describing corrections to the moisture response within each group of particle sizes. It is interesting that the pattern of scores is so different between the three size groups. For the small particle size data, this factor indicates a mostly linear, decreasing, trend over the range of water concentrations. For the medium particle size samples, their scores indicate a nonlinear response in the intermediate water concentrations. For the large particle size samples, the scores indicate a very weak trend from low to high water content. Recall from Table 2-1 that this factor accounts for about 12% of the concentration information and about 8% of the spectral information. Most of that 12% seems to be related to moisture response of the small particle size samples. The corresponding loading vector plot for this factor is shown in Figure 2-4.

The sample scores on the third model factor are shown in Figure 2-5 below. This factor only describes about one percent of the spectral information but accounts for about ten percent of the wt% water concentration information. Again, we can see that this factor of the model is describing very different information for the small particle size data than it is for the medium and large particle size samples. This factor contains little information related to the small particle size samples (their scores are centered around zero and show a slight negative trend from 0.1 to -0.1) other than the sharp discontinuity around samples 9 to 12. For the other two particle sizes, the scores for the samples with added water (recall that the first four samples within each particle size are zero wt% moisture) show a strong trend of increasing score values with increasing sample moisture. From this plot, it appears that most of the 10% of the moisture level information described by this factor is due to the

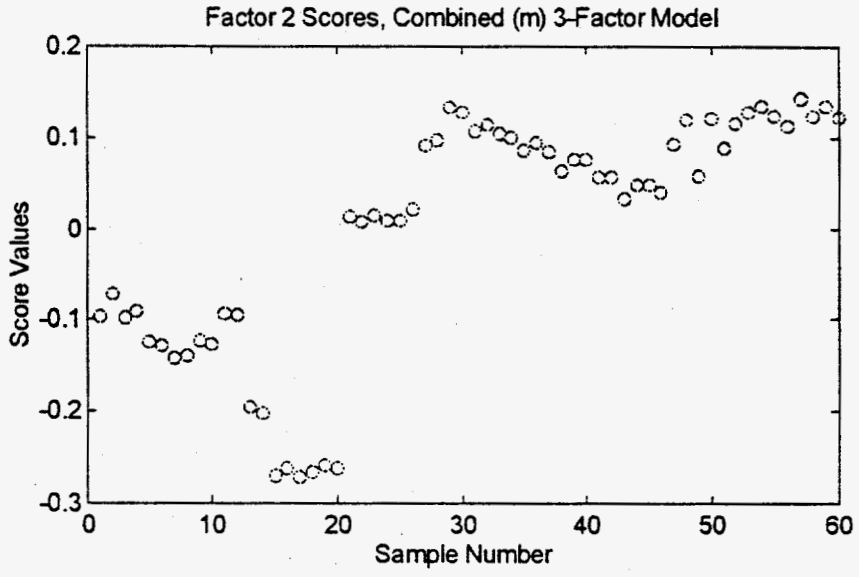


Figure 2-3. Sample scores on the second factor of the combined particle size VIS model.

spectral response of the added water content samples in the medium and large particle sizes.

The spectral loadings for this third model factor are shown in Figure 2-6. This plot shows that the information described by this factor is mainly related to the water peak at 950 nm with additional contributions from the peaks at 700 nm, 510 nm and 575 nm.

Additional interpretation of the PLS model can be obtained by plotting the scores in the spectral and moisture blocks against each other for each factor. Figure 2-7 shows the spectra versus moisture scores for the first factor of the model. There are two interesting things to note in this plot. First, the small particle size samples (sample numbers 1-20) follow the linear inner regression line (the solid line in the plot) between the two blocks but are offset by a constant amount above the line. Secondly, the medium particle size samples (21-40) and the large particle size samples (41-60)

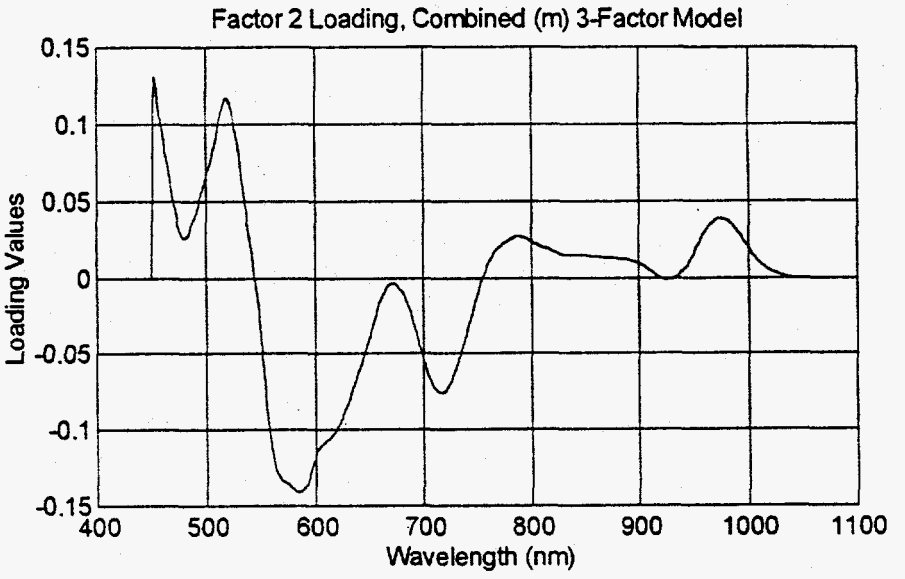


Figure 2-4. Loading vector for the second factor of the combined particle size VIS model.

are not well modeled by this first factor and their nonlinear response can clearly be seen in the curvature they exhibit about the inner relation line.

At this point it should be mentioned that nonlinear response was also observed in the original Phase 1 calibration data set although the magnitude was not as great. It appears that if the medium and large particle size samples were removed from the data, a single factor linear model might suffice to describe the relationship between the spectral responses and the moisture concentration for the small particle size samples.

The spectra versus moisture scores on the second factor of the model are shown in Figure 2-8. This plot is interesting in that it shows the clear difference between the small particle size samples and the medium and large particle size samples. The spectral scores (dispersed along the x-axis) show positive score values for the small particle samples and negative scores for the other two sets of samples. The moisture

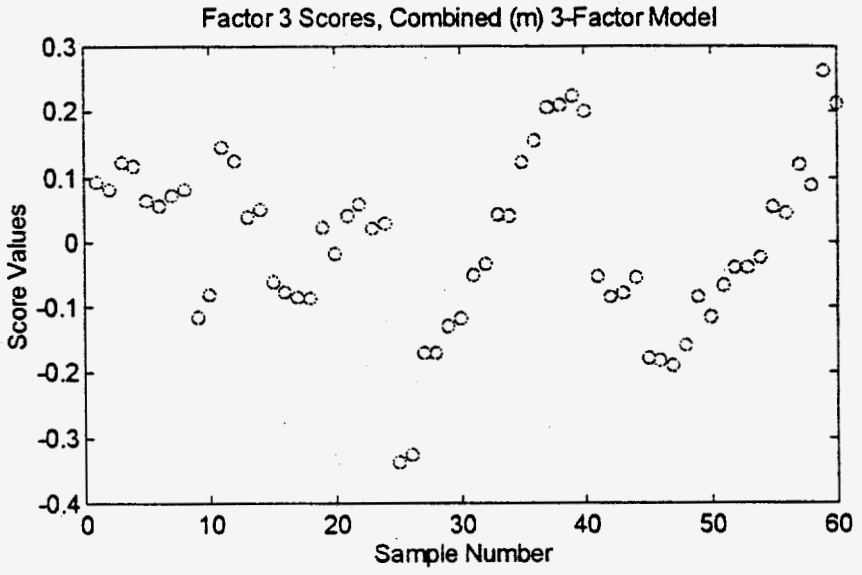


Figure 2-5. Sample scores on the third factor of the combined particle size VIS model.

scores (dispersed along the y-axis) contain much more information about the medium and large particle size samples.

From our interpretation of the spectral scores (discussion of Figure 2-3 above) we concluded that the second factor was mainly describing the difference between the small particle size spectra and the spectra of the other two larger particle sizes. This second factor actually contains very little information which can distinguish between the different water concentrations in the small particle size data (a conclusion obtained from observing the very small range of variance in the Y-block scores for the small particle samples). In this plot it is clear that most of the information correlated to the moisture content is related to the difference in the spectra of the small particle size samples. The small variations in the spectra for the medium and large particle size samples described by this factor are almost totally uncorrelated to the changes in moisture level.

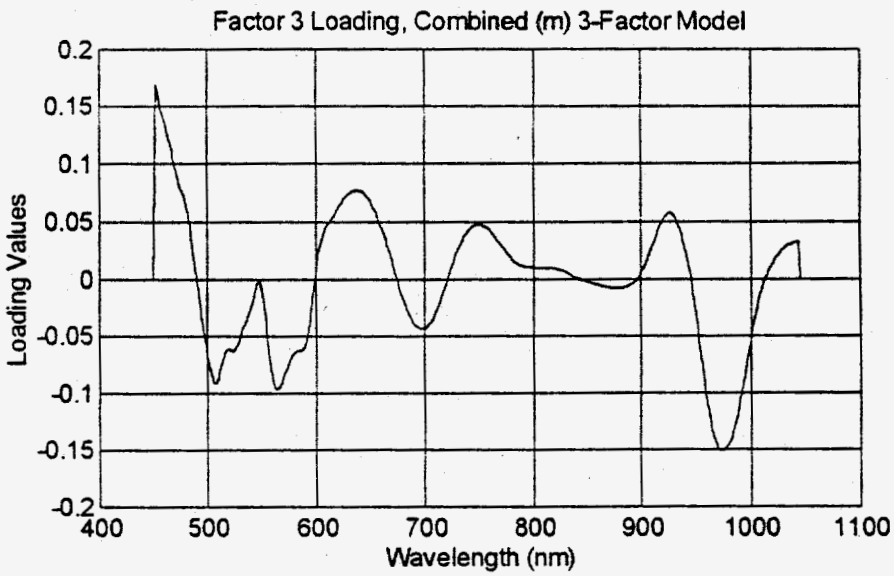


Figure 2-6. Loading vector for the third factor of the combined particle size VIS model.

In the third factor scores we see the linear relationship between the X-block and Y-block (spectra and moisture concentration, respectively) as shown in Figure 2-9. The first factor of the model described the major spectra/moisture relationship for all three particle size samples but was somewhat biased by the nonlinear behavior of the two larger particle size samples. The second factor described mainly the differences between the two larger particle sizes and the small particle size spectra. Once the information from the first two factors was removed, the remaining linear relationship between the spectra and moisture level was exposed for all three particle sizes in the third factor of the PLS model. Recall that this factor is mainly describing the moisture information for the two larger particle sizes and little of the small particle size moisture information.

The three factor PLS model gave a root mean standard error (RMSE) of 0.79 wt% water for the combined data set. This error can be apportioned among the different particle sizes by predicting each set using the combined particle size model. This

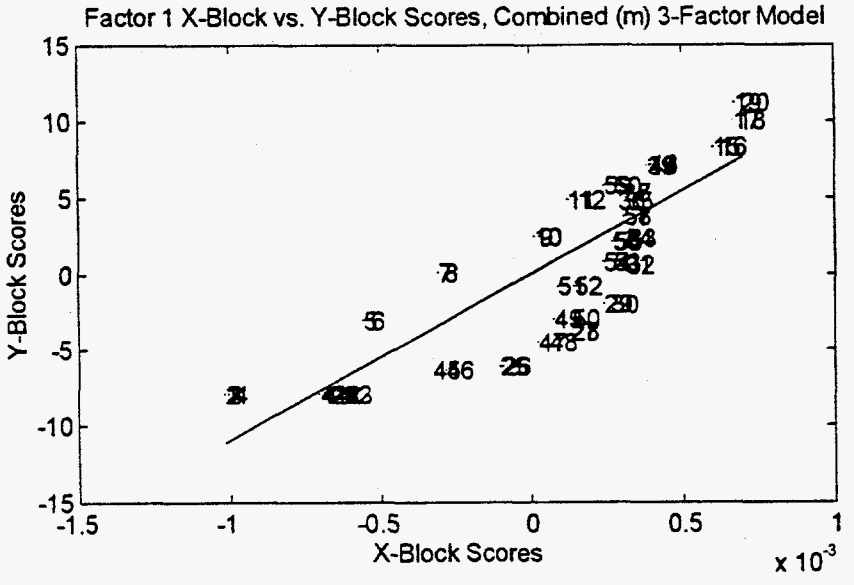


Figure 2-7. Spectral (X-block) versus moisture (Y-block) scores on the first factor of combined VIS model.

gives a RMSE of 0.77, 0.82, and 0.77 wt% water for the small (< 420 μm), medium (420 - 841 μm), and large (0.841 - 2 mm) particle sizes, respectively. Therefore, one can conclude that the combined particle size model was able to model the spectral response to changes in moisture level for all three particle sizes equally well. These prediction results are shown in Figure 2-10. The solid line in the plot represents the ideal 45° linear fit line, the broken lines are the actual least squares fit lines for the various particle sizes.

In order to estimate the effect particle size has on the predictive ability of the PLS models, separate calibration models were developed for each of the three particle size data sets. Then, the model derived from one particle size was used to predict moisture from each of the other particle size spectra. The combined data set was also predicted with each of the single particle size models. The results are shown in Table 2-2 below.

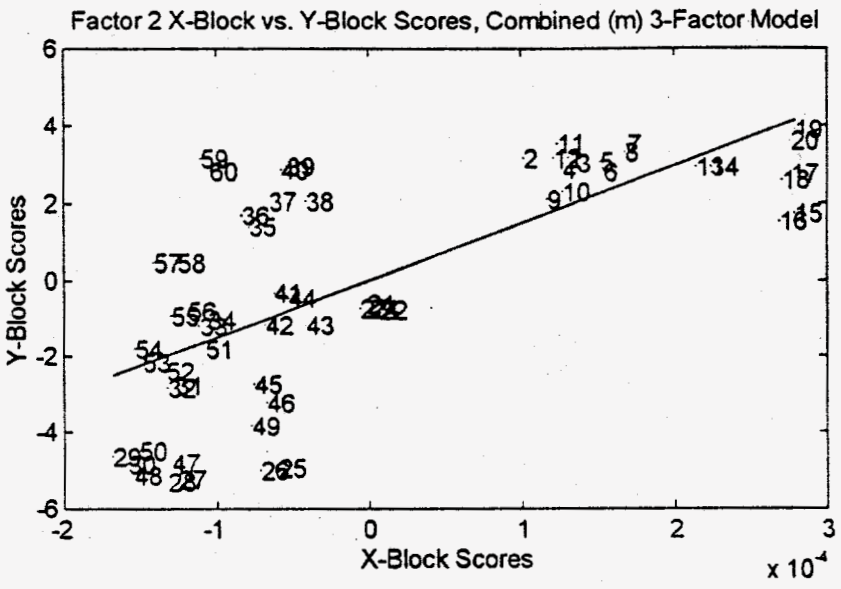


Figure 2-8. Spectra and moisture scores plotted against each other for the second factor of the combined PLS model.

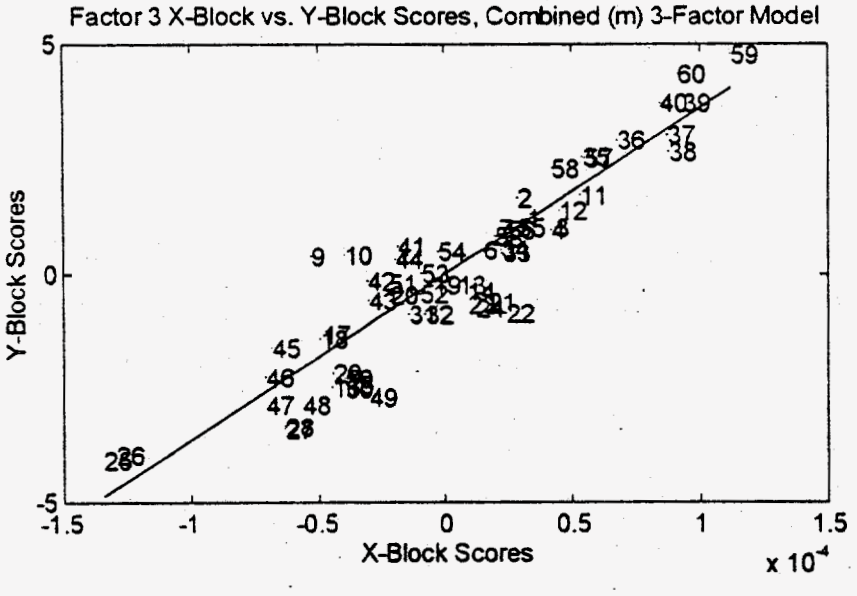


Figure 2-9. Relationship between the spectra and moisture described by the third factor of the combined VIS model.

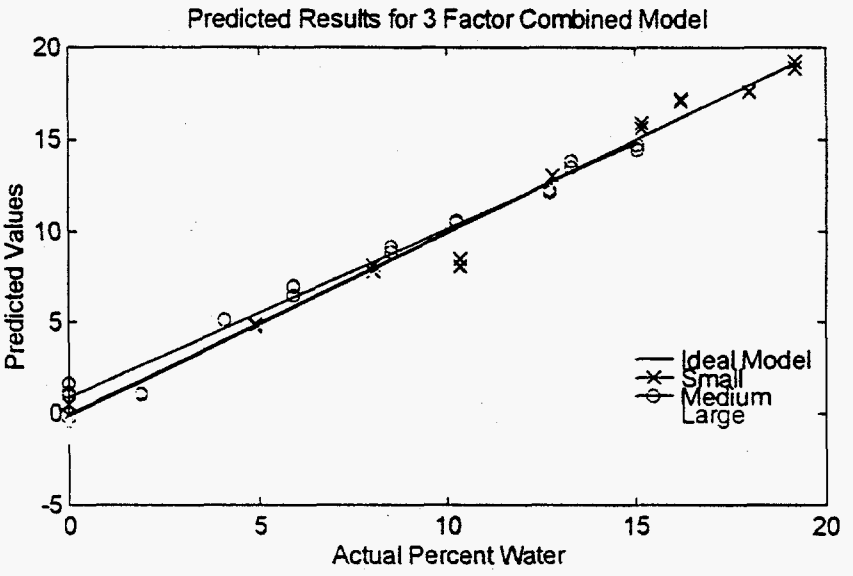


Figure 2-10. Prediction results for the three particle size data sets on the three factor PLS model derived from the combined VIS data set.

# of Factors	Calibration Model Set	Prediction Set			
		Small	Medium	Large	Combined
1	Small	***0.59***	5.17	4.98	4.15
3	Medium	2.23	***0.34***	1.26	1.49
3	Large	3.78	0.95	***0.23***	2.26
3	Combined	0.79	0.83	0.77	***0.80***

Table 2-2 Root Mean Square Error (RMSE) for the prediction of wt% moisture from the visible spectra using PLS. Boldface values are the fit RMSE for the individual models.

In all cases restricting the calibration to a single particle size improved the predictive ability for that particle size relative to the combined model. However, using the model for one particle size to predict the moisture from the spectra of the other particle sizes gave results that were always worse than the combined model. The small particle size model was the worst at predicting moisture from the spectra of the

other particle sizes. Adding additional factors to the small particle size model did not improve its predictive ability for the other sizes. The medium particle size model was the best at predicting the other sizes but still resulted in errors for the small particle size data which were about three times greater than the combined model and about twice as large for the large particle data. The large particle size model was able to predict the medium data set with only slightly larger errors, relative to the combined model, but did quite poorly on the small particle size data. Much of the poor results obtained from the cross prediction studies can be attributed to the relatively large nonlinear component of the spectral response to changes in moisture level observed in the medium and large particle size data. The small particle size data did not have as pronounced a nonlinear behavior and so the small model cannot effectively model the nonlinear response of the medium and large particle sizes. Conversely, the medium and large particle size models need to include strong nonlinear correction terms which are not appropriate for the small particle size data.

Plots of the prediction results are shown in Figure 2-11, Figure 2-12, and Figure 2-13 for the small, medium, and large particle size models, respectively. For the small particle size model, information about the nonlinear spectral response to moisture content was not present in the small particle size data set and so could not be included in the PLS model. The result is that most of the medium and large particle size samples appear to have roughly the same water content when predicted with the small particle size model, as shown in Figure 2-11. Both the medium and large particle size data exhibit almost identical results when predicted with the small particle model. This would seem to suggest that there is some mechanism related to particle size that accounts for the observed nonlinear spectral responses at higher water concentration.

While the exact mechanism that gives rise to the difference between the small particle data and the data for the two larger particle sizes is not clear, there are

several possibilities. It could be simply increased scatter from the larger particles or the change in the surface to volume ratios that indirectly affect the measured spectra. The effect may also be due to some interstitial, or ionic, chemistries between the small particles and the water that causes the spectral response to moisture to behave differently. Differences in surface water and bulk absorbed water may also account for the observed effects. A more detailed study, beyond the scope of this work, would be required to answer these questions.

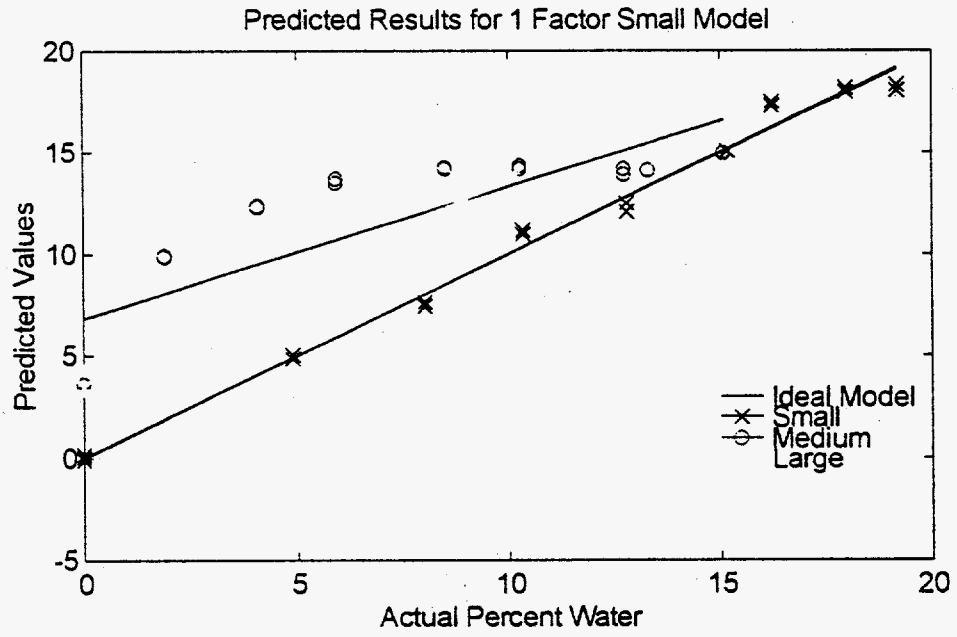


Figure 2-11. Results obtained by predicting the small, medium, and large visible data using the small particle PLS model.

For the medium particle size model, the prediction results for the small particle size data are much worse at the high water concentration, as seen in Figure 2-12. Part of this effect might be due to the fact that the water concentrations for the small particle size data extend beyond the range of the medium data set, requiring extrapolation for the higher water content samples. The more likely cause is the lack

of the nonlinear response at high water content in the small particle size data. From our earlier interpretations for the combined particle size PLS model, we know that the medium and large particle size spectral response to water drops off faster, in a nonlinear way, than the response for the small particle size spectra. Therefore, a model which includes corrections for this nonlinear drop-off would tend to overestimate the water concentration for samples where the nonlinear decrease is not present. The prediction results for the small particle size data are still fairly linear over the concentration range. The prediction results for the larger particle size data, from the medium particle size model, show a significant bias for the low water content samples but fairly good results for the higher water content samples.

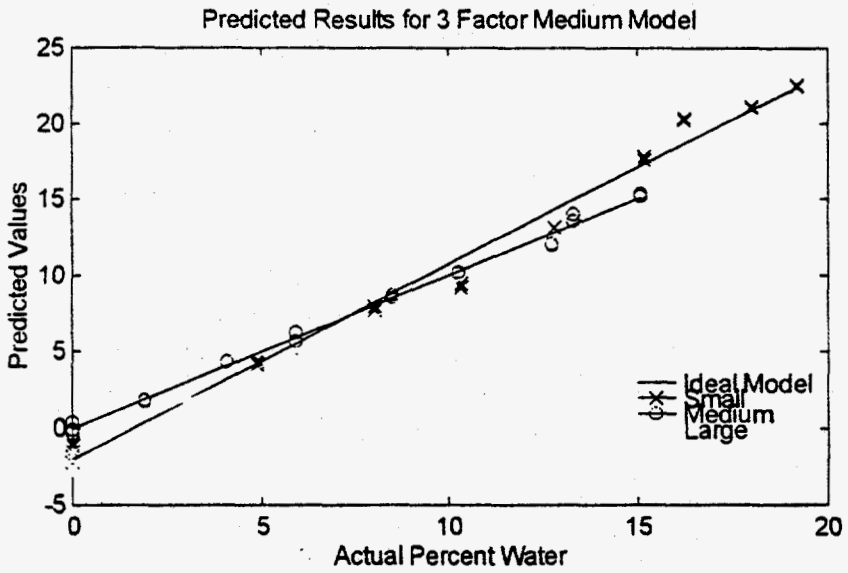


Figure 2-12. Prediction results obtained for the small, medium, and large VIS data using the medium particle size PLS model.

For the large particle size model, the prediction results for the medium particle size data was fairly good over the range of water concentrations, as shown in Figure 2-13. However, for the small particle size data the prediction results were quite poor with a significant offset for the higher water content samples. Even with the poor

performance in applying the large model to the small samples, it is interesting to note that the model still predicts the zero water samples adequately and describes the linear relationship between the spectral response and the water content. It is the difference in slope between the large particle size model and the small particle size prediction results that indicate a significant bias for the small particle size results.

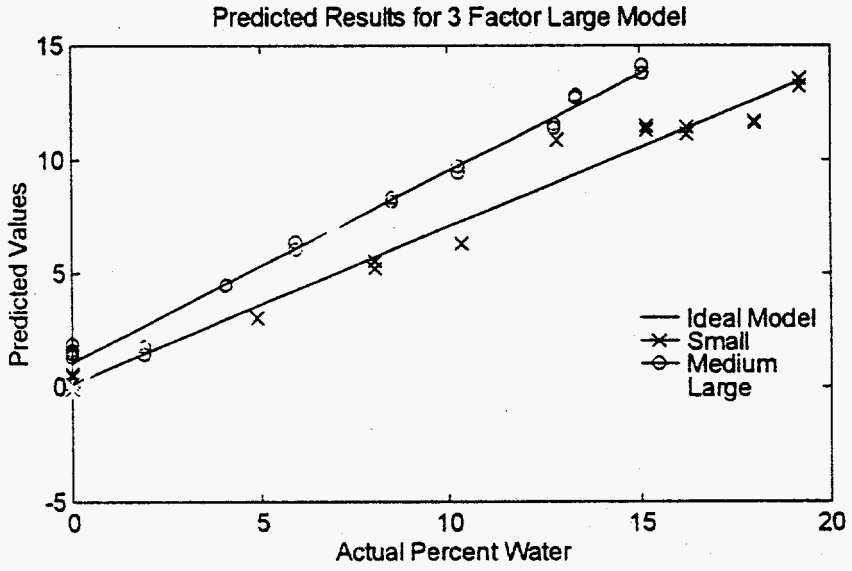


Figure 2-13. Prediction results obtained by predicting the small, medium, and large VIS data using the large particle PLS model.

Based on these results, it is clear that the particle size of the samples will have an important effect on the relationship between the spectral responses and the moisture concentration. It is also clear that by including a range of particle size data into the calibration model, the models are able to give adequate prediction results for all three particle sizes. The models developed from each of the three single particle size data sets, while able to do a much better job of describing the data on which they were built (the prediction errors were typically less than half of what was obtained using the combined particle size data), were not as useful in modeling the other particle

size data as the combined particle size model. This points out the need to include the range of expected particle size distributions in the calibration set data.

NIR Results

For the NIR data (1100 to 2500 nm) we again built a PLS model for the combined data set containing the data for all three particle sizes. In this case, the cross validation studies indicated that a four factor PLS model was needed to produce the best predictive model. The amount of variance, or information, described by each factor of the model is given in Table 2-3 below.

	Spectral Response		Moisture Concentration	
Factor	Individual	Total	Individual	Total
1	98.18	98.18	94.61	94.61
2	1.19	99.30	1.67	96.28
3	0.21	99.51	1.40	97.69
4	0.21	99.73	0.65	98.34

Table 2-3. Percent variance described by each factor of the PLS model for the combined particle size NIR data set.

The first factor of the model describes over 90 percent of the information in both the spectral and water concentration data. The second factor describes about the same amount of information in both the spectral and concentration blocks. The third and fourth factors each describe the same minor amount of spectral information but the third describes about twice as much water concentration information as the fourth factor.

Compared to the visible combined particle size model discussed above, the NIR model includes much more of both the spectral and water concentration information in the first factor. Recall from Table 2-1 that the second and third factors of the visible

model each contained over ten percent of the concentration information. This indicates that the NIR spectra are more closely linear (requiring much smaller corrections from the additional factors) than the visible spectra. This agrees with the results from the earlier studies. In the original study, a single factor PLS model was able to adequately model the NIR spectral response to water concentration variation. While the cross validation studies on this data suggest a four factor model, the contribution of the second, third, and fourth factors is relatively minor in comparison to the first factor and are only expected to provide small corrections to the overall model.

Again, we can look at the sample's score plots for each model factor to gain some insight into the details of the model. The sample scores for the first spectral response (X-block) factor are shown in Figure 2-14 below. In this plot one can clearly see the major trends corresponding to the three different particle size experiments. For each of the particle sizes, the scores start out near -0.2 for the four samples (which correspond to the dry, or zero water added, samples), then the scores show a gradual increase in value as the water content is increased, and finally they exhibit a nonlinear tapering off at the higher water concentrations. The scores for the small particle size samples (samples 1-20) with high water concentration are higher than either the medium (samples 21-40) or the large (samples 41-60) sample scores. This in mainly due to the higher water concentration in the experimental design for the small particle size data.

It is interesting to note that all three particle sizes exhibit similar nonlinear response at higher water concentrations, although the onset of the nonlinear behavior appears to start earlier with increasing particle size. This nonlinear trend was not observed in the original NIR data from the earlier study. This may indicate some systematic error associated with the larger volume water additions and subsequent calculation of the reference weight percent water values in this study. The addition of water

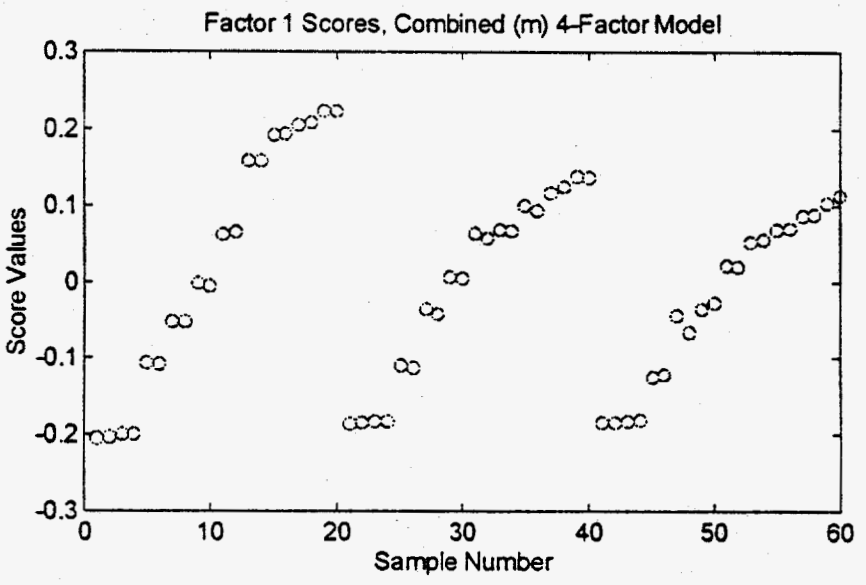


Figure 2-14. Sample scores on the first NIR spectral block factor of the combined particle size NIR model.

using the spray bottle made it easy to end up with unmixed water on the walls of the sample container and this would show up as artificially high wt% water values. This in turn would lead to a seemingly nonlinear response of the spectra to increasing water content since not all of the calculated water would have made it into the sample.

The corresponding loading vector for the first factor of the combined particle size NIR model is shown in Figure 2-15 above. This plot shows that the first factor consists of the information from the spectral regions around 1400 nm (the first OH overtone), 1900 nm (OH combination band), and 2250 nm (OH combination band). This loading vector is clearly the main moisture contribution to the NIR spectra.

The sample scores for the second factor of the combined particle size model are shown in Figure 2-16. This plot is very interesting because it shows different (almost opposite) behavior between the small and large particle size samples. The small

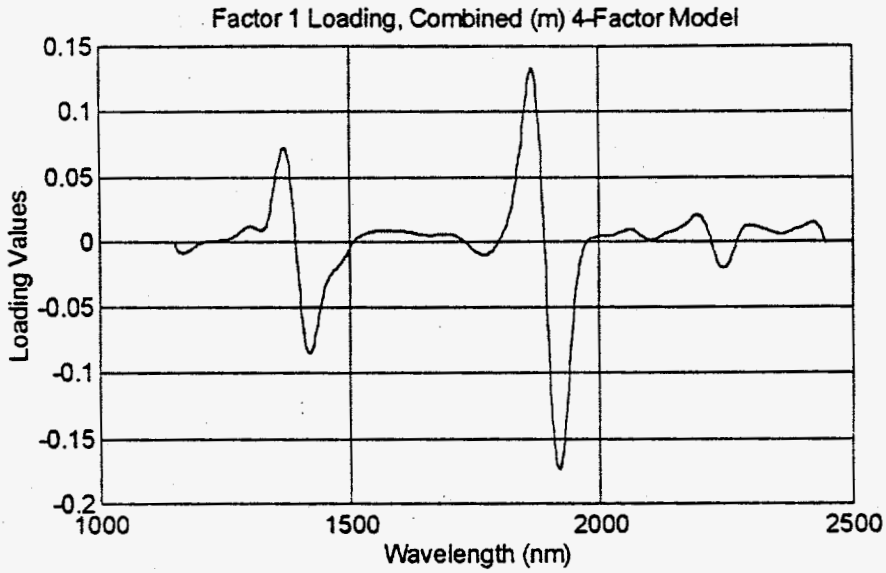


Figure 2-15. Spectral loadings for the first factor of the combined particle size NIR model.

particle size samples have large negative scores for the low moisture content samples, intermediate scores for the mid-range water concentration samples, and near zero scores for the high water content samples. Conversely, the large particle size data have high positive scores for the low water content samples, intermediate scores for the mid range water concentrations, and near zero scores for the high water content samples. The medium particle size samples do not contribute much at all to this factor (scores for samples 21-40 are all near zero). Clearly, this factor is describing some functional difference between the low water concentration sample's spectral response to water in the small and large particle size data. The interesting thing is that the function is inverted depending on the particle size. This may be due to the different surface to volume ratios and its effect on the water spectra between the different particle sizes and could be related to free versus bound water.

The spectral loadings for this second factor are shown in Figure 2-17 along with the NIR second derivative spectra of pure $NaNO_3$. The $NaNO_3$ spectra was scaled (i.e.,

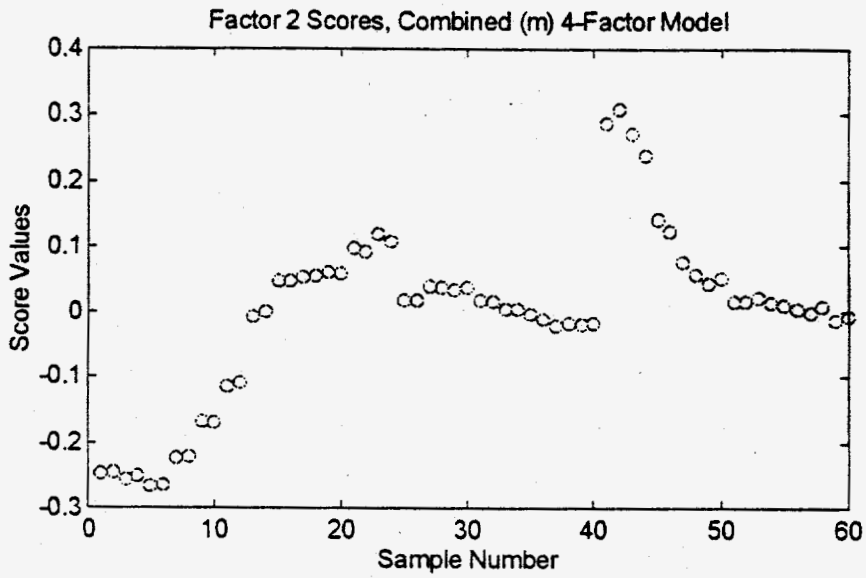


Figure 2-16. Sample scores on the second spectral block factor of the combined particle size NIR model.

multiplied by 200) to match the spectral intensities of the loadings. It is clear that this factor is not describing any moisture level information. Rather it seems to be describing the difference in $NaNO_3$ concentration between the small and large particle size dry samples. Based on this plot, and the corresponding scores shown in Figure 2-16, it seems that the initial large particle size samples are enriched in $NaNO_3$ relative to the initial small particle size samples. Since these samples were drawn from the same lot of the BY-104 simulant, it seems likely that the change in $NaNO_3$ concentration arises from some mechanism associated with the subsequent sample treatment steps. The only sample treatment in this study involved the oven drying and particle size sorting. The samples were oven dried before and after the particle size sorting with the sieves. It seems likely then that the source of the $NaNO_3$ enrichment comes from the migration of the nitrate to the surface of the larger particles during the oven drying step. This has not been tested but the drying studies planned as part of this Task 2A work scope may confirm this hypothesis.

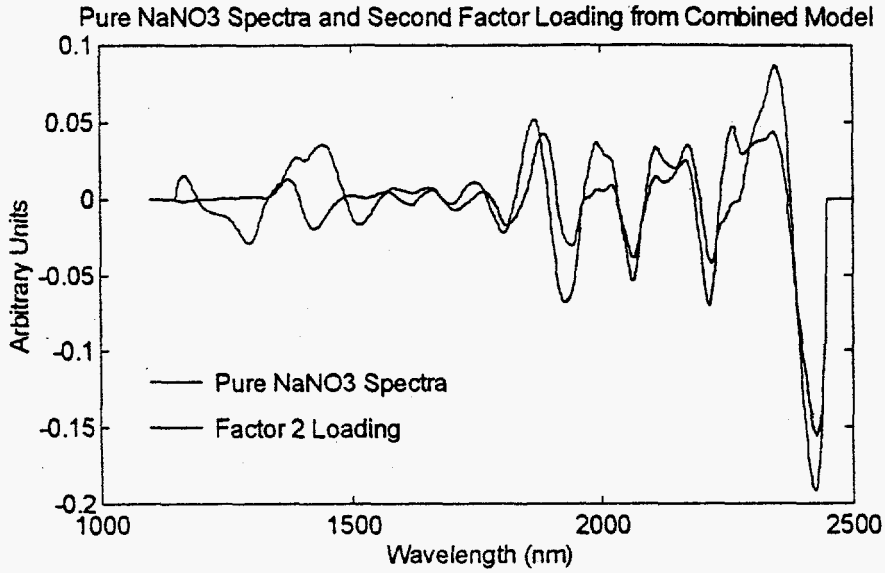


Figure 2-17. Loading vector for the second factor of the combined particle size model along with the spectra of $NaNO_3$.

The sample scores for the third and fourth factors are shown in Figure 2-18 and Figure 2-19, respectively. The third factor clearly describes the nonlinear behavior of the spectral response to changes in the water content. Within each particle size, the high and low water content samples have similar scores and the factor mainly describes the difference between the samples at either extreme of the moisture concentration range and those samples with intermediate moisture content. Therefore, this third factor of the model can be considered a correction term to the overall PLS model to account for nonlinear responses of the spectra to water concentration changes. The fourth factor seems to be describing additional differences between the zero water samples and the samples with water added for the medium and large particle size data. Again, this factor can be viewed as a small model correction factor to correct for specific nonlinear behavior. Both the third and fourth factors of the model contributed relatively minor amount of information to the overall moisture prediction model. The loadings for factors three and four are not shown in this report.

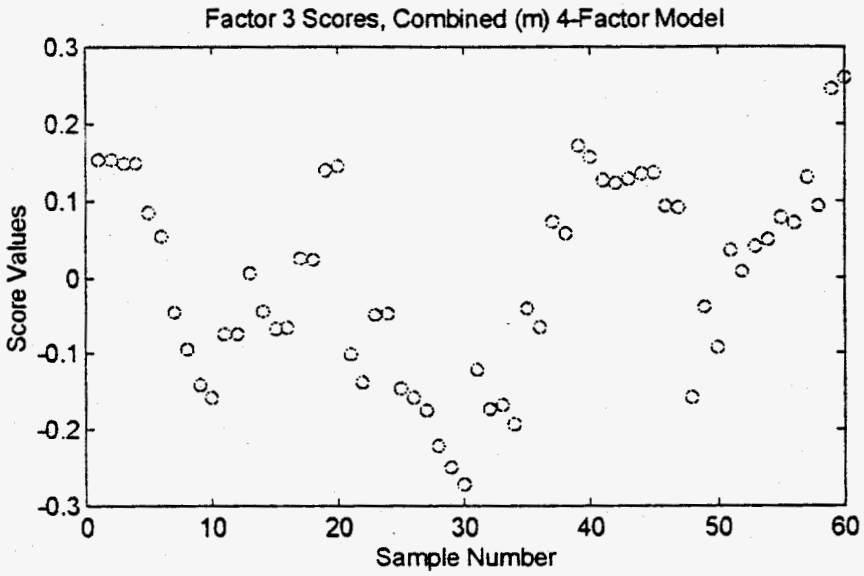


Figure 2-18. Sample scores on the third spectral block factor of the combined particle size NIR model.

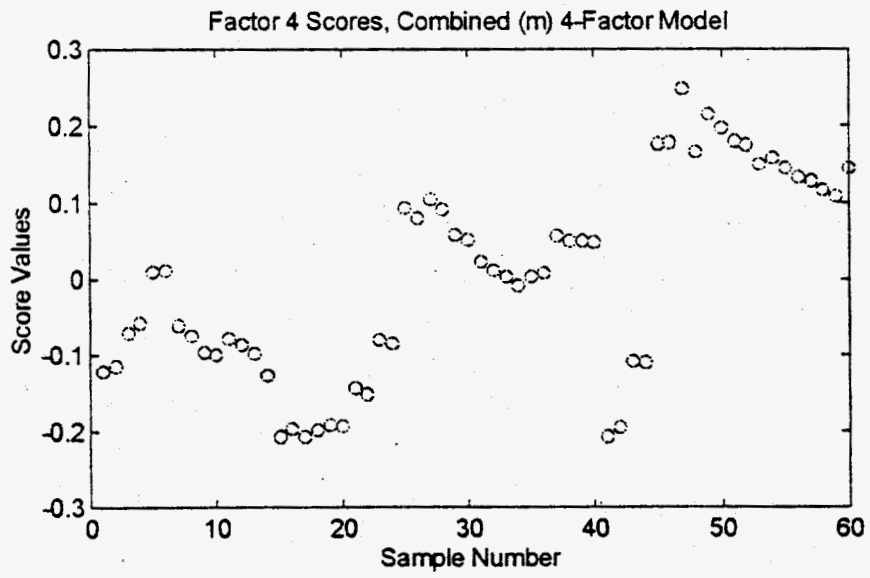


Figure 2-19. Sample scores on the fourth spectral block factor of the combined particle size NIR model.

The plot of the first model factor scores in the spectral block versus the concentration block is shown in Figure 2-20 below. The plotted numbers correspond to the sample number in the data set. The solid line is the least squares "inner relation" between the spectral response and the water concentration blocks. In this plot, one can clearly see that the first factor describes most of the linear relationship between the spectral responses (X-block) and the water concentration (Y-block). We can also see the nonlinear component described by the subsequent factors as scatter, or residuals, off the linear regression line between the two blocks of data. Notice that the residuals due to the nonlinear response are largest for the intermediate water concentration samples, which is what we would expect based on our interpretation of the third model factor described above. Also note that the residuals are above the line for the small particle size samples (samples 5-12) and below the line for the medium and large particle size samples. The X-block versus Y-block plots for the remaining factors of the combined particle size model are less easily interpreted and are not shown in this report.

In evaluating the fit of the model described above to the data, the four factor PLS model had root mean standard error (RMSE) of 0.76 wt% water for the combined particle size data set. By using the combined particle size model described above and the individual particle size spectra, the wt% moisture prediction performance of the combined model can be estimated for each of the three different particle sizes. The results of this analysis gave RMSE errors of 1.03, 0.65, and 0.52 wt% moisture for the small (less than 420 μm, #40 sieve), medium (420 μm to 841 μm, #20 sieve), and large (841 μm to 2 mm, #10 sieve) particle sizes, respectively. The prediction results are shown in Figure 2-21 below. The solid line in the plot represents the ideal 45° linear fit line and the broken lines are the actual least squares fit lines for the various particle sizes. The largest errors are associated with the small particle size samples and the errors for that group of samples are largest for the mid range moisture concentrations. The wt% moisture predictions from the medium and large particle

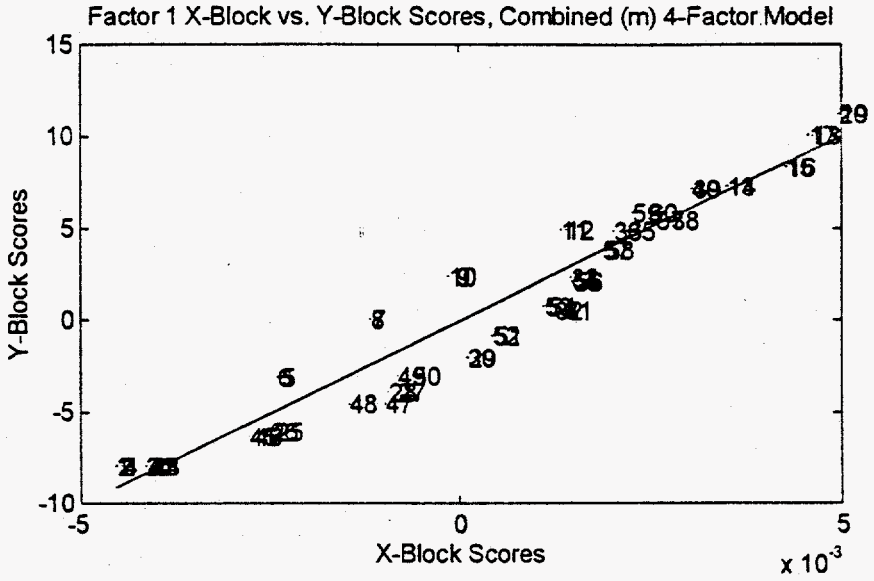


Figure 2-20. Sample scores in the spectral response plotted versus the scores in the water concentration for the first factor of the combined particle size NIR model.

size spectral data had about the same error associated with them, although the large particle size predictions were slightly better. These results indicate that the combined particle size model is biased towards the larger particle sizes and less representative of the small particle size sample's moisture response. Again, this is probably due to the use of additional model factors to account for the nonlinear moisture response for the larger particle sizes and resulting overfitting of the small particle size data as described above.

As was done with the visible data, separate PLS models were developed for each particle size relating the NIR spectra to the wt% water concentration. Each individual particle size model was then used to predict the wt% moisture content from the spectra of each of the other particle sizes. The individual particle size models were also used to predict moisture from the combined particle size data set. This cross prediction analysis provides a basis for estimating the effect of variations

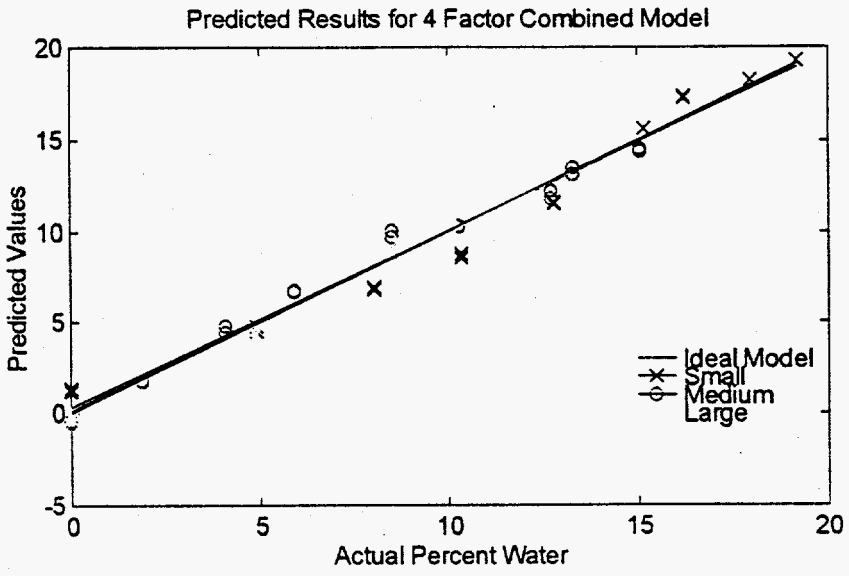


Figure 2-21. Prediction results for the three particle size data sets using the four factor PLS model derived from the combined NIR data set.

in particle size on the actual moisture prediction accuracy. The results of these studies are presented in Table 2-4 below.

# of Factors	Calibration Model Set	Prediction Set			
		Small	Medium	Large	Combined
2	Small	***0.61***	1.95	2.00	1.65
2	Medium	2.38	***0.47***	1.27	1.58
3	Large	4.85	1.86	***0.24***	3.00
4	Combined	1.03	0.65	0.52	***0.76***

Table 2-4. Root Mean Square Error (RMSE) for the prediction of wt% moisture from the NIR spectra using PLS.

In Table 2-4, the diagonal elements (shown in boldface italics) are the RMSE errors obtained for the individual models (predictions from the model made on the same

particle size data) and the off diagonal elements are the RMSE results obtained by using the data set identified by the row label to build the calibration model and the NIR spectra set corresponding to the column labels for the prediction data. [The second row and column labels correspond to the following particle size ranges : Small = < 420 μm; Medium = 420 - 841 μm; Large = 0.84 - 2 mm; Combined = <420 μm to 2 mm].Again, restricting the model development and prediction to a single particle size results in lower errors relative to the combined particle size model. However, the errors obtained when using a single particle size model to predict moisture from the spectra of a different particle size are typically worse than those obtained from the combined particle size model. Generally, the small particle size data had the largest errors when predicted with the other particle size models, probably due to the higher degree of nonlinear behavior observed in the large and medium particle size responses. The first column of Table 2-4 contains the number of factors for each model used in the predictions. This number of factors was determined using cross validation analysis.

The one main reason for the poor prediction across particle sizes stems from the incorporation of variance (information) which is specific to a particular particle size into the individual particle size models. In other words, the PLS models for a particular particle size are too specific to be applied to other size data which may not have the same subtle spectral features. In statistical terms, the individual particle size models overfit the data when applied to other particle size spectra. This is especially true for the medium and large particle size models. In these cases, a single factor model does a better job of predicting the other size data than the model with the number of factor suggested by cross validation. However, the single factor model has larger residual errors for the calibration set used to build the model.

For example, compare the cross validation model results given in Table 2-4 with the results shown in Table 2-5 where the RMSE prediction values are given for different

numbers of calibration model factors for the medium particle size model. In the table, the highlighted value is for the model suggested by cross validation. It is apparent that adding factors to the model will always improve the fit of the model to the calibration data. This is shown in column 3 of Table 2-5 where the model fit error is listed for each additional factor of the medium particle size calibration model. From columns 2, 4, and 5 of Table 2-5, one can also see that adding additional factors to the medium particle size model leads to increased prediction errors when the medium model is used to predict wt% moisture from the spectra of the other particle sizes. Even the one factor medium model did rather poorly at predicting the moisture from the small particle size spectra.

# of Factors	Prediction Set			
	Small	Medium	Large	Combined
1	2.08	1.14	1.01	1.49
2	2.38	***0.47***	1.27	1.58
3	2.47	0.40	1.19	1.61
4	3.09	0.36	1.35	2.03

Table 2-5. Root Mean Square Error (RMSE) for the prediction of wt% moisture from the medium particle size PLS model with different number of factors.

A similar analysis was applied to the large particle size model with the results shown in Table 2-6. Again, by restricting the number of factors in the model to one factor, rather than the cross validation suggested three factors, the cross prediction errors for the small and combined particle size data sets are reduced by a half.

The prediction results from the individual particle size models are shown in Figure 2-22, Figure 2-23, and Figure 2-24 for the small, medium, and large particle size models, respectively. For the small particle size model results, shown in Figure 2-22,

# of Factors	Prediction Set			
	Small	Medium	Large	Combined
1	2.21	1.17	1.00	1.55
2	4.37	1.42	0.40	2.66
3	4.85	1.87	***0.24***	3.00
4	4.87	1.95	0.23	3.03

Table 2-6. Root Mean Square Error (RMSE) for the prediction of wt% moisture from the large particle size PLS model with different number of factors.

one can clearly see that the two factor model does not do an adequate job of describing the nonlinear behavior associated with the medium and large particle size data sets. The small particle size model does a good job of estimating the moisture for the small particle size data over the full range of wt% water concentration. However, the information needed to model the nonlinear spectral response to increasing moisture observed in the medium and large data sets is simply not present in the small particle size data set. This results in the large error, or deviation off the regression line, observed in the predicted results for the intermediate moisture content samples from the medium and large data sets.

The results of predicting all three particle size data sets using the medium particle size model are shown in Figure 2-23. Here we can see that the two factor medium model does a good job of estimating the wt% moisture concentration over the calibration range for the medium particle size spectral data. However, the medium particle size model underestimates the wt% moisture for the small particle size data and overestimates the wt% water for the large particle size data. Notice that the prediction errors for the other particle size data can almost be described by a simple offset term. The slope of the least squares fit line between the actual and estimated values is nearly identical for each of the three particle size data sets. There is some

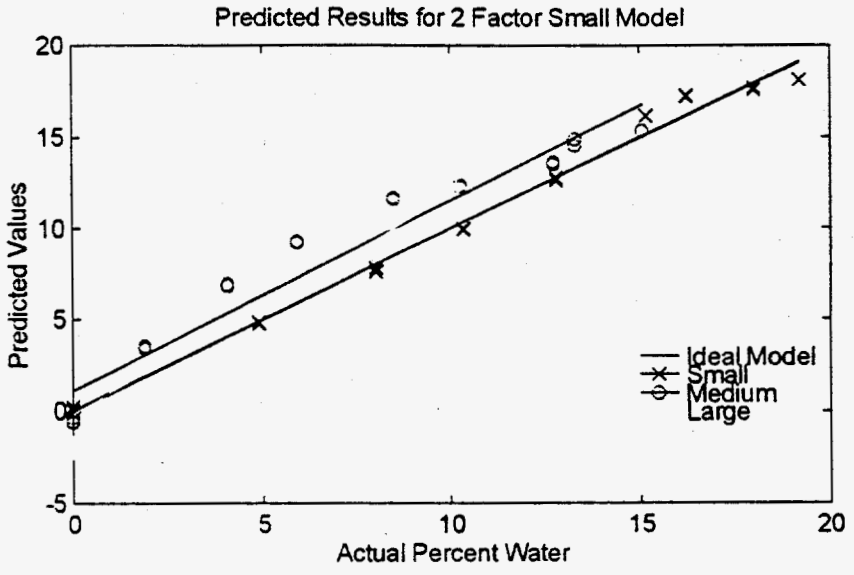


Figure 2-22. Prediction results for the three particle size data sets on the two factor PLS model derived from the small particle size NIR calibration data.

evidence of additional induced error in the small particle size results which is probably due to the nonlinear terms of the medium particle size model. This shows up as a slightly nonlinear deviation from the least squares fit line for the lower water content samples from the small particle size data set.

The results from the prediction of all three particle size data sets using the large particle size PLS model are shown in Figure 2-24. Here we can see that although the large particle size model, using three PLS factors, gives good results for the large particle size data, it underestimates the wt% moisture for the small and medium particle size data. The prediction errors have both a constant offset and a slope bias component for each of the other two particle size data sets, relative to the calibration set model. The errors are most severe for the small particle size data (especially at the higher water concentrations) and somewhat less severe for the medium particle size data. Despite the large magnitude of the prediction errors for the small and

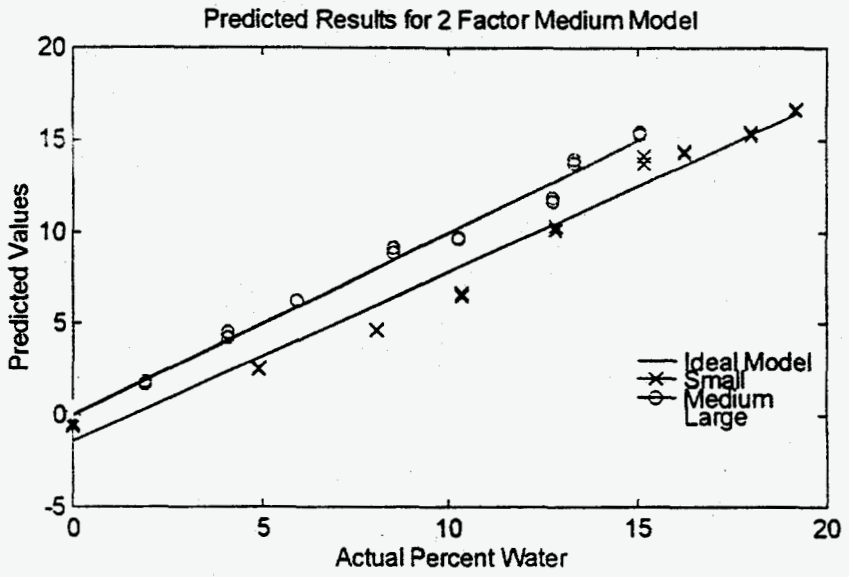


Figure 2-23. Prediction results for the three particle size data sets on the two factor PLS model derived from the medium particle size NIR calibration data.

medium particle size data, the predicted wt% moisture values do follow a fairly linear relationship over the range of water concentrations.

Conclusions

The results of this study showed that in the visible range, the standard error of prediction of the samples when the different particle sizes were analyzed separately was lower than the error when all three particle sizes were analyzed simultaneously. The error when analyzing all sixty samples simultaneously was on the order of 0.8 wt% when applying PLS with three factors to the second derivative visible data, whereas analyzing them separately dropped the error to roughly half a percent or better (0.59, 0.34, and 0.23 wt% for the small, medium, and large particle sizes, respectively). In the NIR, the error for the sixty sample combined particle size model was 0.79 wt% while the individual particle size models had errors of about half a

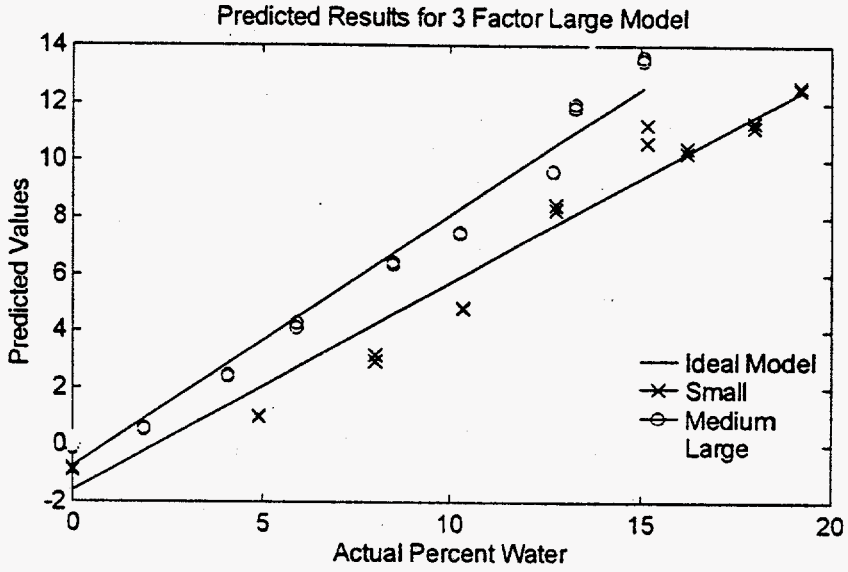


Figure 2-24. Prediction results for the three particle size data sets on the three factor PLS model derived from the large particle size NIR calibration data.

percent of less [0.61, 0.47, and 0.24 wt% for the small, medium, and large particle sizes, respectively].

However, the predictive ability of the individual particle size models, when applied to the spectra of a different particle size, showed increased errors relative to the combined models for both the visible and NIR data sets. In the visible region, the small particle size model gave errors that were almost an order of magnitude larger than were obtained for the calibration estimate (0.59 wt%) when predicting the moisture from the medium and large particle size data (5.17 and 4.98 wt%, respectively) . The errors in predicting the moisture from the small particle size visible spectra were also quite large when either the medium or large particle size models were used for the prediction (2.23 and 3.78 wt%, respectively). The errors between the medium and large particle size data, predicted with the other size model,

were roughly four times greater (about 1 wt%) than the individual size calibration model results.

The same pattern held true for the NIR spectral region, although the magnitude of the cross prediction errors were not as large. In the NIR the small particle model gave moisture prediction errors of about 2 wt% when applied to the medium or large particle size spectra (roughly three and a half times larger than the small particle size data calibration error estimate of 0.61 wt%). The errors were even larger when predicting moisture from the small particle size NIR spectra using either the medium or large particle size models (2.38 and 4.85 wt% respectively). Again, the errors between the medium and large particle size data, predicted with the other size model, were roughly four times greater (about 1.5 wt%) than the individual particle size calibration model results.

For both spectral regions the cause of the poor performance of the individual particle size models when predicting moisture from different particle size spectra is well understood. By limiting the calibration set to a single particle size, better calibration models for that particle size can be developed. However, these individual particle size models include factors which account for subtle spectral responses to concentration variation that are somewhat particle size dependent. The nonlinear response to higher water concentrations observed in both spectral regions is an example of this somewhat particle size dependent spectral feature. The spectral responses from a different particle size may not have the same subtle effects, so when the model tries to fit them, it introduces errors in the estimated concentrations. This is related to the overfitting and extrapolation problems commonly discussed in the regression and statistics literature. Except, in this case we are not overfitting within the calibration data set, but rather, over the entire range of samples we are trying to predict. Stated another way, the individual particle size models are valid over the limited calibration

space of a single particle size, but are not valid over the extended space spanned by the full range of particle sizes we were trying to predict.

Fortunately, the particle size effect is largely removed by including enough variation in the calibration set. When the three particle size data sets are combined, the calibration space spans the expected range of particle sizes and the model can treat the size dependent spectral effects like known interferents and correct for them. Of course, the correction for the size "interferent" is not perfect and it does introduce some additional error into the moisture estimates. Perhaps more correctly, by spanning a wider range of particle sizes, the calibration model is no longer as free to fully account for all the subtle features as a model restricted to a smaller calibration space. As a result, some of the subtle spectral features get regulated to the "noise" of the full calibration space and are not included in the model. For individual samples this may lead to some small decrease in the model's predictive ability, but over the full calibration set the overall prediction efficiency is improved.

The implication for the optical moisture monitoring project is that it will be important for the moisture calibration model to be made with samples covering the expected range of particle sizes. While this may not be possible due to the large range of sample consistencies within the tanks, some effort should be made to insure that the calibration model does not focus too closely on effects that are specific to one particle size or sample consistency. One approach would be to desensitize the model by restricting it to a single factor. We have seen evidence that a single factor model accounts for the major spectral response related to changes in moisture. By limiting the calibration model to the single factor, most of the subtle spectral corrections are eliminated and the model becomes more general. The price you pay is increased error in the predicted moisture values. The results obtained so far indicate that the worst case errors would be on the order of three to four weight percent water. This is

certainly within the limits needed for safety verification and monitoring of the tanks and is probably as accurate as the current moisture determination method.

3. Composition Changes Study

In previous work, CPAC established that optical spectroscopic measurements could be used to accurately predict the moisture content of simulated waste from the Hanford waste tanks. That previous study considered only the single waste composition represented by the BY-104 salt cake simulant. The main purpose of this study was to evaluate the effect of variations in the chemical composition of the waste simulant on the optical moisture determination method developed at CPAC. In this study, the BY-104 salt cake simulant obtained from WHC was again used as the base chemical composition. Experiments were performed where the BY-104 chemical composition was changed by adding amounts of the simulant's individual pure components to the base composition. For each change in the pure component level, the NIR spectral response was measured at different moisture levels. The results of changing the chemical composition were evaluated by observing the effect of the composition change on the moisture prediction from a calibration model derived from the base composition at different moisture levels. Further experiments were conducted to quantitate the composition changes effect on the moisture determination for those components which exhibited large composition sensitivities.

Experimental

All spectra described in this report were collected using the same spectrometer used in the earlier studies. However, the spectrometer has been modified to use a direct insertion fiber optic reflectance probe. This new probe replaces the standard sample cup and cup holder apparatus for the measurement of the spectra. In the new design, reflectance spectra are recorded by simply placing the fiber optic probe into contact with the sample. The fiber optic probe consists of a fiber bundle to deliver the

monochromatic illumination light from the spectrometer, a sampling window, and a collection fiber bundle to transmit the reflected light back to the detector. To couple the light into the source fiber, the individual optical fibers are aligned along the exit slit of the spectrometer's monochrometer. The probe body itself consists of a stainless steel cylinder (approximately 2.5 cm in diameter and 20 cm long) with a sealed window slightly recessed into the body at one end. The fiber optic bundles are encased in a single shielded flexible sheath approximately 120 cm long. At the detector end, the signal fiber bundle is split from the sheath and is positioned such that the light from the bundle illuminates the detector. This fiber probe is shown schematically in Figure 3-1.

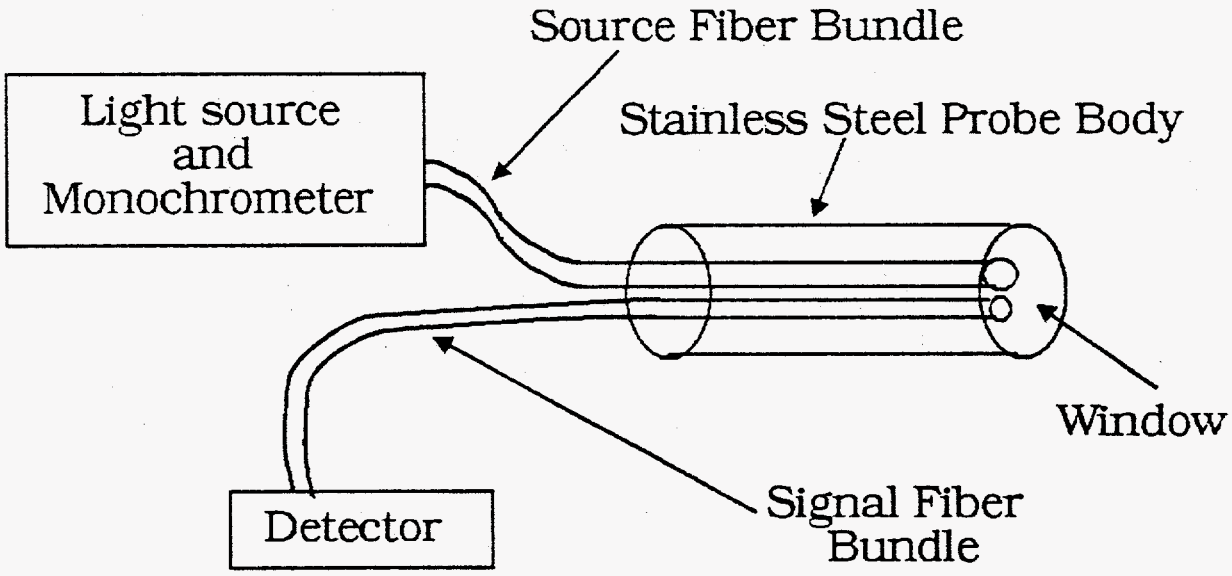


Figure 3-1. Schematic drawing of the fiber optic probe used in the reflectance measurements.

The use of this probe greatly simplifies the sample handling associated with collection of the sample spectra. Using the fiber optic probe allows for smaller sample sizes to be used since one no longer has to fill the sample cup. By getting away from the sample cup arrangement one can also go to higher moisture content samples without

the problems of moisture leakage from the cup experienced in the earlier studies. The use of the fiber optic probe also makes it possible to do *in situ* analysis which is more realistic for eventual hot cell, or in-tank, use. Unfortunately, the current configuration of the probe restricts the spectral range to either the NIR (1100 nm to 2500 nm) or the visible (400 nm to 1100 nm) spectral regions since the signal fiber can illuminate only one of the instruments detectors at a time. In this study, only the NIR spectral region results were collected.

The experimental procedure used in this study consisted of adding a pure component to the dried base BY-104 simulant. In this way, the weight percent (wt%) of each component chemical can be increased, or shimmed, relative to the original BY-104 composition. After adjusting the composition, the sample moisture level is adjusted by the addition of pure water to reach a target moisture level. For each new composition sample, spectra were collected at 0, 5, 10, 15, 20, and 25 wt% moisture levels. The steps of the experimental procedure used to prepare the samples are outlined below.

1. Weigh out a known amount of fully dried base BY-104 simulant.

2. Based on the weight of the base material, the reported composition of the base material, and the target component level; add a known weight of the pure component. Note: the reference BY-104 composition is based on a dry weight percent for each of the nine component chemicals. The amount of a component to add is based on the following formula:

$$wt.\ analyte = \frac{(wt.\ base \times target\ wt\%) - (wt.\ base \times base\ wt\%)}{(1 - target\ wt\%)} \quad (1)$$

where wt. analyte is the amount of pure component to add, wt. base is the amount of starting dry BY-104, target wt% is the desired weight percent of the pure component in the new sample, and base wt% is the weight percent of the component in the original, or starting, BY-104 simulant.

3. To the sample (BY-104 plus added component) add a weight of water needed to bring the sample to the desired moisture level (based on a total weight percent water). The amount of water at each moisture level is based on the water content of the sample before the addition plus the amount of water added at this step. This scheme allows for successive additions of water to the same sample to obtain progressively higher moisture wt% levels. The amount of water to add is calculated by the formula below.

$$wt.\ water = \frac{(wt.\ total \times target\ wt\%) - (wt.\ total \times prev.\ wt\%)}{(1 - target\ wt\%)} \quad (2)$$

where wt. water is the amount of water to add, wt. total is the weight of the starting sample (BY-104, any added component, and any previously added water), target wt% is the desired weight percent of moisture, and prev. wt% is the weight percent moisture prior to the water addition at this step.

4. The actual weight percent analyte and weight percent water were calculated from the recorded weights of material added in the two steps above. Note: the addition of water was limited to a

resolution of about 0.02 g using a standard Pasteur pipette. The pertinent calculations are as follows:

$$wt\%\ analyte = \frac{wt.\ analyte + (wt.\ base \times base\ wt\%)}{(wt.\ base + wt.\ analyte)} \quad (3)$$

$$wt\%\ water = \frac{wt.\ water + (wt.\ total \times prev.\ wt\%)}{(wt.\ total + wt.\ water)} \quad (4)$$

where all the terms are as described above.

5. Thoroughly mix the sample. Since the nature of the BY-104 simulant and most of the pure components are hygroscopic, the time (approximately 30 - 45 seconds) needed to mix the sample is probably sufficient for the sample to become fully hydrated. No provisions were made to exclude atmospheric moisture adsorption, or to prevent the loss of water through evaporation, during the weighing, mixing, or spectral collection steps. These processes are almost certainly occurring over the course of the experiment. By emphasizing consistency in the experimental procedure, it is hoped that these effects would not adversely affect the results. As we will show later, the adsorption of atmospheric moisture was an important effect when dealing with the more hygroscopic pure components.

6. Measure the sample spectra. The spectra were measured by lightly inserting the fiber optic probe into the sample material. This is a simple contact measurement, no effort was made to quantify or control the depth of insertion or the insertion pressure. Two spectra were measured for each sample with the

probe removed and reinserted between scans. Each spectral scan was collected at 2 nm resolution over the 1100 nm to 2500 nm region (NIR) and 32 scans were co-added for each recorded spectrum. Note that some of the sample (simulant and water) was lost after the spectra collection due to adhesion of the sample to the probe body. Efforts were made to minimize this loss by attempting to scrape any material stuck to the probe back into the sample container. The sample was weighed after the spectral data collection in order to estimate the sample loss due to the probe insertion.

The use of the above procedure allows for the preparation of a sample which is very close to a desired experimental design point, in terms of weight percent analyte and moisture. In designing this experiment, there were several goals and constraints that needed to be considered. We wanted to be sure that the concentration changes were large enough to have an observable effect. We also wanted to have the same experimental procedure used for all the samples. This meant that the experimental design was largely driven by concerns about sample homogeneity for the low moisture samples. Sample homogeneity for the higher moisture samples was less of an issue since sufficient water was present to dissolve and mix the pure component additions. The limited amount of the BY-104 simulant, combined with the large number of samples to be prepared, limited the starting sample size to approximately 5 grams of pure BY-104. This in turn, limited the resolution possible for adjustments to the minor components of the simulant. Since many of the minor components of the BY-104 simulant are hygroscopic crystalline solids, it was difficult to obtain small particle size samples of the pure components for the additions. This dictated that the increase in the minor components needed to be rather large (on the order of 0.5 grams) in order to achieve a uniform distribution of the added component in the sample. Table 3-1 shows the design points for the analyte addition based on a five

gram sample of dry BY-104 simulant. The base wt% values are the individual component concentrations in the BY-104 simulant. The initial wt. values are amounts of each component present in a five gram sample of BY-104. The new wt% is the component concentration obtained after adding the weight of component listed in column five to an approximately five gram sample of dry BY-104.

Component	Base wt%	Initial wt.	New wt%	Added wt.
$NaNO_3$	86.1	4.305	90.01	1.9864
$NaAlO_2$	7.9	0.395	16.02	0.4837
NaOH	1.8	0.090	10.01	0.4560
Na_2SiO_3	1.6	0.080	10.16	0.4777
$Fe(NO_3)_3$	1.1	0.055	10.07	0.4992
Na_3PO_4	0.7	0.035	11.11	0.5860
$Ca(NO_3)_2$	0.4	0.020	10.08	0.5395
$Mg(NO_3)_2$	0.2	0.010	10.32	0.5659
$Mn(NO_3)_2$	0.2	0.010	10.16	0.5561

Table 3-1. Initial and final component concentrations for the composition experiment design points.

For each of the nine samples, corresponding to the addition of a single pure component, six samples at different moisture levels (0, 5, 10, 15, 20, and 25 wt%) were prepared by successively adding water according to the procedure outlined above. For each of these samples, two replicate spectra were measured. Table 3-2 shows the resulting moisture level data for the full experiment. The resulting composition data set consisted of 108 spectra (nine compositions, six moisture levels per composition, and two replicate spectra per moisture level). The zero wt% moisture design point (corresponding to the composition change but no water addition) is not listed in the table and is assumed to be zero for all samples.

In addition to the composition variation data set described above, spectral data was collected at the same six moisture levels for the base BY-104 simulant. This data was obtained following the same procedure outlined above, but the chemical composition of the BY-104 simulant was not changed. This experiment was replicated three times. This data, which we will call the Control data, was used to generate the pure BY-104 moisture calibration model. It was decided to generate a new moisture calibration model rather than use the Phase 1 model for two reasons. First, we wanted to avoid any errors which might be introduced through the use of the fiber optic probe instead of the sample cups for data collection. Secondly, we wanted to generate the Control calibration model using data obtained with the same experimental procedure used for the composition variation experiment. This should allow us to evaluate the effect of the composition changes free from any extraneous sources of errors due to the calibration model.

Pure Component	Resulting Moisture Level (weight percent water)				
	5 wt%	10 wt%	15 wt%	20 wt%	25 %
$NaNO_3$	5.15	10.03	14.87	20.71	25.04
$NaAlO_2$	5.10	10.12	15.04	20.34	24.92
NaOH	5.24	10.03	14.97	19.98	25.06
Na_2SiO_3	4.96	10.64	14.97	20.01	25.13
$Fe(NO_3)_3$	5.32	10.00	14.97	20.19	25.00
Na_3PO_4	4.99	9.93	15.23	20.12	24.94
$Ca(NO_3)_2$	5.20	10.64	15.09	19.77	25.00
$Mg(NO_3)_2$	5.18	9.90	15.03	19.88	24.77
$Mn(NO_3)_2$	5.04	10.01	15.23	20.84	25.01

Table 3-2. Moisture level experimental design points for each of the new composition samples.

A third data set, which we will call NO_LOSS, was generated using the pure BY-104 simulant at the same six moisture levels. However, for this data the moisture levels

were not obtained by the successive water addition procedure outlined above. Instead, the different moisture levels were achieved by direct addition of the required water to six separate samples of the dry base BY-104 simulant. The purpose of this data set is to evaluate any bias introduced by the successive water addition procedure and the effect of sample loss due to adhesion to the spectrometer probe.

In the experimental procedure the sample loss is assumed to be homogeneous (i.e., a proportional loss of both the BY-104 material and moisture). For example, a loss of 0.05 grams before and after the spectral measurement at the 20 wt% moisture level is assumed to be a loss of 0.01 grams of water and 0.04 grams of dry sample. While at lower moisture levels (where the water is fully adsorbed) this assumption seems reasonable, at the higher moisture levels one might expect a proportionately higher loss of water than sample. The net result could be a bias in the calibration model for the higher water samples. As the results presented below show, this bias does exist in the experimental data.

In addition to the experimental data, the pure chemical components of the BY-104 simulant were measured spectroscopically. The spectra of the pure components were measured without drying and should include some adsorbed and hydrated water. These pure component spectra, and their second derivative spectra, are shown in Appendix A for the NIR and VIS spectral regions.

Results and Discussion

First we will consider the PLS calibration model for the Control data. This data set replicates the experimental procedure but does not adjust the composition of the BY-104 simulant. The results of this model will provide the benchmark by which the composition variation experiments can be evaluated. In evaluating the Control model, we will also be interested in model changes related to the use of the fiber optic

probe. Next we compare the Control model to the NO_LOSS data to evaluate the experimental procedure. Finally, we will present the results from the composition variation experiment.

Control Model

For the Control model calibration data set, we combined the data from the three replicated experiments mentioned above. The resulting data set contained 36 spectra (three experiments with six moisture levels per experiment and two replicate spectra at each moisture level.) Cross validation studies indicated that a two factor model provides the best predictive model for this data set without overfitting. The amount of variance explained by the PLS model factors are given in Table 3-3 below. The first factor of the model accounts for almost all of the moisture level (Y-block) information using about 96 percent of the spectral response (X-block) information. Recall that the Phase 1 NIR model consisted of a single factor. However, the second factor of the Control model only describes a minor amount of the spectral response which correlates to a very minor percentage of the moisture level variance.

	X-Block (Spectra)		Y-Block (Moisture)	
Factor #	This Factor	Total	This Factor	Total
1	96.55	96.55	99.01	99.01
2	2.63	99.18	0.29	99.31

Table 3-3. Percent variance described by each factor of the PLS model for the combined Control experiment data.

At this point a brief review of the cross validation procedure may be in order. Cross validation seeks to estimate the maximum complexity (i.e., the number of model factors) for a model derived from a given calibration data set. The criteria for this estimate is the prediction error of the model. As more factors are added, the

resulting model will always fit the calibration data better and have lower fit errors. In the extreme, a full factor model will use 100% of the spectral variance to describe 100% of the moisture concentration information. However, if the calibration data contains noise, or other sources of variance which are uncorrelated between the spectral and concentration blocks, including that variance in the model is undesirable and is referred to in the statistics literature as overfitting. Since future sample are likely to have slightly different "noise" structure, the resulting predictions from an overfitted model will likely have inflated prediction errors. Cross validation attempts to guard against overfitting by developing many different models with increasing complexity from a portion of the calibration set. The number of model factors corresponding to the lowest prediction errors when using another portion of the calibration set (not used in building the model) is selected for the final model. However, cross validation is still an empirical method which is based on the calibration data set itself and is a substitute for having a fully independent validation data set for testing the models. In other words, cross validation tries to develop models that utilize the relevant predictive information in the spectra which is consistent within the calibration data set. Therefore, the cross validation results will depend to a large extent on the quality of the experimental design and the quality of the calibration data set itself.

Given the discussion above, it is not alarming that cross validation indicated the Control model should use two factors for the moisture prediction while a single factor model was indicated for the Phase 1 study. Especially since the second factor of the Control model described only about 0.30% of the moisture concentration information. In the Control experiment the moisture levels were much more evenly spaced than in the Phase 1 study and covered a slightly wider range (up to 25 wt% moisture) of moisture concentrations. Also, the experimental procedures and spectra collection methods were different between the two studies. The minor amount of additional information described by the second factor of the Control model can easily be

attributed to a better experimental design and data collection setup. In cross validation, as in most everything else, the number of model factors issue reduces to a signal-to-noise consideration. In the Control experiment there were actually three replicate experiments. In this case, we actually had independent validation data sets which could be used to test the calibration models. When one replicate was used to build the models and the other two replicates were used as validation sets, the prediction results indicated that the best model was an average between the one and two factor models. This agrees very well with the cross validation results for the combined Control model.

The model fit results for the combined Control model are shown in Figure 3-2. From this plot, one can see that the model describes a linear fit over the full moisture calibration range with no noticeable bias or offset. Some summary statistics for the model fit are given in Table 3-4 below. Also listed in Table 3-4 are the fit statistics for the separate PLS models built from each of the three replicate Control experiments. Note, the RMSE from the Control model is slightly larger for the combined data than we observed for the NIR calibration model from the Phase 1 study (0.734 wt% versus 0.42 wt%). However, for each of the three replicate Control experiments, the individual two factor models gave RMSE values which agreed well with the Phase 1 results. Much of the increased error for the combined Control model seems to be due to variability between the three replicate experiments and not the use of the fiber optic probe.

When the Control model was used to predict the moisture from the Phase 1 calibration set it had an RMSE of 1.667. Conversely, when the Phase 1 model was used to predict moisture using the Control data the RMSE was 1.620. In both cases, the main contribution to the errors was due to an increasing bias at higher water concentrations. Recall that in the Phase 1 study the higher moisture content data was somewhat suspect due to water leakage from the sample cup. The bias observed

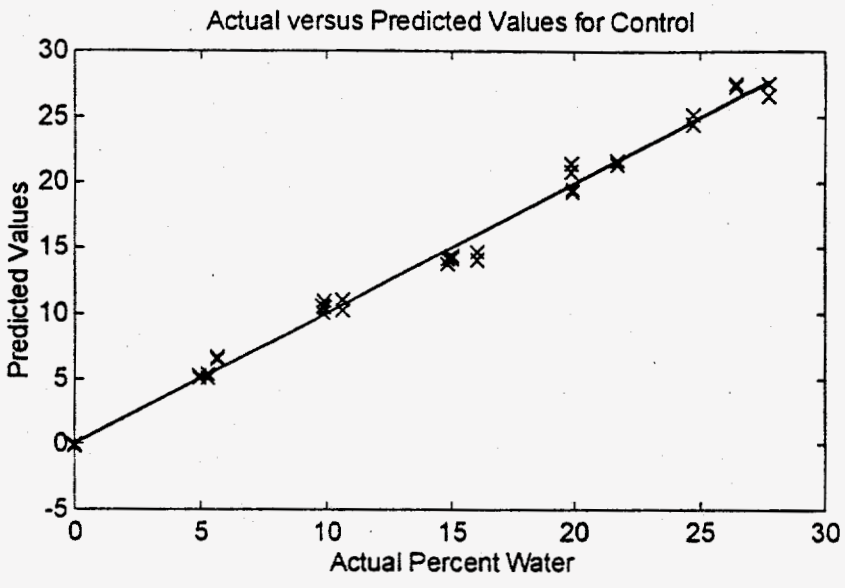


Figure 3-2. Actual versus predicted fit results for the two factor PLS model derived from the Control data set.

in the moisture predictions from the Phase 1 data using the Control model is consistent with systematic errors in the high water reference values. The systematic errors are likely due to differences in the experimental procedure between this data and the Phase 1 data. Given the nature of the cross prediction errors, it is unlikely that they are due to the use of the fiber optic probe itself.

The PLS model scores for the first factor of this model are shown in Figure 3-3 below. In this plot we can see that this factor is describing the experimental design related to the six moisture levels. Recall that this factor accounted for over 99% of the moisture information in the data. Each cluster of two points corresponds to the replicate spectra collected at each moisture level. The three trends of twelve samples correspond to the three replicate Control experiments. The corresponding loading plot for the first factor is shown in Figure 3-4 below. The loading vector looks very much like the first factor from the Phase 1 model with the familiar intensities in the OH first overtone region (1450 nm) and the OH combination band regions (1950 and

	Combined	Exp. 1	Exp. 2	Exp. 3
Slope	0.993	0.994	0.996	0.995
Y-Intercept	0.090	0.072	0.053	0.033
R-squared	0.993	0.994	0.996	0.995
Std. Error Est.	0.743	0.661	0.589	0.743
Pearson's Corr.	0.997	0.997	0.998	0.999
RMSE	0.734	0.661	0.589	0.437

Table 3-4. PLS model fit statistics for the Control experiments. Combined is for the Control model, Exp.1, 2, and 3 are for the individual replicate experiment models.

2250 nm) which we attribute to sample moisture level.

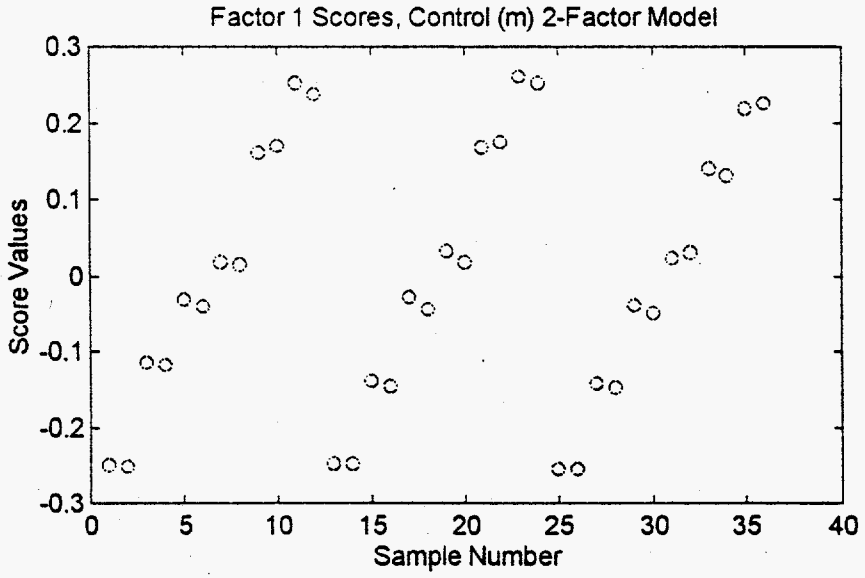


Figure 3-3. Spectral scores on the first factor of the Control PLS model.

The second factor scores and loadings are shown in Figure 3-5 and Figure 3-6, respectively. This factor mainly provides a large correction to the 25 wt% moisture samples from the first replicate experiment (samples 11 and 12). Recall that this

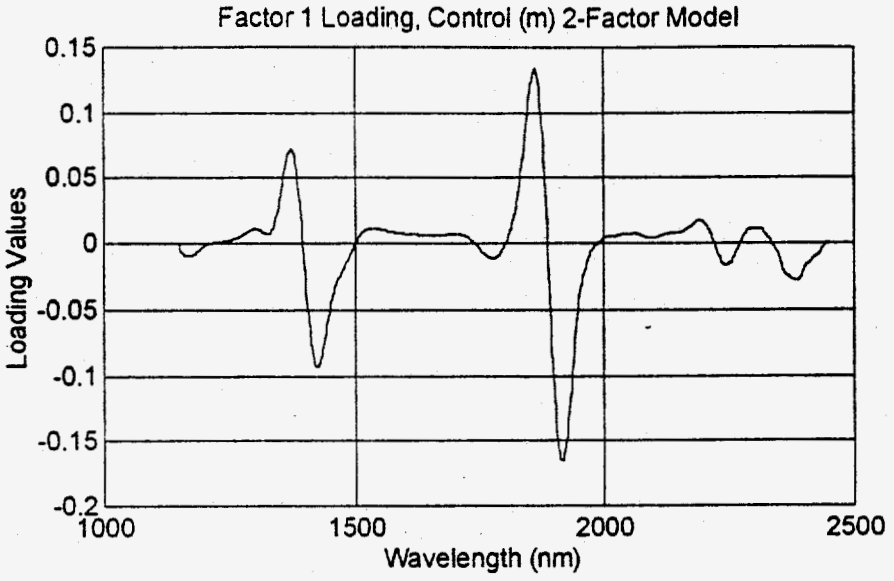


Figure 3-4. Spectral loadings on the first factor of the Control model.

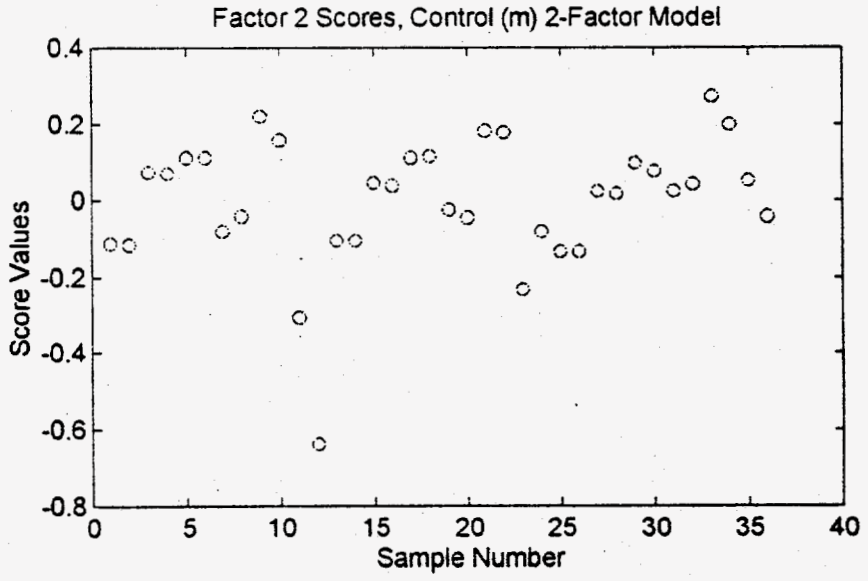


Figure 3-5. Spectral scores on the second factor of the Control model.

factor is only describing a small amount of the full spectral variation which accounts for very little of the moisture predictive ability of the model. Therefore, the dominance of the two samples (11 and 12) in this factor indicate that their spectra

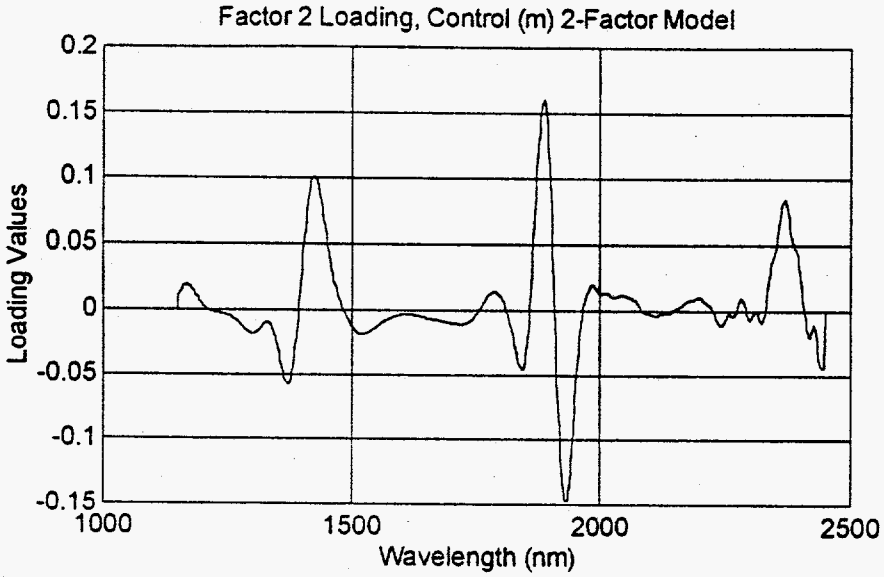


Figure 3-6. Spectral loadings on the second factor of the Control model.

are slightly different than the rest of the samples. This factor is also describing a small correction to the PLS model to help account for minor variation associated with the middle water concentration samples. Although the pattern is somewhat obscured by the high (negative) scores for 25 wt% moisture samples from the first experiment, the scores increase for the first three moisture level samples (0, 5, and 15 wt%), drop sharply for the 15 wt% samples, and then decrease for the 20 and 25 wt% samples. The loading vector for this second factor looks very much like the first factor loading (inverted) except for a slight shift towards smaller wavelengths of the peak at 1900 nm and increased contribution of the peak at 2425 nm.

The regression coefficient vector for the Control model is shown in Figure 3-7 below. Comparing the regression coefficient vector with the factor one loadings, one can see that the second factor provides only a slight modification. Consequently, the regression vectors from the Control model and the Phase 1 are nearly identical.

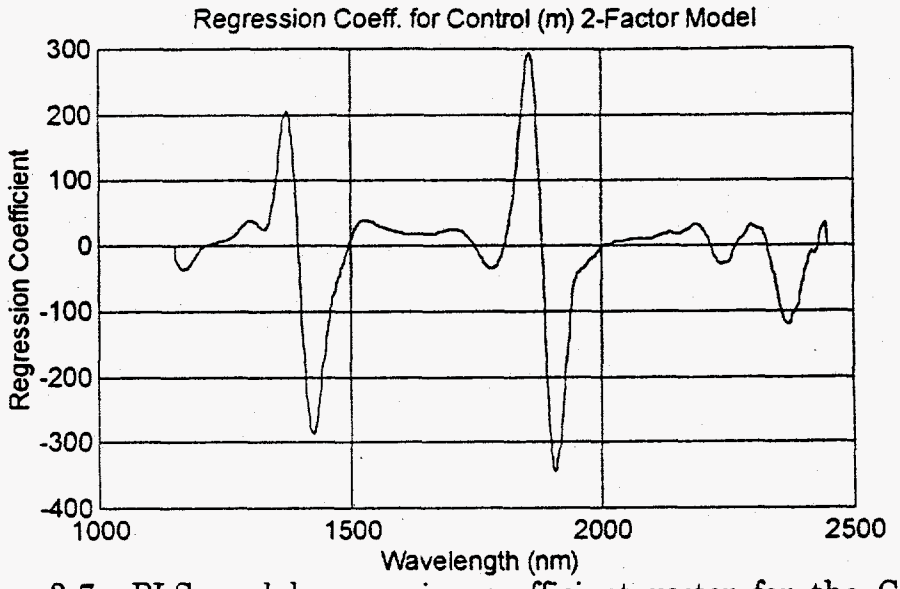


Figure 3-7. PLS model regression coefficient vector for the Control model.

Recall that an additional experiment was performed to attempt to isolate and quantify possible bias in the reference moisture weight percent values due to sample loss related to the insertion and removal of the fiber probe. The data from this experiment, which we will call NO_LOSS, was obtained by adding water to the dry BY-104 simulant for each moisture level measurement rather than successive additions of water to the same sample. Since a completely new sample was generated for each measurement, there should be no variation due to the loss of sample adhered to the probe body. When the wt% moisture from this spectral data was predicted using the Control data model, the results shown in Figure 3-8 were obtained. Here one can clearly see that the Control model accurately predicts the lower moisture (less than 15 wt%) content samples but shows increasing positive bias for the higher moisture samples.

These results are consistent with our earlier discussion about the effects of sample loss due to the insertion of the fiber probe into the sample. While we attempted to correct for the sample loss by weighing the sample before and after the spectroscopic

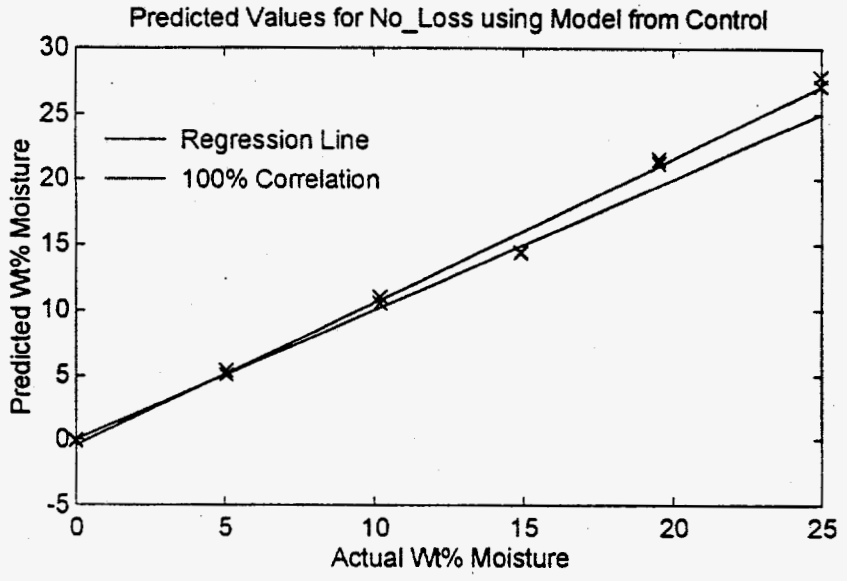


Figure 3-8. Prediction results from the NO_LOSS data using the Control model. Two lines are shown, the ideal 100% correlation line (dotted) and the actual regression line.

measurements, our correction was based on an equal proportion of the water and sample being lost. As expected, this experiment suggests that more water than sample was lost when dealing with samples above 15 wt% moisture. Therefore, our Control model was based on data where the reference moisture value was artificially lower than it should be. In effect, we have introduced a bias into our model by training the model to see a 23 wt% moisture spectra as having 25 wt% water. When the model is presented with spectra from a sample which is really 25 wt% moisture, it must predict a higher value. This illustrates an inherent danger, or limitation, of these indirect calibration methods; the model is only as good as the reference values you use to develop the model. To verify these conclusions, we developed a PLS model based on the NO_LOSS data and used it to predict the moisture from the Control spectra. The results are shown in Figure 3-9 below. As expected, the predicted high moisture values were consistently lower than the reference value and the model bias was larger at the higher water concentrations.

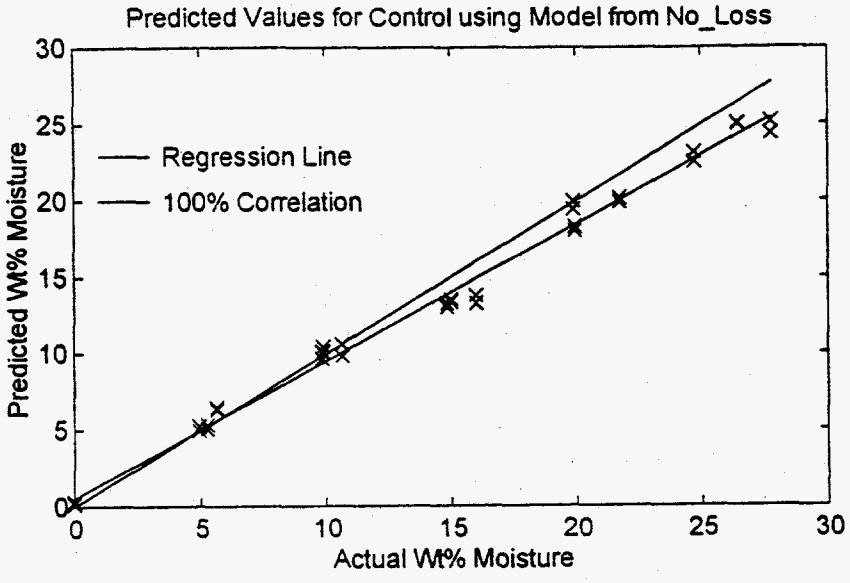


Figure 3-9. Prediction results from the Control data using the NO_LOSS model. Two lines are shown, the ideal 100% correlation line (dotted) and the actual regression line.

The conclusions from this experiment are that a more sophisticated correction for the sample loss needs to be implemented if successive water additions are to be used in the experimental procedure outlined above. Alternatively, the problem can be avoided if new samples are produced for each spectral measurement, although that can be quite costly in terms of simulant use and waste generation. In terms of evaluating the current data, the Control model can be used for all the composition variation data since the same procedure was followed in both cases. We will assume that since the same care was used in removing sample from the probe for the Control and composition variation experiments, the same relative partitioning of sample and water loss will have occurred in both sets of data.

Composition Change Effects

Now we will begin looking at the composition change data. Recall that in this experiment, we varied the concentration of each pure component of the BY-104 simulant and measured the NIR spectra at the same six moisture levels discussed above. The main goal was to determine the effect changes in composition would have on the moisture determination. In order to evaluate this effect, the spectral data corresponding to changes in each pure component concentration was used with the Control model to predicted the moisture levels. In the absence of any composition effects the predicted moisture values should fall along the regression line of the Control model. Any errors in the moisture prediction introduced by the composition changes will show up as deviations, or scatter, off the Control model regression line.

As we will see, the effects of varying the sample composition will usually not be constant over the range of moisture levels. Therefore, it is somewhat meaningless to try and compare the composition change effects based on some overall RMSE or similar statistic. Instead, one has to examine the actual versus predicted moisture level plots to discern the magnitude of the errors introduced by changes in composition and their dependence on the moisture level of the sample. In order to reduce the clutter of the plots, while trying to conserve the space needed to present the results, we chose to plot the prediction results in three groups of three component changes. The groupings are purely arbitrary and simply follow the relative abundance of the components in the BY-104 simulant. The results of these predictions are shown in Figure 3-10, Figure 3-13, and Figure 3-16. In these plots, the broken lines correspond to the least squares regression line between the actual and predicted moisture levels for the different compositions. The solid line represents the Control model regression line. The prediction results are also summarized in Table 3-5 below. In the table, the errors are expressed as actual minus predicted wt% moisture and the shaded entries are within the RMSE of the Control model.

Pure Component	Moisture Level					
	0%	5%	10%	15%	20%	25%
$NaNO_3$	-0.363	-1.203	-1.790	-2.119	1.066	0.271
$NaAlO_2$	0.102	0.646	1.398	3.131	4.444	8.474
NaOH	-0.148	2.320	4.215	6.727	7.058	8.556
Na_2SiO_3	0.050	0.804	2.062	3.120	3.114	5.719
$Fe(NO_3)_3$	-0.076	-2.958	-3.505	-3.128	-3.031	-0.810
Na_3PO_4	0.104	0.951	2.732	3.171	4.565	2.478
$Ca(NO_3)_2$	-0.697	-2.650	-2.572	-2.120	-1.957	0.356
$Mg(NO_3)_2$	-1.901	-2.457	-3.663	-3.499	-2.583	0.141
$Mn(NO_3)_2$	-5.584	-5.552	-3.808	-2.713	-4.713	-2.056

Table 3-5. Average wt% moisture prediction errors (at each moisture level) for each pure component predicted from the two factor Control model.

From Table 3-5, one can see that the effect of the pure component changes had no real effect at the zero wt% moisture level, except for the $Mg(NO_3)_2$ and $Mn(NO_3)_2$ changes. As mentioned above, this effect is most likely due to moisture adsorption from the atmosphere of the hygroscopic pure components $Mg(NO_3)_2$ and $Mn(NO_3)_2$. From the results shown in Table 3-5, it is clear there are two main type of composition effects For the $NaAlO_2$, NaOH, Na_2SiO_3, and Na_3PO_4 components, the prediction errors are positive (i.e., the predicted wt% moisture is lower than the reference values) and increasing with increasing sample moisture level. The $Fe(NO_3)_3$, $Ca(NO_3)_2$, $Mg(NO_3)_2$, and $Mn(NO_3)_2$ components all had negative errors (i.e., higher predicted wt% moisture than the reference, or actual, values) and the effect was largest at the intermediate (10 and 15 wt%) moisture levels. Interestingly, the composition effect for these components disappeared at the 25 wt% moisture level, except for the $Mn(NO_3)_2$ component. Finally, the $NaNO_3$ composition change effect seemed to be intermediate between these two trends with negative errors up to the 15 wt% moisture level and positive errors for the two remaining high moisture points.

These general trends are more apparent in the plots below. Discussion of the possible causes for these errors is also deferred to the discussion of the plots below.

For the first group of three components, one can see from Figure 3-10, that the change in the $NaNO_3$ level did not have a large effect at the high or low moisture levels but has an effect at the intermediate moisture levels. The large variation at the 25 wt% moisture is most likely due to settling of the solid sample material in the slurry between the recording of the spectra. In all subsequent experiments, the 25 wt% moisture slurry samples were stirred between spectral measurements to eliminate this effect.

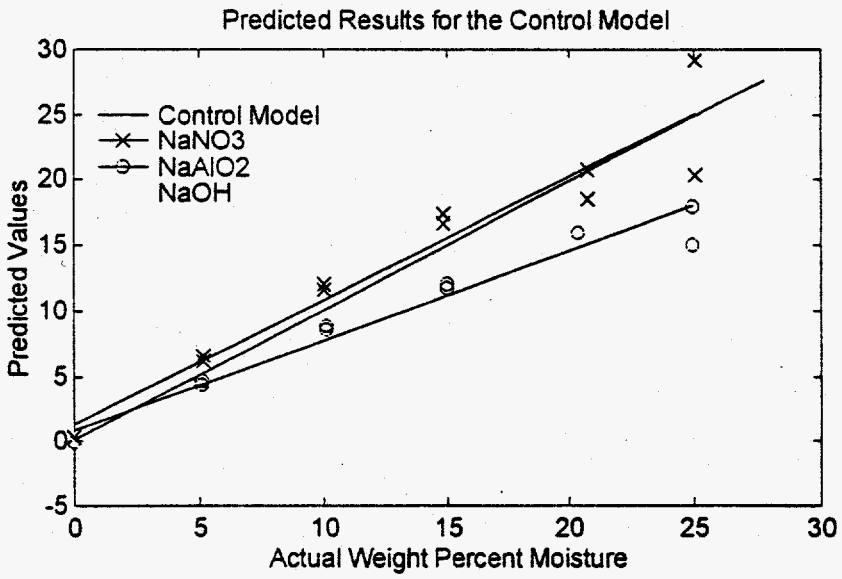


Figure 3-10. Predicted moisture results from $NaNO_3$, $NaAlO_2$, and NaOH spectral data using the Control model.

The $NaAlO_2$ and the NaOH composition changes had similar effects on the moisture prediction. For both these components, there was a gradually increasing bias towards low moisture prediction as the water level increased. This effect was more pronounced for the NaOH addition but the regression line for both had nearly

identical slopes. In order to explain the observed influence of the NaOH and $NaAlO_2$ composition changes, it is useful to look at the high moisture (25 wt%) spectra for the pure BY-104 and the samples with NaOH, and $NaAlO_2$ added. These spectra are shown in Figure 3-11 below.

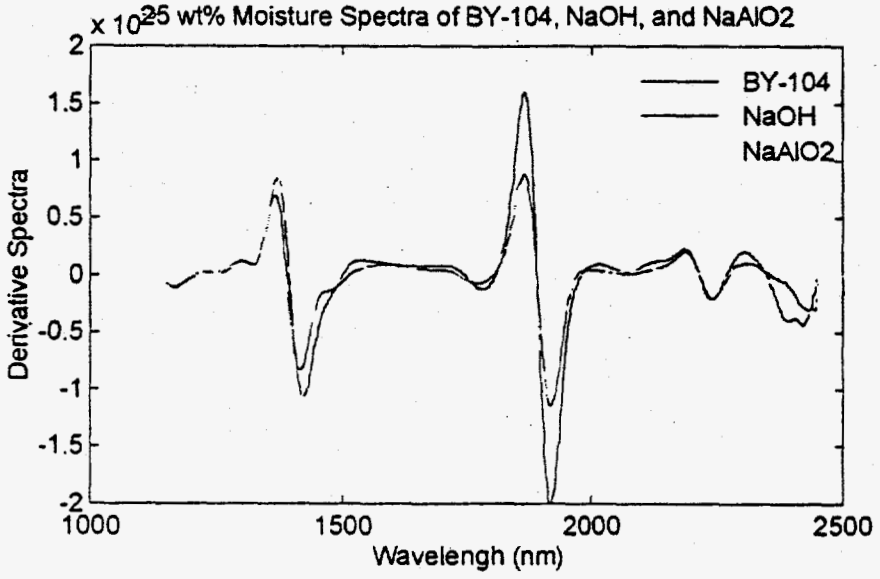


Figure 3-11. High moisture derivative spectra for the BY-104, NaOH, and $NaAlO_2$ samples from the composition experiment.

From Figure 3-11, one can see that the main difference between the original BY-104 spectra and the spectra of the samples with added NaOH and $NaAlO_2$ is associated with the peak at 1925 nm. The pure BY-104 has a much larger intensity than the corresponding peaks in the NaOH and $NaAlO_2$ spectra. One can also see a more subtle difference in the peak at 1400 nm. Here, the spectra for NaOH and $NaAlO_2$ both show a narrower peak than the pure BY-104. Since it is often difficult to interpret changes in the second derivative spectra and relate then to features in the original spectra, the same three spectra shown in Figure 3-11 are also plotted in Figure 3-12 in their original form. From this plot, it is clear that the reduced intensity for the 1925 nm peak observed in the second derivative spectra is really due

to an increased broad background adsorption in the NaOH and $NaAlO_2$ spectra which is not present in the pure BY-104 spectra. This background absorbance is most noticeable above 1900 nm in the NaOH spectra.

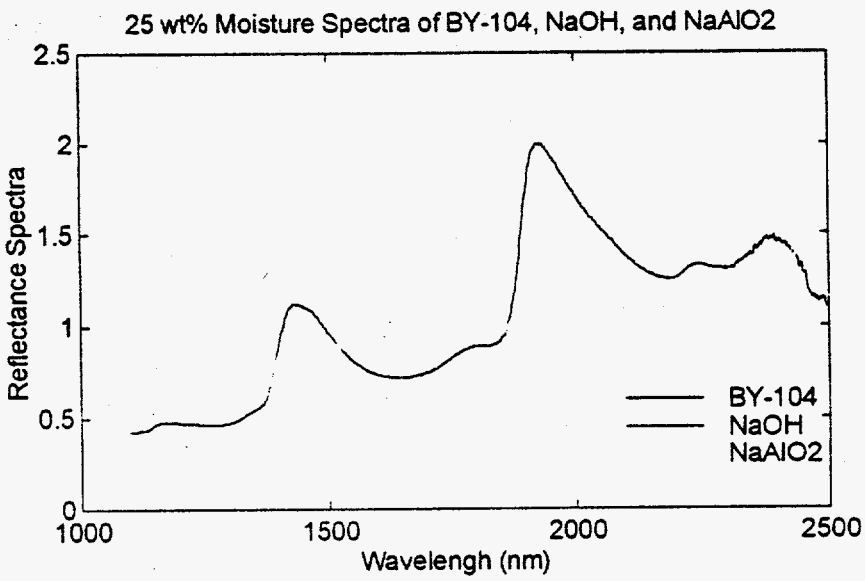


Figure 3-12. Original high moisture spectra for the BY-104, NaOH, and $NaAlO_2$ samples from the composition experiment.

For the next set of three pure components, the moisture prediction results obtained from the Control model are shown in Figure 3-13. In this plot we see that the effects of the Na_2SiO_3 and Na_2PO_4 component additions were similar to that observed for the NaOH and $NaAlO_2$ components. Again, both these components exhibited increasing bias towards low moisture predictions as the moisture level of the sample was increased. The magnitude of the prediction bias was not as large for these two components as it was for the NaOH and $NaAlO_2$ additions however. The effect of increasing the $Fe(NO_3)_3$ concentration produced quite a different effect in the predicted moisture level as shown in Figure 3-13 below. For this component, the effect on predicted moisture seems to be an almost constant offset, or bias, towards higher moisture levels for the samples in the 5 wt% to 20 wt% moisture range.

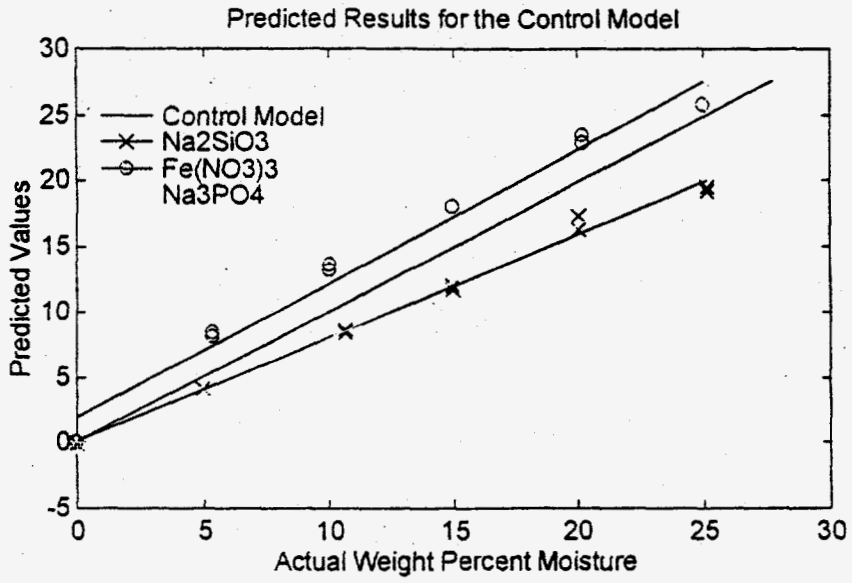


Figure 3-13. Prediction results from Na_2SiO_3, $Fe(NO_3)_3$, and Na_3PO_4 spectral data using the Control model.

The spectra for the pure BY-104 and $Fe(NO_3)_3$ samples with 10 wt% moisture are shown plotted together in Figure 3-14 and Figure 3-15 as the second derivative and original spectra, respectively. In the second derivative plot, the main spectral differences due to the increased $Fe(NO_3)_3$ content is an increased intensity for the peaks in the 1800 to 1950 nm region relative to the BY-104 spectra. There is also a slight intensity increase for the peak at 1400 nm, and a shift of the 1400 nm peak towards higher wavelengths in the $Fe(NO_3)_3$ spectra. In the plot shown in Figure 3-15, the spectral changes described above correspond to a general decrease in the background adsorption, and increased peak definition, in the original spectrum of $Fe(NO_3)_3$ relative to the BY-104 spectrum. The Na_2SiO_3 and Na_3PO_4 spectra are not plotted but their spectral response to the change in composition is similar to the NaOH and $NaAlO_2$ spectra described above, except that the spectra changes are much more subtle.

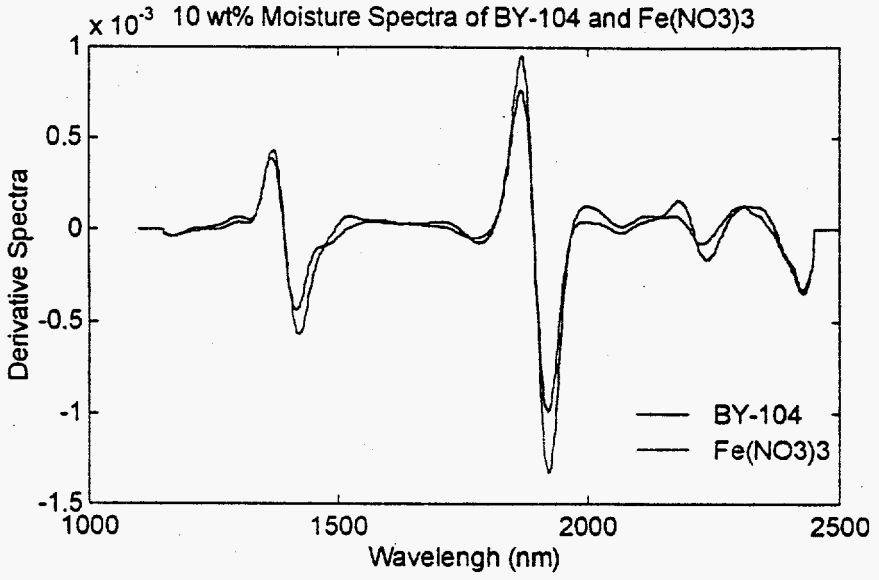


Figure 3-14. Derivative 10 wt% moisture spectra for the BY-104 and $Fe(NO_3)_3$ samples from the composition experiment.

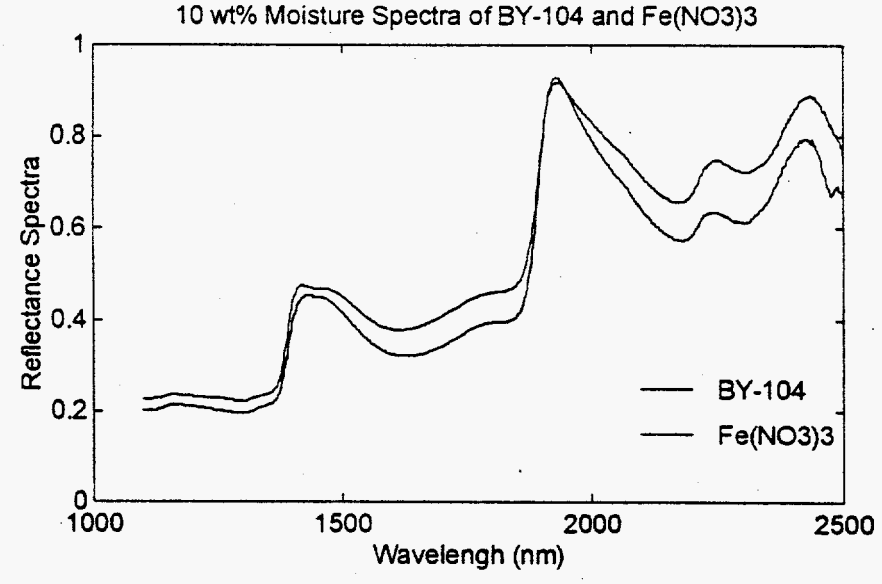


Figure 3-15. Original 10 wt% moisture spectra for the BY-104 and $Fe(NO_3)_3$ samples from the composition experiment.

For the third set of pure components, $Ca(NO_3)_2$, $Mg(NO_3)_2$, and $Mn(NO_3)_2$, the

prediction results are shown in Figure 3-16. As before, the predicted wt% moisture values were obtained by applying the Control model to the measured spectral data for each pure component addition. For each of the three components, the predicted moisture levels were higher for the low moisture samples and became progressively closer to the reference values with increasing moisture content. The $Ca(NO_3)_2$ and $Mg(NO_3)_2$ effect appears to be very similar to the effect seen for the $Fe(NO_3)_3$ composition change above. The original spectra of these two components (not included in this report) show the same type of suppressed, or reduced, background absorbance as we saw in the $Fe(NO_3)_3$ spectra plotted in the previous Figure 3-15.

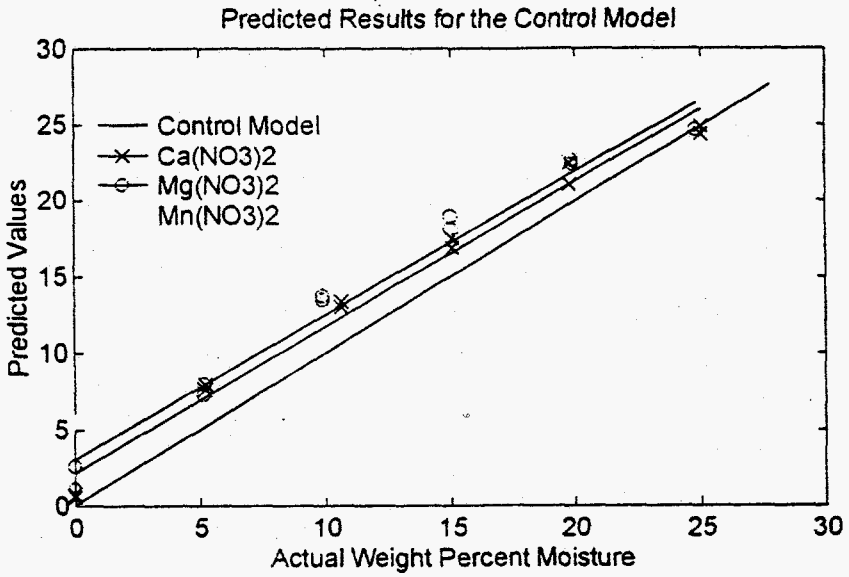


Figure 3-16. Predicted moisture results from the $Ca(NO_3)_2$, $Mg(NO_3)_2$, and $Mn(NO_3)_2$ data using the Control model.

For this group of three components, the difference between the actual moisture level and that predicted from the two factor Control model was highest for the $Mn(NO_3)_2$ data. Much of this effect can be directly attributed to the very hygroscopic nature of the pure components. When making up the zero wt% water sample for $Mn(NO_3)_2$, it was visibly apparent that all the pure component was dissolved into water adsorbed

from the air by the time the sample was weighed and mixed. After mixing, the sample appeared to be very similar to samples prepared at the 5 wt% moisture level instead of a zero wt% moisture sample. In the experimental procedure, the influence of the adsorbed atmospheric moisture on the total wt% moisture in the sample will naturally decrease as more water is successively added to achieve the higher wt% moisture samples of the experimental design. This is reflected in the pattern of decreasing errors at higher moisture levels seen in Figure 3-16 and Table 3-6.

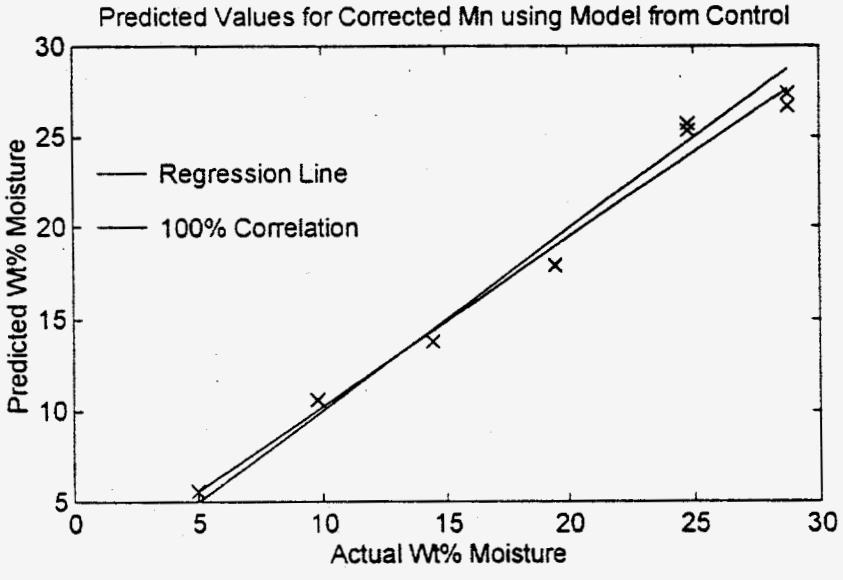


Figure 3-17. Corrected wt% moisture for $Mn(NO_3)_2$ versus the wt% moisture predicted from the Control model.

If we assume that the true wt% moisture for the starting base BY-104 plus the additional $Mn(NO_3)_2$ was actually 5 wt% due to water adsorption from the atmosphere, we can calculate a weight of adsorbed water equal to 0.2931 grams. If we correct the added water weights to account for this amount of extra water, the corrected $Mn(NO_3)_2$ moisture wt% values agree very well with the predicted moisture levels, as shown in Figure 3-17. This suggests that most, if not all, of the effect observed for the $Mn(NO_3)_2$ composition change is due to atmospheric water

adsorption. This is purely an experimental artifact which is not likely to occur in the analysis of real samples. A similar argument could be made for the observed large initial bias in the $Mg(NO_3)_2$ results, which also exhibited some water adsorption for the pure component.

It is somewhat encouraging that, even with the rather large adjustments to the compositions studied in these experiments, the effect on the moisture prediction is relatively small. The largest errors were seen for the NaOH data where the difference between the actual and predicted moisture values was about 8.5 wt% moisture. The corresponding difference in the NaOH concentration was 8.2% on a dry weight percent basis or an almost 500% increase in an absolute weight basis (from 0.09 grams to 0.46 grams in a five gram sample). Based on the results from this experiment, it was decided to attempt to develop sensitivity factors for the composition change of the three components (NaOH, $NaAlO_2$, and $Fe(NO_3)_3$) which exhibited the largest composition effects.

Composition Sensitivities

An additional set of experiments were run with several new composition levels of NaOH, $NaAlO_2$, and $Fe(NO_3)_3$. Spectra for each new composition was recorded at two different wt% moisture levels. The experimental design points for the new experiment are shown in Table 3-6. The original BY-104 composition design points (the 1.8 wt% NaOH samples at 15 and 25 wt% moisture) are not shown in this table. The same procedure used for the original composition experiment was followed in preparing the samples for this new experiment. For the NaOH component, six new samples were created consisting of three different compositions and two moisture levels per new composition. When combined with the additional composition data from the first composition experiment and the zero NaOH composition change data from the Control experiment, the expanded data set contained six different NaOH

composition levels ranging from 1.8 wt% to 10.01 wt% NaOH. For the $NaAlO_2$ and $Fe(NO_3)_3$ components, two additional compositions were evaluated at 15 wt% and 25 wt% moisture. The $NaAlO_2$ compositions ranged from 7.9 wt% to 16.02 wt% and the $Fe(NO_3)_3$ compositions ranged from 1.1 wt% to 10.07 wt% when the respective data from the earlier experiments were included.

Component Wt. %	Moisture Level	
	15 wt%	25 wt%
NaOH		
2.89	14.91	25.32
6.14	14.90	25.14
9.03	14.94	25.05
10.01*	14.97	25.06
$NaAlO_2$		
10.04	15.08	24.95
14.03	14.15	25.30
16.02*	15.04	24.92
$Fe(NO_3)_3$		
3.25	14.94	25.24
6.24	14.80	25.24
10.07*	14.97	25.00

Table 3-6. Experimental design for the extended composition experiment. The starred entries are the design points from the earlier composition experiment.

For each component, the new compositions spectral data were added to the spectra from the earlier experiments. This expanded data set was then used with the Control model to predict the wt% moisture. The difference between the reference values (calculated from the weight of water added in the experimental procedure above) and the predicted moisture from the Control model was used to estimate the effect of the composition changes. The moisture prediction results for the expanded NaOH composition data set are shown in Figure 3-18 below. From this plot it is clear that

the intermediate NaOH compositions from the additional experiments have proportionately less bias, or error, as the wt% NaOH is reduced. It is also clear that the effect of the NaOH composition changes is more apparent at the 25 wt% moisture level than at the 15 wt% moisture level.

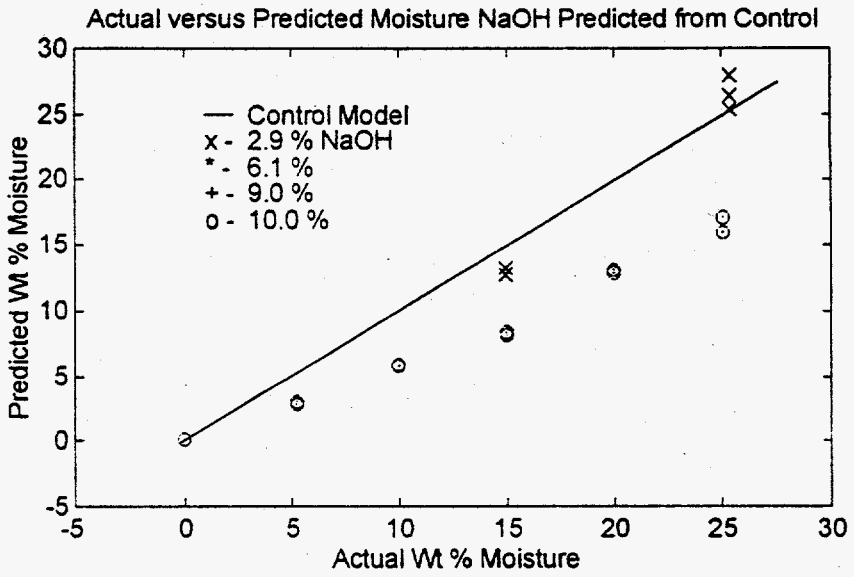


Figure 3-18. Moisture prediction results from the extended NaOH composition data using the Control model.

In order to obtain a sensitivity factor relating the prediction error to the weight percent NaOH change, the NaOH composition values were regressed against the prediction error (the difference between the actual and predicted moisture) at the 15 wt% and 25 wt% moisture levels. The results of these two regressions are shown in Figure 3-19. The solid line corresponds to the least squares regression equation for the 15 wt% moisture prediction errors and the dashed line the 25 wt% moisture errors versus wt% NaOH regression equation. For the pure BY-104 sample (the 1.8 wt% NaOH point) the RMSE of 0.7 wt% (plus or minus) for the Control model was used as the error. The resulting least squares model at 15 wt% moisture gave the slope of the regression line as 0.67, the intercept was -0.28 and the R^2 statistic was

0.98. At the 25 wt% moisture level, the slope was equal to 1.06 and the intercept was -2.98. Clearly, the sensitivity of the calibration model to variations in the NaOH composition is highly dependent on the moisture level of the sample. The higher the sample moisture, the more sensitive the model is to the NaOH concentration. At the 15 wt% moisture level, a change in the NaOH composition of one weight percent from the nominal value used in the calibration would increase the moisture prediction error by 0.67 wt%. This is very close to the RMSE of the calibration model itself and so probably would not be statistically significant. At 25 wt% moisture, a one percent change in the NaOH concentration will introduce about a one percent error in the predicted moisture. At moisture levels below 15 wt% the sensitivity to NaOH variations is expected to have a smaller, but still significant, effect. At zero wt% moisture the NaOH concentration change has no effect since there are no water peaks in the spectrum. However, the effect of NaOH variations are expected to be noticable, even at very low water concentrations.

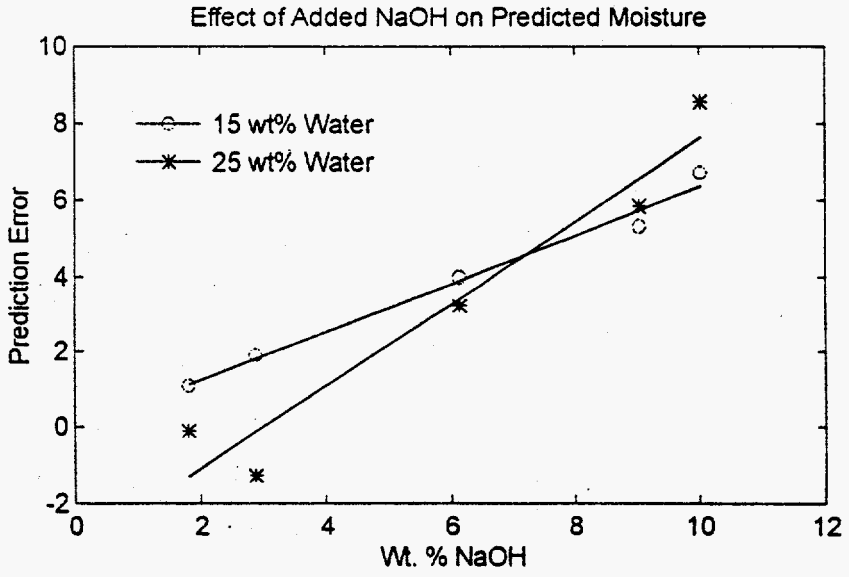


Figure 3-19. Regression models for wt% NaOH versus moisture prediction error at the two sample moisture levels.

For the $NaAlO_2$ extended composition data the moisture prediction results, using the Control model for the prediction, are shown below in Figure 3-20. The moisture prediction errors were regressed against the sample's wt% $NaAlO_2$ concentration to estimate the sensitivity of the model at the same two moisture levels described above. At the 15 wt% moisture level, the resulting least squares model had a slope of 0.32, an intercept of -2.18, and a R^2 statistic of 0.91. At the 25 wt% moisture level, the slope was 0.93 and the intercept was -8.61. The least squares regression results for both moisture levels are plotted in Figure 3-21. From this plot, and the regression parameters, it seems that the calibration model is about three times as sensitive to $NaAlO_2$ variations at the 25 wt% moisture level than it is at the 15 wt% moisture level. Below 10 wt% moisture the calibration model is less sensitive to $NaAlO_2$ concentration changes than at the higher moisture levels. At 15 wt% water, a change in the $NaAlO_2$ concentration of less than 2 wt% would have no significant impact on the moisture prediction accuracy. At 25 wt% moisture, the same 2 wt% change in the $NaAlO_2$ concentration would result in a statistically significant increase in the moisture prediction errors.

The moisture prediction for the extended $Fe(NO_3)_3$ data from the Control model is shown in Figure 3-22. For the intermediate $Fe(NO_3)_3$ compositions, the prediction errors are quite small. Regressing the 15 wt% moisture prediction errors against the wt % $Fe(NO_3)_3$ gives a least squares model with a slope of -0.43 and an intercept of 1.42. The same model for the 25 wt% moisture errors gives a model with a slope of -0.01 and an intercept of 0.95. The least squares regression models are shown in Figure 3-23 for both the 15 wt% moisture (the solid line) and 25 wt% moisture (dashed line) predictions. The negative slope of the regression models indicate that the predicted moisture values tend to be higher than actual when affected by the $Fe(NO_3)_3$ concentration. Clearly at the 25 wt% moisture level, the model is insensitive to changes in the $Fe(NO_3)_3$ concentration. At the 15 wt% moisture level, a two percent change in the $Fe(NO_3)_3$ concentration will be needed before the

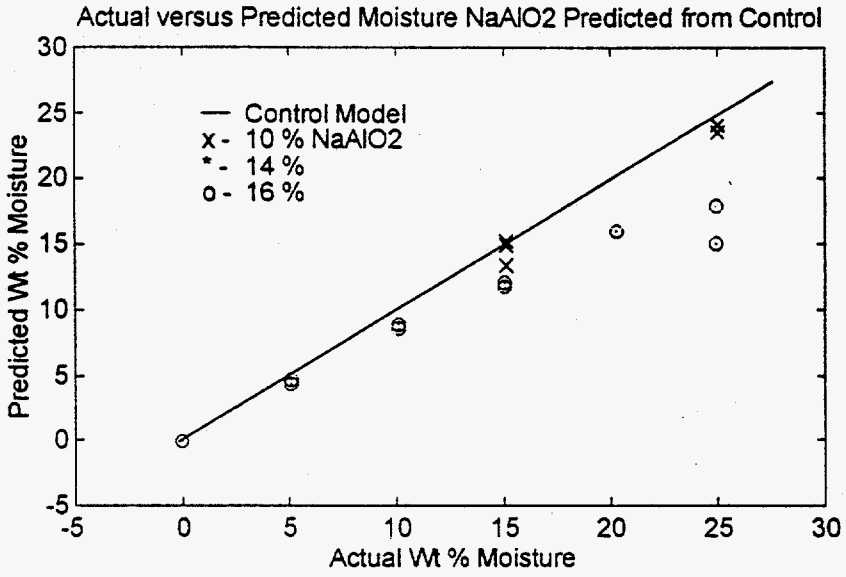


Figure 3-20. Moisture prediction results from the extended $NaAlO_2$ composition data using the Control model.

predicted moisture values will be affected above the RMSE of the model.

In examining the remaining components, Na_2SiO_3 and Na_3PO_4 had similar response to concentration changes as NaOH and $NaAlO_2$; although the magnitude of the sensitivity was less. The sensitivity of the moisture prediction to changes in the $Ca(NO_3)_2$ concentration was similar to that observed for $Fe(NO_3)_3$, although again the magnitude was less. The sensitivities for all the components are shown in Table 3-7 below. The numbers in the table correspond to the expected moisture prediction error (in wt% moisture) to a one percent increase in the wt% concentration of the component. Note that $Mg(NO_3)_2$ and $Mn(NO_3)_2$ are not listed in the table since the moisture prediction was essentially unaffected by concentration changes for these components.

Including Interferent Effects

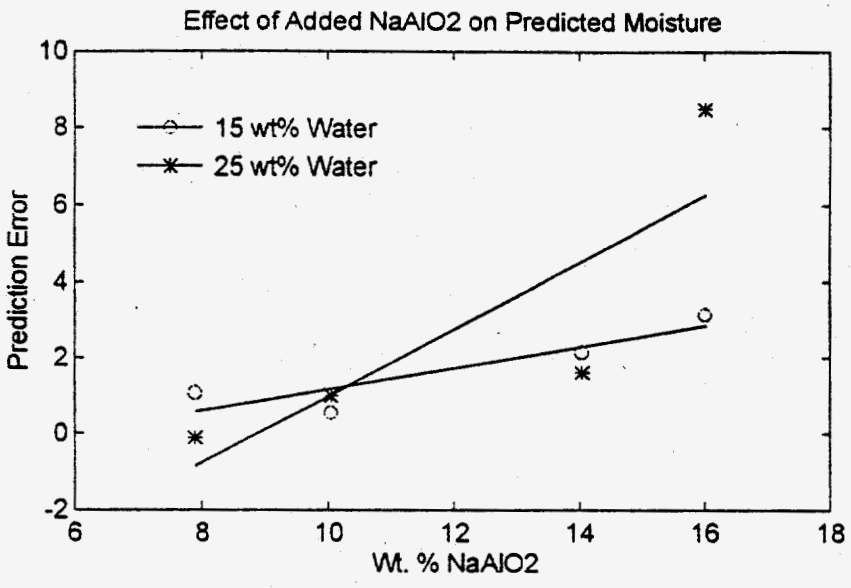


Figure 3-21. Regression models for wt% $NaAlO_2$ versus moisture prediction error at the two sample moisture levels.

Pure Component	Moisture Level	
	15%	25%
$NaNO_3$	-0.90	0.39
$NaAlO_2$	0.32	0.79
NaOH	0.67	1.02
Na_2SiO_3	0.29	0.61
$Fe(NO_3)_3$	-0.42	-0.01
Na_3PO_4	0.24	0.16
$Ca(NO_3)_2$	-0.29	-0.13

Table 3-7. Sensitivity of moisture prediction to changes in component concentration.

The results discussed above are all based on the Control model. In this model only the water concentration was varied, the calibration set spectra contained only the single BY-104 composition at different water levels. Since changing the sample composition affects the sample spectra in regions that overlap with the moisture level

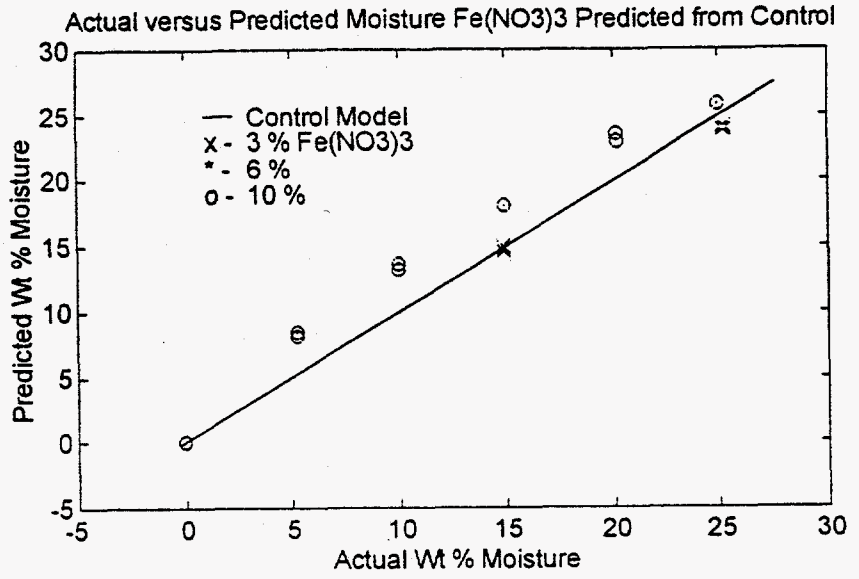


Figure 3-22. Moisture prediction results from the extended $Fe(NO_3)_3$ composition data using the Control model.

response, changing the BY-104 composition is really equivalent to adding, or changing, an interferent to the water calibration. Since the calibration model has no information about that interferent, larger prediction errors occur. Fortunately, correcting for known, or unknown, interferents is a strength of multivariate calibration methods such as PLS. In order to correct for the interferent(s), however, the calibration data must contain the interferent's response. It is not necessary to actually know the concentrations of the interferent(s) in order for the correction to occur, it is only required that the interferent(s) level change over the expected range of the interferent(s) in future prediction, or unknown, samples.

As an example of the interferent correction, the spectra and moisture information corresponding to the NaOH composition changes were added to the Control calibration set. A new calibration model was then generated from this expanded data set. Note that the actual wt% NaOH concentrations were not included in the data used to generate the model. The cross validation procedure indicated that for the

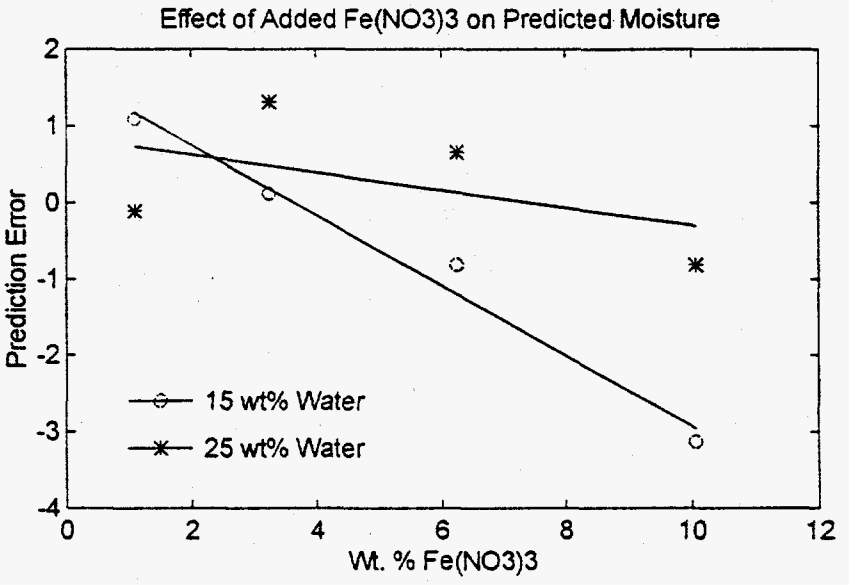


Figure 3-23. Regression models for wt% $Fe(NO_3)_3$ versus moisture prediction error at the two sample mositure levels.

expanded data set a four factor model should be used. The amount of spectral and moisture information described by each factor of this new model is shown in Table 3-8 below. The first factor of this new model describes 91% of the spectral information and 88% of the moisture information. Recall from Table 3-3 that the first factor of the Control model described 96% and 99% of the spectral and moisture information, respectively. Therefore, by adding the spectral variation related to the NaOH concentration changes, we have reduced the amount of spectral information described by the first factor by about 5% and the amount of moisture information by almost 11%. The second factor of the new model describes about 9% of the remaining moisture level information using about 6% of the spectral information. The third factor uses only a very small amount (about 1%) of the spectral information to describe about the same amount of the remaining moisture level variance. The fourth factor is a very small incremental increase to the variance described by the model in both blocks. Note that the new final model describes about the same total amount of information as the Control model.

	X-Block (Spectra)		Y-Block (Moisture)	
Factor #	This Factor	Total	This Factor	Total
1	91.49	91.49	88.21	88.21
2	6.51	98.00	9.09	97.30
3	1.15	99.15	1.44	98.74
4	0.30	99.45	0.31	99.05

Table 3-8. Percent variance described by each factor of the PLS model for the Control plus NaOH change data.

The fit of the new model to the expanded data set is shown in Figure 3-24 where the actual versus predicted moisture level values are plotted for those samples with NaOH composition variations. The new model has a RSME of 0.860 which is slightly higher than the original Control model's RMSE value of 0.734. Comparing Figure 3-24 with Figure 3-18 shows that, by including information about the NaOH variations, the PLS model was able to correct for the large prediction bias observed at the higher moisture levels for the samples with NaOH added. The correction was not able to fully account for all the variation in the 25 wt% moisture samples however. The RMSE for just the samples with NaOH variation was 0.956 wt% moisture when predicted with this new model.

The interesting thing about the new model is how it corrects for the interferent due to the different NaOH levels. This can be revealed by looking at the scores and loadings plots for the first few individual model factors. The spectra scores for the first model factor are shown in Figure 3-25. In this plot one can clearly see the progression from negative to positive scores for the first 36 samples. This pattern corresponds to the increasing moisture level for the three replicate Control experiments described above. The scores for samples 37 - 48 follow the same general pattern and correspond to the six moisture levels of the 10.01 wt% NaOH samples from the original composition variation experiment. Notice that the scores for these

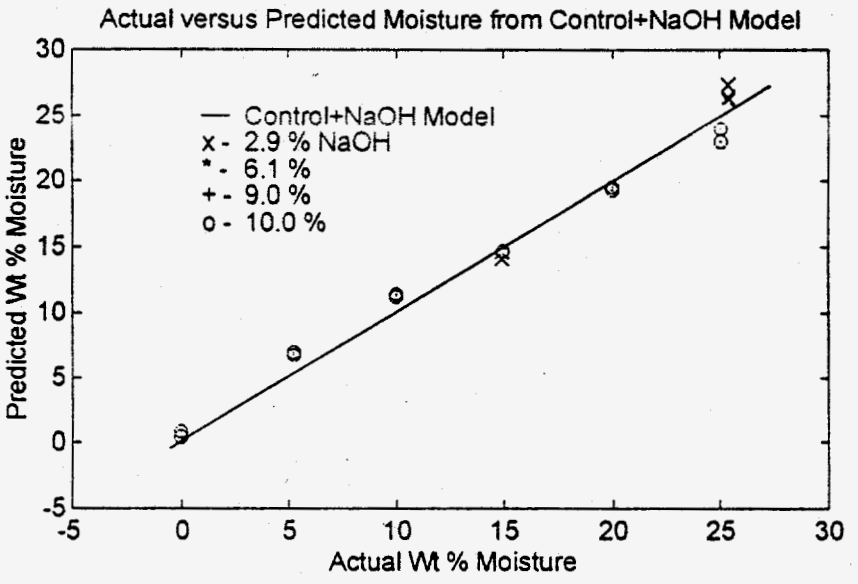


Figure 3-24. Actual versus predicted fit results for the three factor PLS model derived from the Control and NaOH data set.

samples only go slightly positive (to about the same value as the 15 wt% water samples from the Control experiment). Samples 49 - 63 correspond to the 3 wt%, 6 wt%, and 9 wt% NaOH samples from the second experiment (there were two spectra recorded at the 15 wt% moisture level and three at the 25 wt% level for each of these NaOH concentrations). The scores for these samples show the difference between moisture levels which is roughly equivalent to the difference between the corresponding moisture levels for the Control samples. However, notice that there is a general trend towards decreasing scores with increasing NaOH concentration. Given this interpretation of the score plot, it seems reasonable to assign to this factor the same moisture response as was observed in the original Control model. If this is true, then the loading plot for this factor should be almost identical to the loading on the first factor of the Control model.

The loading vector for the first factor of the new model is shown in Figure 3-26 along with the first factor loading vector from the Control model. It is clear that the first

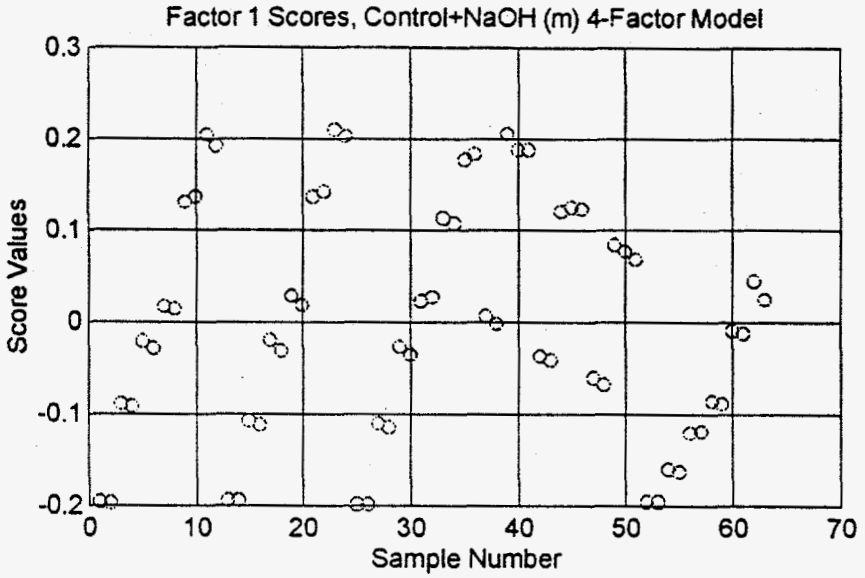


Figure 3-25. Spectra scores for the first factor of the new model built with the Control and NaOH composition data.

factor loadings for both the Control and Control+NaOH models are identical. This confirms our earlier analysis that this factor is describing the same moisture response information in both calibration data sets. Therefore, adding in the spectral variation related to the changes in the NaOH concentration has no effect on the first factor of the PLS model. This is not unexpected since the largest source of systematic variation in the spectral response is still due to changes related to the six different moisture levels of the experimental design. Recall that although the loadings for this factor are identical for both models, this factor is describing less of the total spectral and moisture information in the extended calibration data than it did for the Control data set. One can also conclude that since the first factor of the model contains none of the information related to the NaOH concentration variations, that the remaining factors of the model must contain this information.

The second factor scores for the spectra from the expanded calibration data set (Control samples plus the NaOH composition change samples) are shown in Figure

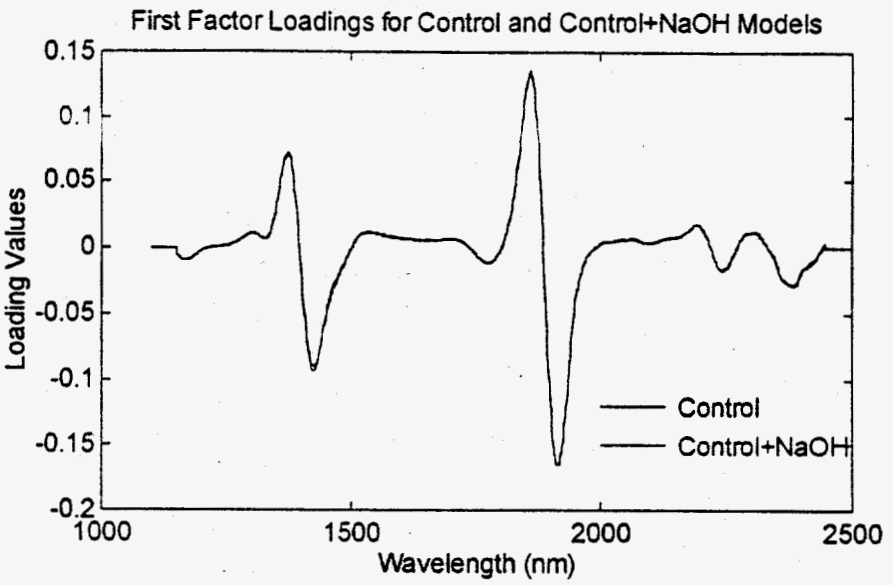


Figure 3-26. Loading vectors for the first factor of the Control and Control+NaOH models.

3-27 below. From this plot one can see that the second factor of the new model is mainly describing the difference between the spectra of the Control samples (original BY-104 composition) and the samples with different NaOH compositions. Again, samples 1 - 36 correspond to the original BY-104 composition at the six different moisture levels and samples 37 - 63 correspond to the different NaOH composition samples at the various moisture levels described above. In general, the original BY-104 samples all have similar scores centered around 0.1 and the added NaOH samples all have similar score values around -0.2. Recall from Table 3-7, that this factor describes about 6% of the spectral information and about 9% of the moisture level information in the Control+NaOH calibration data set. Therefore, this factor is describing not only the difference between the BY-104 and BY-104 plus NaOH spectra, but also the moisture level information which is correlated to that spectral difference

The loadings for the second factor of the Control and Control+NaOH models are shown in Figure 3-28. It is clear that the addition of the NaOH composition variation to the calibration set has significantly changed the second factor of the new model relative to the Control model. The main difference is related to the relative intensities of the peaks around 1400 nm (increase) and 1900 nm (decrease). In the new model, the peaks around 1400 nm have higher loading values than the peaks around 1900 nm. In the Control model, the relative loadings for these two spectral regions are reversed, with the 1900 nm region having higher loadings than the 1400 nm region. The Control model loadings in the region just above 1400 nm also tail off much more slowly than the corresponding Control+NaOH model loadings in the same region. Finally, in the spectral region around 2250 nm, the loadings for the second factor of the Control+NaOH model show new structure that is not present in the same factor from the Control model. All these observations are consistent with the change in the measured spectra between the original BY-104 composition, and the composition with added NaOH shown in Figure 3-11 and Figure 3-12 above.

The remaining two factors of the Control+NaOH model describe minor corrections to the moisture prediction model which are not clearly interpretable as either moisture level or NaOH concentration effects. Consequently, the score and loading plots for these two remaining factors are not presented in this report.

The regression coefficient vectors for the two factor Control model and the two factor Control+NaOH model are shown in Figure 3-29. [The remaining two factors of the new model were omitted from the calculation of the Control+NaOH model regression coefficient vector to simplify the comparison of the two model results.] In the new model, the peaks around 1400 nm have much higher coefficients than those around 1900 nm. This indicates that the Control+NaOH model is using more of the spectral information from the 1400 nm region, and less of the information from the 1900 nm region, to predict the moisture level than the Control model. Since the 1900 nm

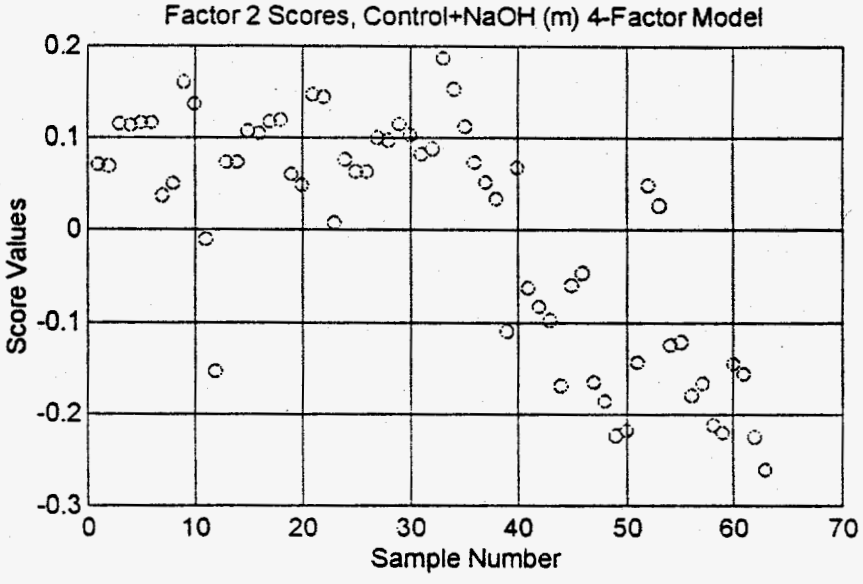


Figure 3-27. Spectral scores from the second factor of the new Control+NaOH PLS model.

region contains both the moisture response and the increased background absorbance due to varying NaOH concentration, the new model effectively discriminates against using as much of this spectral region in the final prediction. One can also see that the coefficients in the 1900 nm region are slightly shifted to lower wavelengths in the new model relative to the Control model coefficients in the same region. This indicates that the new model is using more of the information on the left shoulder of the 1925 nm OH combination band peak than the Control model. Again, this is due to the higher background absorbance attributed to the NaOH spectra in the region above 1925 nm. The first overtone band of OH (centered at 1425 nm) is also much sharper in the Control+NaOH model regression coefficients than in the original Control model. Finally, the regression coefficient from the Control+NaOH model includes information from the water combination band at 2250 nm whereas the Control model does not give much importance to that band.

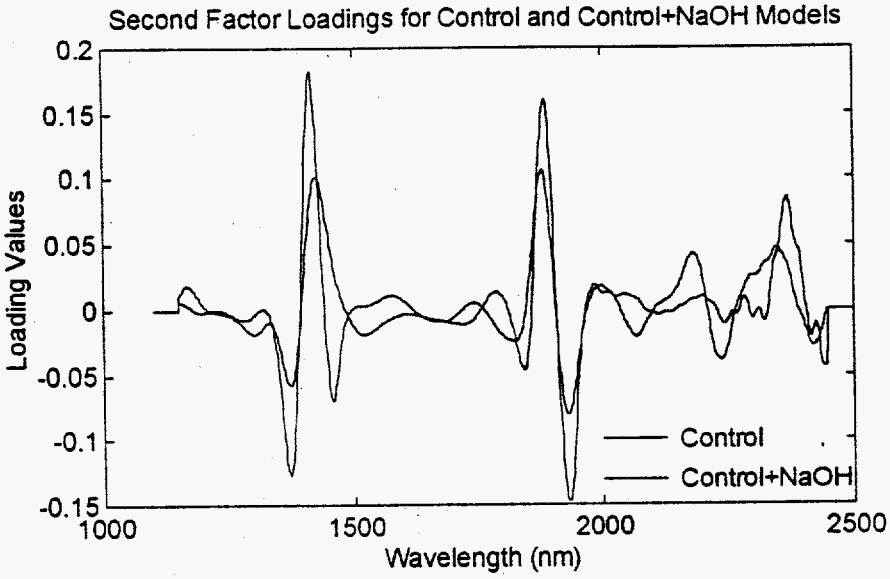


Figure 3-28. Spectral loading vectors from the second factor of the Control and Control+NaOH models. [Note the change at 1400 nm due to the added NaOH information.]

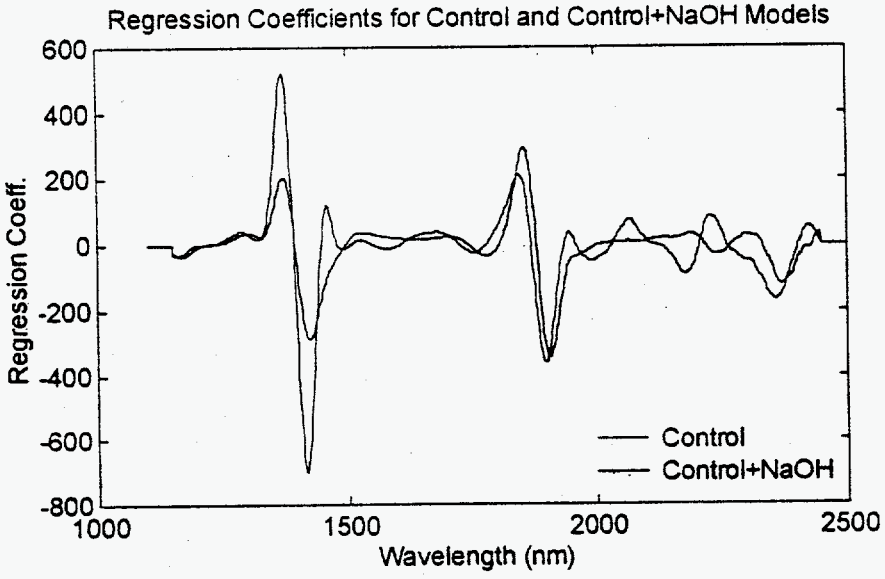


Figure 3-29. Regression coefficient vectors for the two factor Control and Control+NaOH PLS models.

Similar arguments could be made for the other component variations, where new models which include composition variation information, as well as moisture variation, could be developed to correct for the interfering nature of the composition change on the moisture prediction. The results of these models will not be presented in this report. Instead, if one attempts to correct for all the "interferents" related to variations in the BY-104 simulant composition, by including all the composition variations described above in the calibration data set, one can obtain an new estimate of the RMSE for the moisture prediction from this new overall model. We will call the model derived from all the composition variation spectra plus the Control spectra, the Overall model. The Overall model was determined to have 5 factors from cross validation studies. The five factor model accounted for 99% of the spectral block information and 97% of the moisture level block information.

The RMSE fit error for the prediction of wt% moisture was 1.41 using the Overall model. This is compared to the RMSE of 0.734 obtained from Control model prediction of the single BY-104 composition and the RMSE of 2.96 wt% moisture when predicting all the composition data from the Control model. By including all the composition variation into the model, the moisture prediction errors were roughly doubled from what we obtained for the single composition Control model. However, if the composition variation was not included in the model, the prediction errors were roughly four times greater than the single composition. In other words, including the composition variations degraded the model performance by a factor of two, but it resulted in a factor of two improvement over the results obtained with no correction for the composition variation interferences. Perhaps more importantly, the corrected Overall model results showed no bias at different moisture levels. Figure 3-30 shows the moisture prediction results for all the composition spectra using the Control model for prediction.

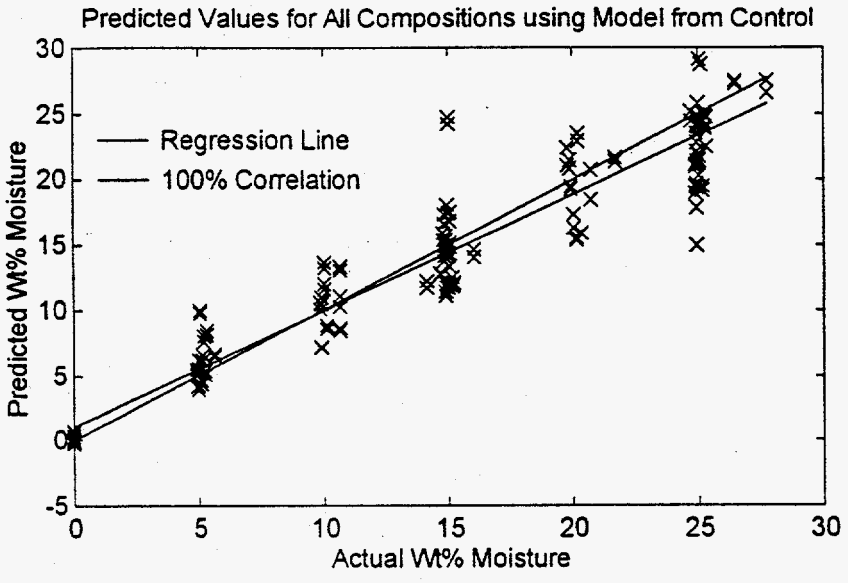


Figure 3-30. Moisture prediction values from all composition variation spectra using the Control model.

Figure 3-31 shows the corresponding plot using the Overall model for the prediction. From the plot in Figure 3-30 one can clearly see that without the composition interferent corrections, many of the higher moisture sample show a large negative bias. The source of this bias has been explained above and is due mainly to the increased background absorption due to the pure component's spectra. Notice that for the Overall model, the prediction errors are not only lower, but more uniform over the moisture range. There is still a higher prediction error for the high moisture samples, but the errors are more randomly distributed about the true moisture value. Notice that the zero moisture samples are now predicted as negative due to the corrections as shown in Figure 3-31.

The individual component sensitivities derived from the Overall model are listed in Table 3-9. The RMSE values also listed in Table 3-9 are for the moisture prediction for the samples with composition changes for that component using the Overall model. The numbers in the table correspond to the expected moisture prediction

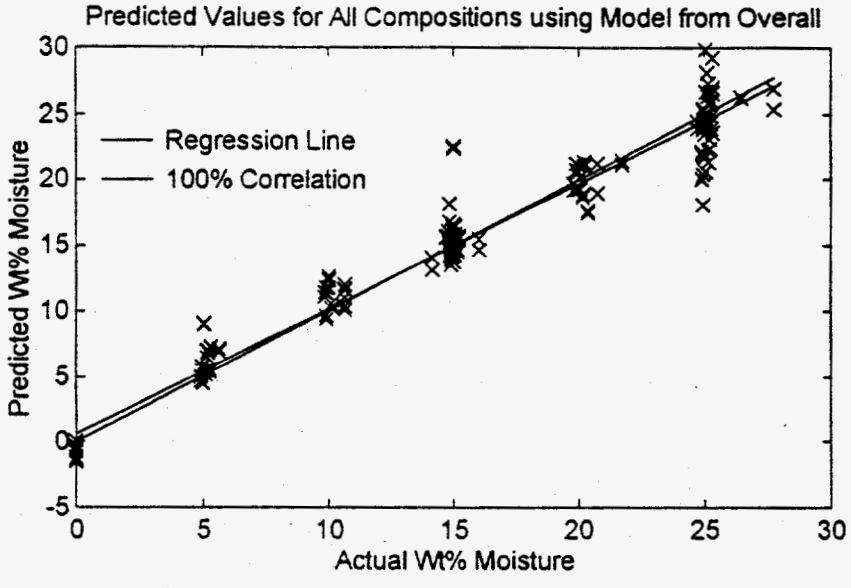


Figure 3-31. Moisture prediction values from all composition variation spectra using the Overall model.

error (in wt% moisture) to a one percent increase in the wt% concentration of the component. Compared to the sensitivities listed in Table 3-6 the use of the Overall model reduced the sensitivity of the moisture prediction errors to changes in the composition by at least a half relative to the Control model results. For most components, the new sensitivities from the Overall model are so low that changing the composition no longer has a discernible effect on the prediction error.

The obvious conclusion from this study is that including a range of expected compositions is necessary for developing good calibration models. If one limits the calibration data to a single, or too narrow a range of, compositions, the resulting model will be unable to accurately predict future samples with slightly different compositions. By increasing the range of variability in the calibration model compositions, the resulting model may not be able to fit the data as well, but this tradeoff is worth making in terms of increased robustness of the model to predicting future unknown samples. The implication for application of the spectroscopic

moisture determination method to waste tank samples is that a range of tank waste compositions may need to be measured in order to achieve reliable moisture predictions. Of course, the importance of having a range of compositions in the calibration set will depend on the magnitude of the composition variation within the waste tanks and the range of moisture within the waste.

		Moisture Level	
Component	RMSE	15%	25%
$NaNO_3$	2.31	-0.45	-0.08
$NaAlO_2$	1.58	-0.04	0.32
NaOH	1.26	-0.06	0.16
Na_2SiO_3	1.55	0.03	0.29
$Fe(NO_3)_3$	1.69	-0.12	0.03
Na_3PO_4	0.77	0.00	-0.02
$Ca(NO_3)_2$	1.54	-0.01	0.08

Table 3-9. Sensitivity of moisture prediction to changes in component concentration for the Overall model.

NaOH Concentration Prediction Model

As an aside, if one were to replace the moisture level values in the Y-block with the NaOH concentrations and build a new model using the Control+NaOH spectra, the resulting model could be used to predict the NaOH concentration in new samples. Cross validation indicated a three factor model should be used for the NaOH prediction. The amount of variance explained by each factor of this new NaOH model is shown in Table 3-10 below. The first factor accounts for 78% of the spectral information but only 24% of the NaOH concentration information. The second factor uses 19% of the spectral information to describe 58% of the NaOH concentration

information. The third factor describes only 1% of the spectral information and 13% of the NaOH concentration variance.

	X-Block (Spectra)		Y-Block (NaOH conc.)	
Factor #	This Factor	Total	This Factor	Total
1	78.45	78.45	24.12	24.12
2	19.49	97.95	58.42	82.55
3	1.17	99.12	13.40	95.95

Table 3-10. Percent variance described by each factor of the NaOH model from the Control and NaOH composition change data.

As before, the first factor of the NaOH model describes the water addition experimental design and is the exact inverse of the first factor obtained from the Control+NaOH model. Even though we have replaced the moisture level information in the Y-block with the NaOH wt% composition values and told the model building algorithm nothing about moisture levels, the main source of variation in the recorded spectra is still due to the moisture addition experiments. In other words, the moisture related variation in the spectra was still present as an unknown interferent to the NaOH calibration. This first factor can then be interpreted as a moisture level correction to the overall NaOH calibration model.

The second factor describes more of the NaOH concentration information and turns out to be identical to the second factor of the Control+NaOH model described above. However, for this model the second factor is not an interferent correction for the NaOH concentration effects on the moisture prediction, but rather the main factor used to describe the correlation between the spectral responses and the NaOH concentration. This subtle difference can be seen when one looks at the relative percentage of the Y-block information described by this factor in the two models. For the Control+NaOH model the second factor described only 9% of the moisture level

information. For the NaOH model, the same factor (the scores and loadings are identical between the two models) now describes 58% of the NaOH concentration information. In developing the model, the independent factors, or unique sources of variance, in the spectral data is largely governed by the experimental design used to collect the data. The relative importance of each of these factors to the calibration is then determined by their correlation to the quantity being predicted. The score and loading plots for the first two factors are not shown since they are essentially the same as Figure 3-25, Figure 3-26, Figure 3-27, and Figure 3-28 and their interpretation would be the same as for the Control+NaOH results discussed above. The third factor represents a correction to the model to describe mainly nonlinear moisture effects and the scores and loading plots are not readily interpreted and are also not shown in this report. What is interesting is to look at the X-block (spectral data) versus Y-block (NaOH wt% concentrations) scores for each factor of the NaOH model.

For the first factor of the model the spectral versus concentration scores are plotted in Figure 3-32. In this plot one can clearly see the five different NaOH concentration levels. The large collection of points with Y-block scores less than -2 correspond to the 1.8 wt% NaOH concentration samples from the pure BY-104 Control experiment. The samples with Y-block scores between -2 and zero are the 2.89 wt% NaOH samples, the Y-block scores at 2 correspond to the 6.14 wt% NaOH samples, and those with Y-block scores near 5 are the 9.03 wt% NaOH samples from the additional composition experiments. The samples with Y-block scores near 6 are the 10.01 wt% NaOH samples from the composition experiment. It is obvious that the spectral, or X-block, information described by this factor (seen as the dispersion of the sample points along the horizontal axis) is not very correlated to the NaOH concentration information. If the X-block information described by this factor was perfectly correlated to the Y-block information, the samples points would all fall along a 45° line through the plot. The solid line shown in Figure 3-32 is the "inner regression"

line from the PLS model which describes the fit of the X-block scores to the Y-block scores. It is clear from this plot that the first factor is describing spectral variation information which is largely uncorrelated to the different NaOH concentration levels of the samples. Again, this first factor is related to the water information in the spectra and serves to correct the NaOH model for the moisture level interferent.

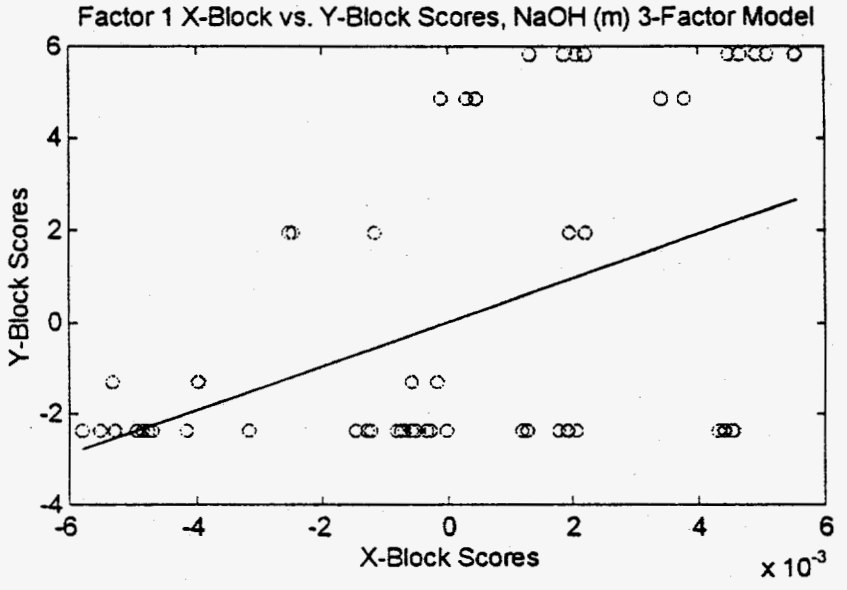


Figure 3-32. Spectral block scores versus the NaOH concentration block scores for the first factor of the NaOH model.

For the remaining two factors, the X-block versus Y-block scores are shown in Figure 3-33 and Figure 3-34 below. From Figure 3-33 one can see that the second factor is describing more correlated information from the spectral and NaOH concentration blocks. However, there is still considerable scatter off the inner regression line in this plot. The corresponding plot for the third factor of the NaOH model is shown in Figure 3-34. Here one can see that the linear correlation between the spectral and concentration blocks is fairly good. This factor is strongly influenced by a few samples with very low scores.

The resulting regression coefficient for the NaOH model is shown in Figure 3-35. Notice that the NaOH predictions are based almost entirely on the first OH overtone band around 1425 nm. The OH combination band at 1925 nm has only a small contribution to the NaOH prediction, mainly due to the strong overlap of the moisture variation interferent in that region. There is also a minor contribution from the 2250 nm OH combination band. The lack of features in the spectral regions above the 1400 nm region is mainly due to the moisture interferent.

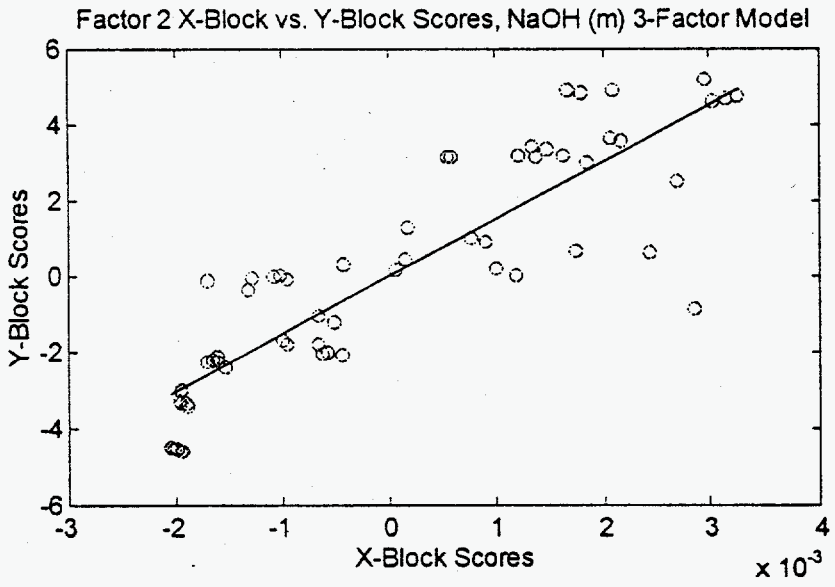


Figure 3-33. Spectral versus NaOH concentration scores for the second factor of the NaOH model.

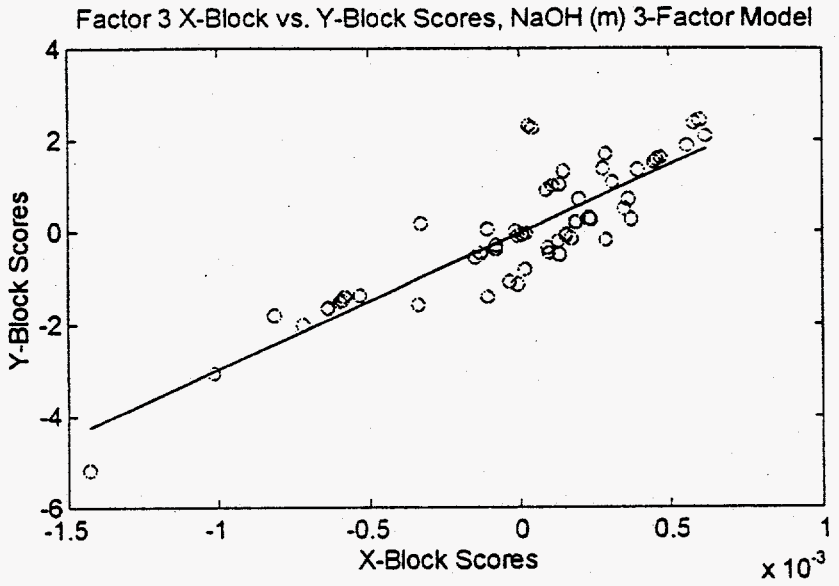


Figure 3-34. Spectral scores versus NaOH concentration scores for the third factor of the NaOH model.

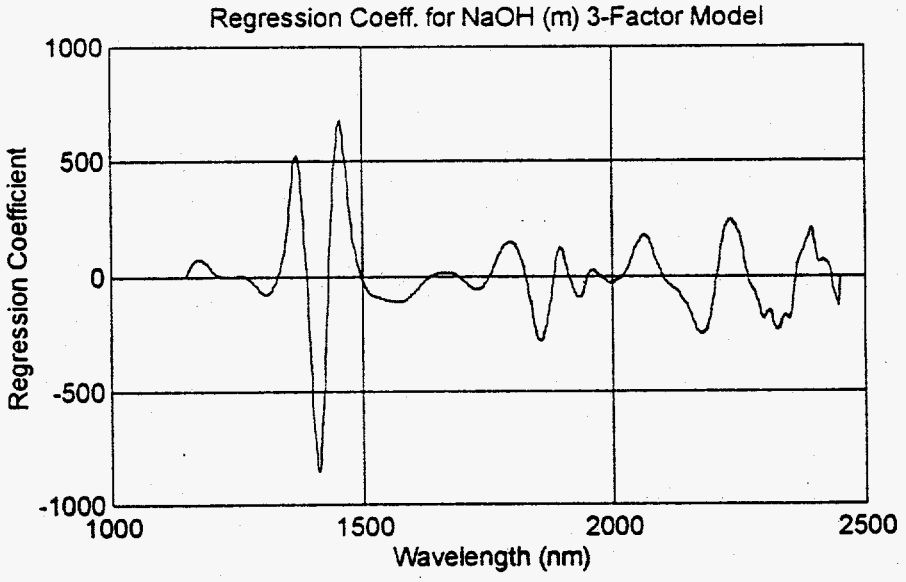


Figure 3-35. Regression coefficient vector for the three factor NaOH model.

The NaOH prediction results are shown in Figure 3-36 where the actual NaOH concentrations are plotted versus the NaOH concentrations estimated by the NaOH model. The RMSE for the NaOH prediction was 0.68 wt% NaOH with a R^2 statistic of 0.96. The predicted values show slightly higher variability at the higher NaOH concentrations, but in general the prediction results show little bias over the NaOH concentration range. Remember, that the NaOH model already implicitly includes corrections for the moisture level but does not include correction for other composition variations. Since we really don't have sufficient data to build a calibration model for NaOH including all the composition variations, we cannot estimate the effect varying composition would have on the NaOH predictions. However, we can speculate that the effects would be smaller for the NaOH prediction model than were observed for the moisture model, This speculation is based on the fact that most of the pure component spectra do not show as strong features in the 1400 nm region primarily used for the NaOH prediction as they do in the 1900 nm region used for the moisture prediction.

Conclusions

In this study the effect of composition variations on the moisture prediction were investigated. The BY-104 simulant was used as the base composition and the individual components of the simulant were varied to achieve different sample compositions. An experimental procedure was developed to generate samples with different moisture and composition values. This study also involved the first use of a fiber optic probe for the measurement of the sample spectra. The experimental design used in this study allowed for the moisture level and composition effects to be studied independently and together. In all cases, the experimental data was analyzed using Partial Least Squares (PLS) regression to generate appropriate calibration models. Some conclusions from the experimental and data analysis results are presented below.

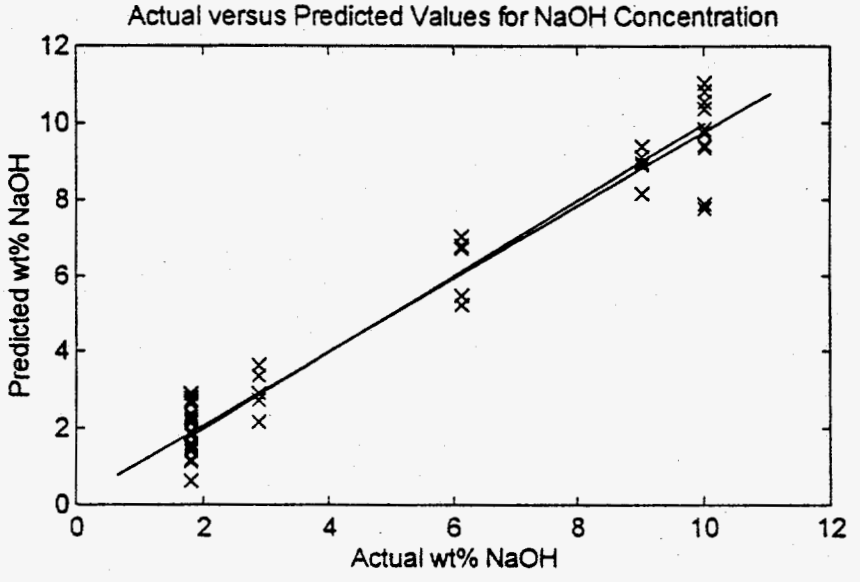


Figure 3-36. Actual versus predicted wt% NaOH from the three factor NaOH model.

When considering the data without composition changes the two factor model was able to predict moisture with a root mean square error (RMSE) of 0.734 wt% moisture. This prediction error estimate was slightly higher than the estimate from the previous Phase 1 study (0.42 wt% moisture) but most of the difference was attributed to between experiment variability in the experimental design. The use of the fiber optic probe did not seem to adversely effect the moisture predictions and has several advantages over the previous method of using sample cups for measuring the spectra. Unfortunately, the experimental setup restricted the spectral region to the NIR region (1100 nm to 2500 nm) with the use of the fiber probe. It was observed that the experimental procedure did introduce an error, or bias, in the high moisture values due to sample sticking to the probe during the spectra measurement process. This bias was purely an experimental artifact and should not present a problem when using the probe for routine moisture determinations.

In analyzing the composition variation data, it was observed that the effect of changing the BY-104 component compositions did adversely effect the moisture predictive ability of the calibration model. The nature of the composition effect was dependent both on the simulant component which was changing and the moisture level of the sample. There were two general effects that were observed. For the NaOH, $NaAlO_2$, Na_2SiO_3, and Na_3PO_4 components, the effect was a bias towards lower than actual moisture predictions. This effect was larger at the higher moisture levels and was most pronounced for the NaOH component. The source of this bias was related to the increased spectral background due to the component spectra around the 1424 nm first OH overtone and 1925 nm OH combination band. The second main effect was observed for the $Fe(NO_3)_3$, $Ca(NO_3)_2$, and $Mg(NO_3)_2$ components which exhibited increased predicted moisture values relative to the actual sample moisture values. This effect was strongest for the $Fe(NO_3)_3$ composition changes and was mainly observed at the low to intermediate moisture levels. The main cause for this observed sensitivity was a decrease in the background adsorption in the 1400 nm to 1900 nm region due to the pure component spectra. The $NaNO_3$ composition changes mainly seemed to increase the nonlinear component of the sample's spectral response to moisture level increases while the effect of the $Mn(NO_3)_2$ component was negligible (other than an experimental effect due to the highly hygroscopic nature of the pure $Mn(NO_3)_2$).

It was observed that the effect of the pure component additions to the BY-104 simulant on the resulting spectra strongly overlapped the moisture response spectral regions. Therefore, the changes in composition were effectively unknown interferents to the moisture calibration model. To illustrate the multivariate model's ability to correct for these interferents, a model was developed which included both moisture variation and NaOH concentration variation effects in the spectra. This model was able to correct for the NaOH composition change interferent and predict the moisture level without the bias observed for the pure BY-104 calibration model. The RMSE

for the corrected model was 0.860 wt% moisture versus a RMSE of 4.80 wt% moisture for the model without the correction (the errors for the high moisture samples were even larger due to the strong bias of this model). The corrected model regression coefficient vector indicated that correcting for the interference caused by the NaOH concentration variations resulted in an increased emphasis on the 1425 nm spectral region (OH first overtone band) and a decreased emphasis on the 1925 nm spectral region (OH combination band), relative to the single composition BY-104 moisture model.

When all the composition variations were included in the model, the resulting model had a moisture RMSE of 1.41 wt% moisture. This compares to a RMSE of 0.734 wt% moisture for the Control model fit to the calibration data set and a RMSE of 2.96 wt% moisture for all the composition's data predicted with the single BY-104 composition Control model. Therefore, by including the spectral variation due to the composition changes along with the moisture variations, the resulting model was able to correct for the interferences due to composition changes and reduce the prediction error by more than a factor of two. The price paid for the increased model complexity was that the moisture prediction errors were twice as large. However, the resulting model is now much more robust to composition variations. The sensitivities for most of the BY-104 components were below the statistically significant level over the composition variation ranges studied when the corrected model was used.

The implication for future development and deployment of the spectroscopic moisture monitoring system is quite clear. It will be important to develop the moisture calibration models with a range of compositions expected to be encountered in the routine analysis of waste from the waste tanks. The exact sensitivity of the moisture prediction model will depend on the magnitude of the composition variation within the waste tanks, the waste components which are varying, and the desired moisture range of the calibration model. These studies indicate that by including composition

variation in the calibration data, the resulting model can be made relatively insensitive to composition variation.

Finally, it appears that the same spectra used for the wt% moisture prediction can be used to predict the wt% NaOH in the same sample. The NaOH calibration model was seen to use mainly the first overtone of the OH band for the wt% NaOH prediction. The RMSE of the NaOH prediction was 0.68 wt% and already included corrections for the water level interferent. Of course, it should be remembered that the NaOH calibration models only considered NaOH concentrations above the BY-104 concentration and did not include corrections for the other component composition changes. Even so, the resulting NaOH prediction errors were quite good and suggests that the potential exists to monitor both the wt% moisture and wt% NaOH in a waste sample using a single spectroscopic measurement.

4. Sample Drying versus Water Addition Study

In this study we investigated the effect of two different sample treatments in preparing the calibration data set for moisture determination. The first sample treatment involved the complete drying of the BY-104 simulant followed by successive additions of water to obtain samples with different moisture levels. This was the method used to obtain the calibration samples for the Phase 1 studies. The second sample treatment involved drying the wet BY-104 simulant sample for increasing periods of time. As the drying time was increased, the sample contained progressively less moisture. This second sample treatment, which we will call successive drying, was investigated to determine if there is a quantitative difference in the moisture model calibration results related to the two different methods of preparing the calibration samples. The first method of water additions to the dry sample is faster, easier to implement, and allows one to more easily control the experimental design. However, the successive drying treatment may be more likely to approximate the mechanism in the waste tanks whereby the waste could have different moisture levels.

Experimental

Three data sets were obtained from the successive drying experiments for both the VIS (400 - 1100 nm) and NIR (1100 - 2500 nm) spectral regions. The first data set, which we will call Run_1, contained 9 discrete sample points corresponding to increasing drying times. In this experiment the sample was dried in the oven for 5 minutes, allowed to cool for 15-20 minutes in a desiccator, and then placed in the spectrometer where the spectra were recorded. The sample remained in a metal sample cup with a glass window throughout the drying and spectral measurement

process. The weight of the sample was recorded immediately prior to each spectra collection step. For each sample point, two spectral scans were recorded with the sample cup repositioned between replicate scans. The original wet BY-104 simulant was independently determined to contain 10.82 wt% water by complete oven drying. The moisture range was calculated to be from 10.22 wt% water to 4.16 wt% water for this data set (determined from the original moisture level and weight difference between measurement points). It should be noted that this experimental procedure was not continued for the length of time required to fully de-hydrate the BY-104 simulant.

The second data set, which we will call Run_2, consisted of 10 sample points with a 10 minute drying period between the cooling and spectral collection steps. Again, two spectra were collected at each sample point and the sample remained in the sample cup throughout the drying and data collection cycle. For the Run_2 data, the moisture content calculated from the weight differences ranged from 10.97 to 0.75 wt% water. Again, the total length of the experiment (approximately 6 1/2 hours) was probably not long enough to completely dry the simulant.

The third data set consisted of two full replicates of 15 different drying times. This data set, which we will refer to as Run_3, was collected with two differences in the experimental procedure from Run_1 and Run_2. First, the BY-104 simulant was dried outside the sample cup and repacked in the cup for each sample point. The rationale was that keeping the sample in the cup throughout the experiment introduces the possibility of sample inhomogeneity as the material at the top of the cup is open to the oven environment while the material at the bottom of the cup (against the spectral window) is closed to the environment. By repacking, the sample would be mixed prior to taking the spectra. The second change was that the drying times were increased as the moisture level decreased. The first five sample points were taken at two minutes drying time between the cooling and data collection steps.

The second five samples were obtained with 5 minutes drying time and the last five sample points were dried for 10 minutes. For all the sample points, the sample was removed from the oven and allowed to cool for approximately 20 minutes prior to collecting the spectral data. The moisture levels for this run ranged from 12.09 to 0 wt% water.

All experimental spectra were collected using the same spectrometer used in the Phase 1 studies. The spectrometer was operated in the reflectance mode using the standard sample cup attachment and controlled from an PC computer. This spectrometer is capable of scanning both the VIS and NIR spectral regions. The collected spectra were transferred to a computer workstation for analysis using CPAC developed software.

Results and Discussion

In discussing the drying study results, we will first concentrate on the NIR spectral region. While the VIS region results were similar for the individual models, the correlation of the drying results to the water addition results previously obtained were quite poor. Possible reasons for this will be presented after the NIR discussion.

Recall that the main goal of this study was to determine if there was any advantage, or disadvantage, to preparing the moisture calibration models using the method of successive drying of the simulant rather than the method of water addition to dry material used in the previous studies. In order to answer this question, separate PLS models were developed for each of the three drying runs described above. The performance of these individual models were then compared to the PLS model results obtained from the water addition experiments from the Phase 1 study. In addition, the effect of the two different sample treatments was checked by cross prediction of

the data obtained with one treatment using the PLS models derived from the other sample treatment's data.

NIR Spectral Region Results

Rather than presenting all the information from the PLS models for each of the three runs in this study, the results are summarized and only the significant, or interesting, features of the models are discussed. For each of the runs, cross validation studies indicated that a two factor PLS model gave the best predictive model without overfitting the data. The information content of each factor of the individual run's PLS model is shown in Table 4-1 below. Runs 2 and 3 are very similar in the amount of information described by each factor. For the Run_1 model, the first factor describes more of the spectral response (X-block) and the second factor describes more of the concentration information (Y-block) than the other two models. This is the first clue that Run_1 might be different from the other two runs and that the concentration information is less correlated to the spectral variation than in the subsequent runs.

The PLS model performance for the three sample drying runs, and the previous results from the Phase 1 study, are summarized in Table 4-2 below. For all three runs the performance is essentially the same as the previous results reported for the Phase 1 NIR moisture model which used the water addition method for generating the calibration samples. Therefore, in terms of modeling ability, there does not seem to be any real difference between the water addition and the successive drying sample treatments; although the drying data does require an extra factor in the model to achieve the same results. Of course, the modeling ability is not the only criteria (or even the best one) for evaluating the effect of the two sample treatments. The discussion that follows looks at differences between the models for the various treatments and then explores the relative predictive ability of models built with the

	X-Block (Spectra)		Y-Block (Moisture)	
Run_1				
Factor #	This Factor	Total	This Factor	Total
1	99.45	99.45	87.74	87.74
2	0.42	99.87	9.50	97.24
Run_2				
Factor #	This Factor	Total	This Factor	Total
1	97.22	97.22	95.28	95.28
2	1.68	98.90	3.38	98.66
Run_3				
Factor #	This Factor	Total	This Factor	Total
1	97.60	97.60	96.83	96.83
2	1.09	98.69	2.10	98.93

Table 4-1. Percent variance described by each factor of the PLS models for the thre successive drying experiments.

two sample treatments.

	Run_1	Run_2	Run_3	Phase 1
SEP (cross val.)	0.532	0.537	0.457	0.476
Slope	0.972	0.986	0.989	0.996
Y Intercept	0.205	0.048	0.070	0.031
R-squared	0.972	0.986	0.989	0.996
Std. Error Est.	0.409	0.365	0.434	0.424
Pearson's Corr.	0.986	0.993	0.994	0.998
RMSE	0.409	0.365	0.434	0.424

Table 4-2. PLS model statistics for the three drying experiments. All models used 2 factors. The standard error of prediction (SEP) value is from the cross validation studies.

The PLS model fit results are shown in Figure 4-1, Figure 4-2, and Figure 4-3 for Run_1, Run_2, and Run_3, respectively. For all three runs, the PLS model was able to accurately model the variation in the water concentration with no significant bias. However, for Run_1 there appears to be some non-modeled nonlinear relationship between drying time and water concentration. Also note that the Run_1 water values never reach a moisture concentration below about 4 wt%. For the Run_2 data, it is clear that the 10 minute drying time between sample points results in relatively even spacing along the moisture axis for the early samples but, as the sample became drier, the moisture values began to cluster near each other. In the final refined experiment, Run_3, the drying time was graduated with longer drying times as the sample material contained less moisture. This allowed us to cover the full range of moisture levels with the data points spread fairly evenly across the moisture scale. As we can see from Figure 4-3, this design objective was met in Run_3.

The regression coefficients for the three runs are plotted versus wavelength in Figure 4-4. From this plot we can see that Run_2 and Run_3 have very similar regression coefficients while the regression coefficient for the Run_1 model is quite different. In

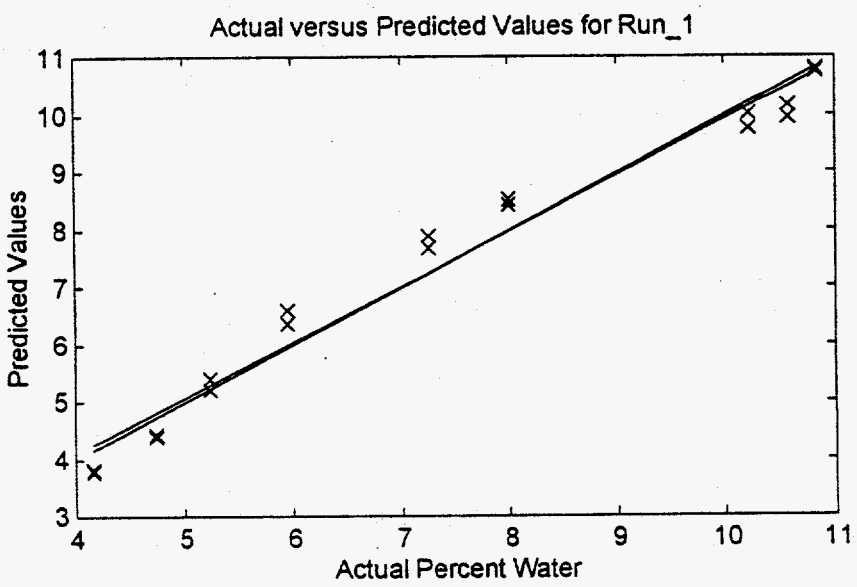


Figure 4-1. Actual versus estimated moisture for the Run_1 NIR data set.

order to explain this observation it is instructive to look at the individual PLS factor loadings for each model. Recall that the regression coefficient is a weighted linear combination of the individual model factors. The individual factor loadings for the X-block represent the portion of the spectral information used to derive the calibration model.

Figure 4-5 shows the first PLS model factor loadings for the spectra from each of the three runs. From this plot it is clear that the three models have almost identical loadings for the first spectral factor. Therefore, the observed difference in the regression coefficients cannot be due to this factor and must be due to differences in the spectral information associated with the second factor of each model as shown in Figure 4-6. From inspection of the score plots, and the information from Table 4-1, the first factor of each model describes the bulk of the spectral response related to the moisture variation in each of the experiments.

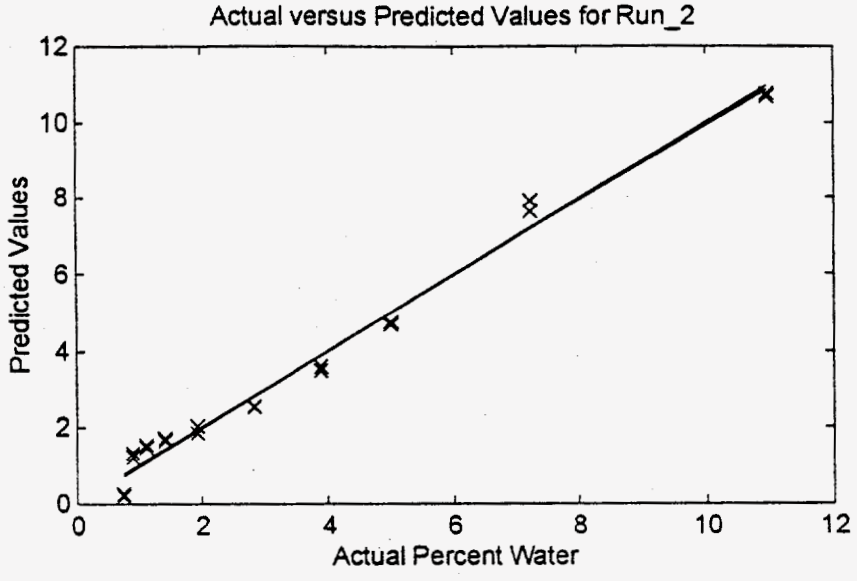


Figure 4-2. Actual versus estimated moisture for the Run_2 NIR data set.

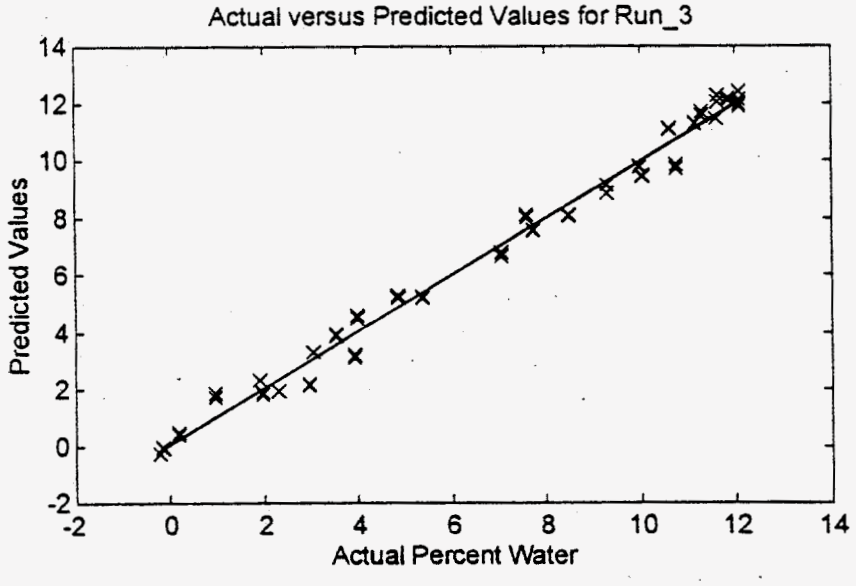


Figure 4-3. Actual versus estimated moisture for the Run_3 NIR data set.

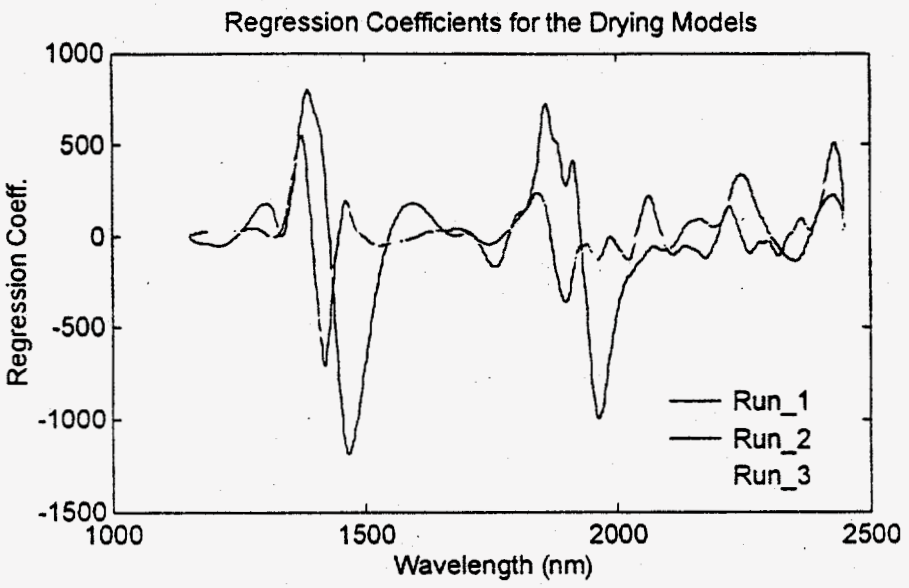


Figure 4-4. Regression coefficient vectors from the 2-factor NIR models for Run_1, Run_2, and Run_3.

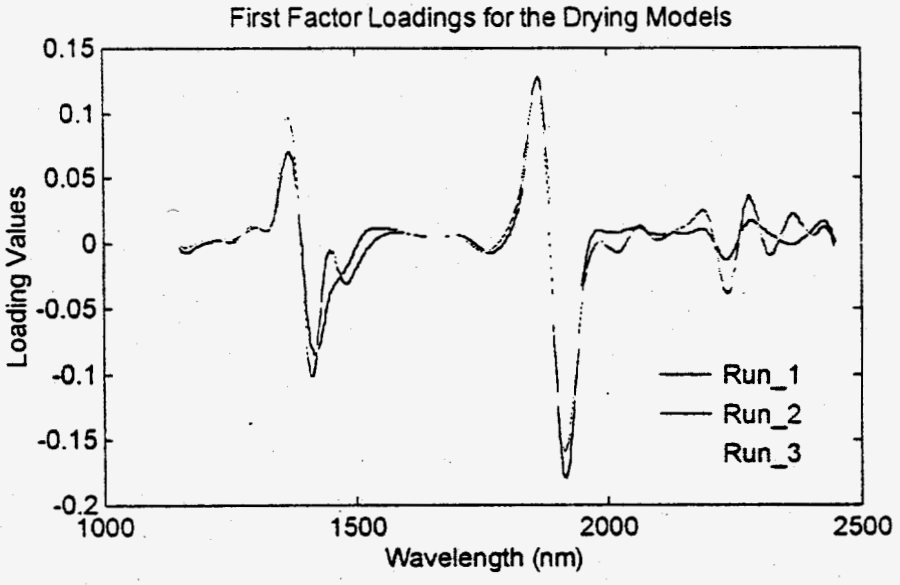


Figure 4-5. Spectral loading vectors for the first factor of the NIR models for Run_1, Run_2, and Run_3.

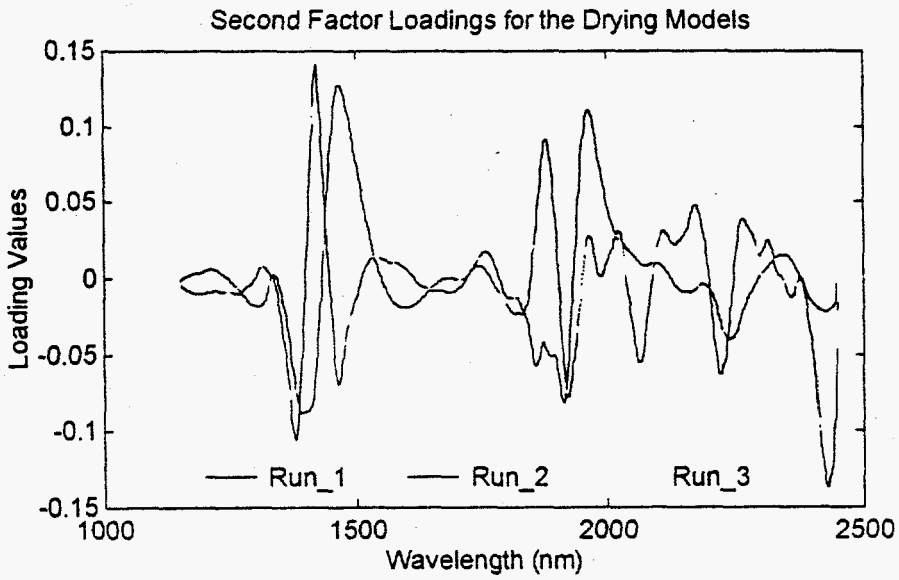


Figure 4-6. Spectral loading vectors for the second factor of the PLS models for Run_1, Run_2, and Run_3.

The second factor of the Run_1 model is only describing information related to the difference between the first sample point (the BY-104 before drying) and the third (after two 5 minute drying cycles, 8 wt% water). The second factor of the Run_2 and Run_3 models are describing additional spectral variation related to the very dry samples from those runs. Figure 4-7 and Figure 4-8 show the X-block (spectral) versus Y-block (wt% moisture) scores for the two factors of the Run_3 model. In Figure 4-7 one can see that the first model factor is describing the mostly linear response of the spectra to moisture decrease for the first 11 sample points (no drying to ~3 wt% moisture) while the last 4 sample points (~2 wt% to 0 wt%) are not well described by this factor. Remember that each sample point corresponds to two data points in the plot since replicate spectra were recorded for each drying cycle. In fact, as the sample is dried below ~4 wt% moisture, there seems to be curved, or exponential decay, component in the spectral response to increased drying of the sample. The second factor of the model is mainly describing the information associated with the very dry samples as is shown in Figure 4-8. The Run_2 model

shows nearly identical behavior, except that the higher moisture samples are not as well represented due to the constant drying interval in that experiment.

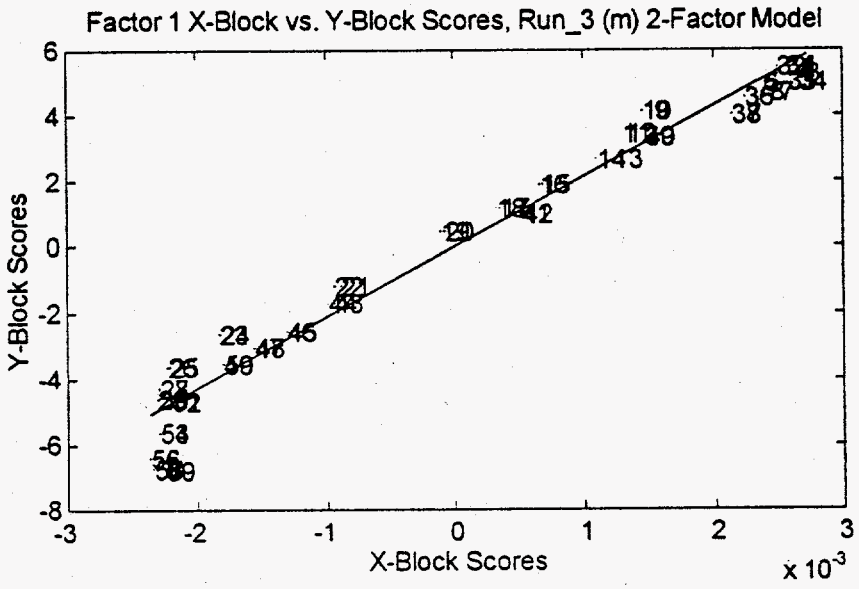


Figure 4-7. Relationship between the spectra and moisture explained by first PLS factor from the NIR Run_3 model.

By comparing the loadings of the second PLS model factor for Run_2 and Run_3, described above, with the spectra of the pure BY-104 chemical components one can make the following observation. This second factor loading vector is very similar to the second derivative NIR spectra of pure $NaNO_3$ plus the negative of the NaOH second derivative NIR spectra. This is shown graphically in Figure 4-9 where we have plotted the second factor loading vector for Run_2 and Run_3 along with a composite spectrum consisting of the pure sodium hydroxide and pure sodium nitrate spectra. The composite spectrum consists of the sodium nitrate spectra plus the negative of the sodium hydroxide spectra, scaled by an arbitrary 200 units.

Based on the discussion above, one can conclude that, as the BY-104 simulant is dried in the oven, the particle surface is enriched in $NaNO_3$ and depleted in NaOH.

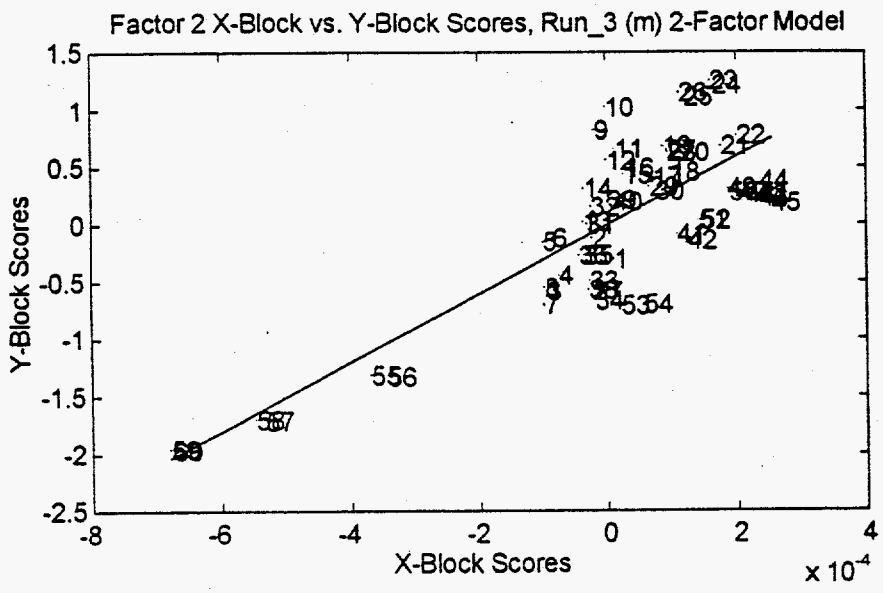


Figure 4-8. Relationship between the second NIR model PLS factor scores in the spectral responses and the moisture level for Run_3.

There is not sufficient information to speculate on the exact mechanism whereby the nitrate would migrate to the surface and the hydroxide becomes depleted. However, there is information to suggest that this mechanism is only active once the moisture level drops below a certain threshold level (roughly 2-3 wt%).

The last point to be considered in this study relates back to the question of which is the best method for preparing the calibration samples for the moisture model determination, the successive drying method or the water addition method. In order to evaluate this issue, moisture prediction models were developed from both the drying experiment data and the water addition data obtained in the Phase 1 study. We then looked at the relative performance of these models at predicting the moisture from samples prepared with the other sample treatment. The goal is to discover if one, or both, of the sample preparation methods introduce appreciable bias in the moisture estimation. The results of these studies, in terms of the RMSE of prediction are summarized in Table 4-3 below. PLS models from the full length

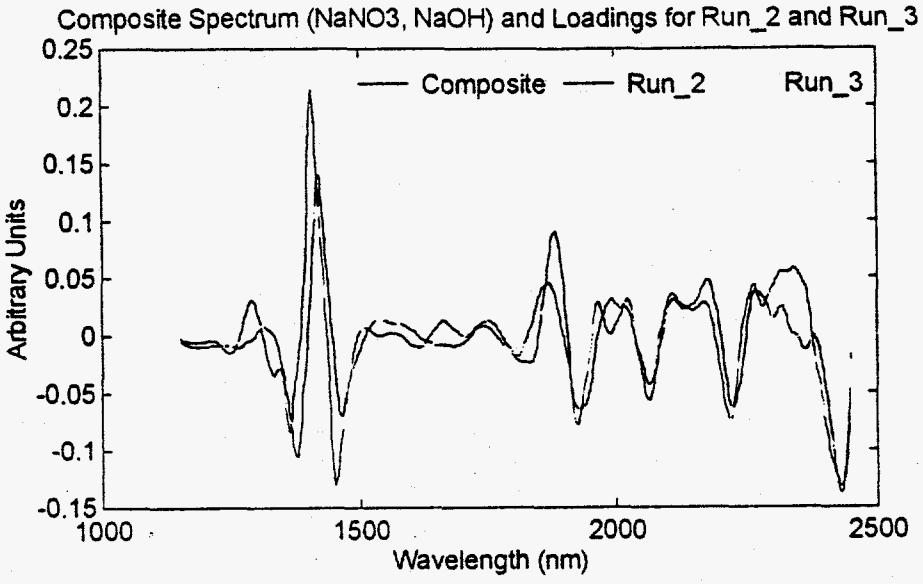


Figure 4-9. Run_2 and Run_3 PLS model spectral loadings for second factor and a composite spectrum made from the pure spectra of NaOH and $NaNO_3$.

columns were used to predict moisture for the data sets in columns 3, 4, 5, and 6. The boldface values are the model fit results (self prediction).

Although the results of the cross prediction experiments are somewhat confusing, several generalizations are possible. It is clear that the Run_1 model is not very useful at predicting either the other drying runs or the water addition data. Conversely, none of the other models do a very good job of predicting the Run_1 moisture levels from the Run_1 spectra. In fact the predicted moisture for the Run_1 data is consistently higher than the actual moisture values calculated from the before and after drying weight values, as shown in Figure 4-10 below. Only the Phase 1 samples with moisture values in the range covered by the drying experiments (zero to ~14 wt%) are plotted. This is probably due to the short drying times and the fact that the sample remained in the sample cup throughout the Run_1 experiment. The calculated reference moisture values for this run are based on the total weight

	Prediction Data Set			
Model	Run_1	Run_2	Run_3	Phase 1
Run_1				
1 Factor	0.862	1.334	1.636	1.367
2 Factor	*0.409*	1.162	1.615	3.347
Run_2				
1 Factor	2.845	0.684	0.900	1.391
2 Factor	2.493	*0.365*	0.597	0.999
Run_3				
1 Factor	2.084	0.806	0.612	0.909
2 Factor	1.998	0.571	*0.313*	0.983
Phase 1				
1 Factor	1.660	1.478	1.007	*0.366*

Table 4-3. Comparison of the RMSE for different sample treatments.

difference before and after drying of the bulk sample in the cup. However, since the sample cup is closed next to the spectral window, it is conceivable that the portion of the sample material actually sampled spectroscopically during the experiment has a higher moisture level than the material at the other side of the sample cup (which is open to the oven environment).

The Run_2 and Run_3 data sets agree very well with each other. While the model errors for each of these runs were about 0.3 wt%, the errors for predicting moisture from one run using the model from the other run was only slightly higher (about 0.6 wt%). Furthermore, both Run_2 and Run_3 models had similar performance when predicting moisture from the Phase 1 data (the water addition to dried simulant method). For both these runs, the prediction errors for the Phase 1 data were about one wt% compared to about 0.4 wt% for the Phase 1 model fit value. Note, from Figure 4-10 above that the Phase 1 predicted moisture values are very close over the range from around 1 to 5 wt% moisture but then are biased low for the higher and lower moisture levels. This may be due mainly to the specific information related to

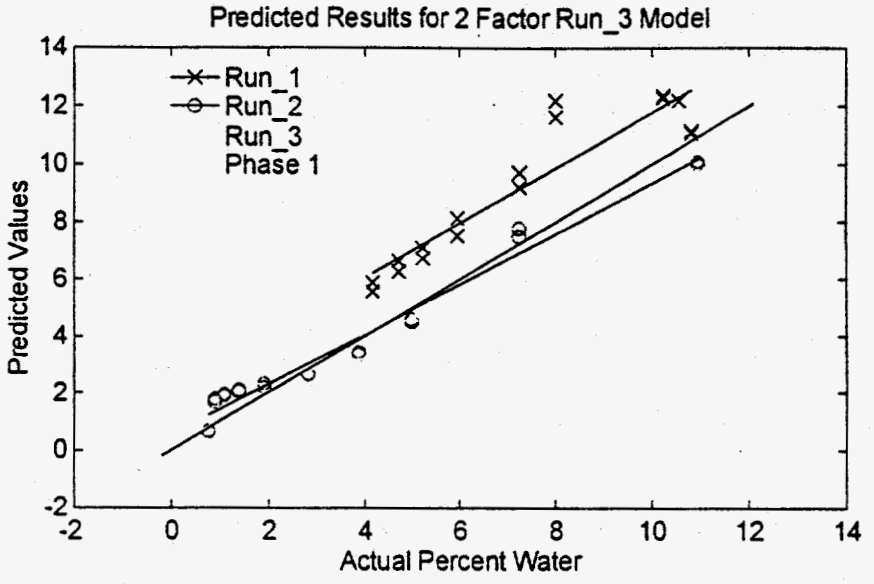


Figure 4-10. Prediction results for Run_1, Run_2, Run_3, and Phase 1 using the 2 factor Run_3 PLS model.

dry samples from Run_2 and Run_3 contained in the second factor of the model.

Finally, the Phase 1 model (derived from the water addition sample preparation method) did rather poorly at predicting moisture from the drying experiment spectra. There are several factors which might lead to this result. First, the Phase 1 data was collected several months before the drying experiment data and several modifications to the spectrometer were made during that time, including replacement of the spectrometer lamp and detection electronics. So, some of the differences may be due to instrumental artifacts. Secondly, from the discussion above, it seems clear that as the sample is dried, there is selective segregation of some components between the bulk and surface of the simulant. This was seen quite clearly for the very dry samples from Run_2 and Run_3 in the second factor of the PLS models for those runs. Since the Phase 1 starting material was also oven dried to complete dryness, it is reasonable to assume that the same segregation also took place. Indeed, in Figure 4-11 below, the dry samples from Run_2, Run_3, and Phase 1 all have similar

predicted moisture levels. Then, as water was added, the enriched nitrate and depleted hydroxide layer was dissolved, altering the composition of the sample slightly. Again, from the Figure below, we see the largest prediction errors for those samples in the 0-2 wt % range as a constant offset, or bias, from the Phase 1 prediction line. At any rate, the Phase 1 one factor PLS model contains no term to explicitly model this segregation. At the higher moisture levels, the Phase 1 and the drying data from Run_2 and Run_3 again start to converge. Again, for the Phase 1 model, the largest errors in the prediction of the Run_1 moisture occur as high bias for the early drying period samples. As mentioned above, this is probably due to a moisture gradient present in the Run_1 sample as it is being dried.

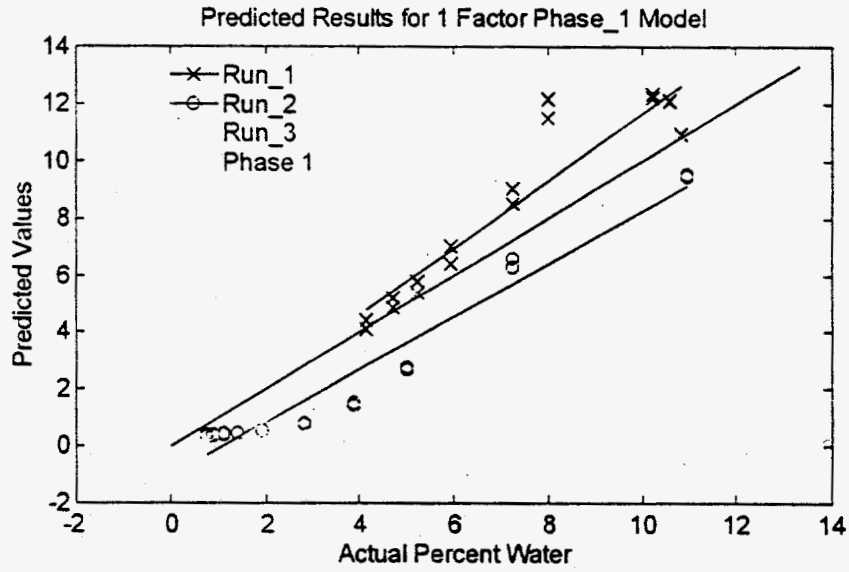


Figure 4-11. Prediction results for Run_1, Run_2, Run_3, and Phase 1 using the 1 factor Phase 1 PLS model.

Visible Spectral Region Results

The results for the individual PLS models on the VIS (400 to 1100 nm) spectral data for the three drying runs are given in Table 4-4 below. The number of factors in the

model was determined by cross validation studies. The Run_1 model required three factors for the VIS data compared to only two for the NIR model. This was due to a stronger nonlinear response of the VIS data to the moisture. This nonlinear response was even stronger in the Run_3 data where four factor were required.

	X-Block (Spectra)		Y-Block (Moisture)	
Run_1				
Factor #	This Factor	Total	This Factor	Total
1	75.606	75.606	79.649	76.649
2	17.915	93.520	7.666	87.315
3	1.529	95.050	9.545	96.860
Run_2				
Factor #	This Factor	Total	This Factor	Total
1	97.574	97.574	83.340	83.340
2	1.137	98.710	6.884	95.224
Run_3				
Factor #	This Factor	Total	This Factor	Total
1	95.638	95.638	73.627	73.627
2	2.368	98.006	15.192	88.818
3	0.633	98.639	8.778	97.594
4	0.293	98.931	0.775	98.576

Table 4-4. Percent variance described by each factor of the VIS PLS models for the three successive drying experiments.

All three PLS models gave fit results which were very similar to the NIR results, as shown in Table 4-5. The RMSE of the model fit was about half a weight percent moisture for all three models. The RMSE for thePhase 1 VIS data was higher (1.327 wt% moisture) when the full range of moisture concentration was included in the model. When the Phase 1 model was restricted to the moisture range covered in the drying experiments (0 to ~14 wt% moisture) the RMSE was only 0.501 wt%, which agrees quite well with the values from the drying experiments.

	Run_1	Run_2	Run_3	Phase 1
SEP (cross val.)	0.641	1.020	0.569	1.496
Slope	0.969	0.952	0.986	0.967
Y Intercept	0.234	1.172	0.094	0.303
R-squared	0.969	0.952	0.986	0.967
Std. Error Est.	0.436	0.689	0.501	1.327
Pearson's Corr.	0.984	0.976	0.993	0.983
RMSE	0.436	0.689	0.501	1.327

Table 4-5. PLS model statistics for the three VIS drying experiments. The standard error of prediction (SEP) value is from the cross validation studies.

	Prediction Data Set			
Model	Run_1	Run_2	Run_3	Phase 1
Run_1				
1 Factor	1.110	13.443	6.797	11.136
3 Factor	*0.436*	10.613	6.834	32.401
Run_2				
1 Factor	2.548	1.076	2.351	3.876
2 Factor	2.077	*0.689*	3.517	2.041
Run_3				
1 Factor	2.548	1.182	2.118	5.492
4 Factor	2.123	1.423	*0.501*	8.227
Phase 1				
1 Factor	4.802	4.213	5.324	1.445
2 Factor	5.939	4.631	6.315	*0.501*

Table 4-6. Comparison of the RMSE for different sample treatments. PLS models from column 1 were used to predict moisture for the data sets in columns 3, 4, 5, and 6.

However, when the PLS models from the individual drying runs were used to predict moisture from the spectral data of other runs, or the Phase 1 spectra, the results are

quite poor as shown in Table 4-6. Results in Table 4-6 are presented for both the cross validated models and for models using a single factor. The italicized values in the table are the model fit results (self prediction). The Phase 1 data and model was restricted to the concentration range covered by the drying experiments. As in the NIR results discussed above, the Run_1 model did the worst at predicting moisture from the other run's spectra. Again, this is most likely due to the limited moisture range and problems with moisture homogeneity present in the Run_1 experiment as mentioned in the discussion of the NIR results.

None of the drying run models did a very good job of predicting moisture from the Phase 1 data. An example of the cross prediction results are shown in Figure 4-12 for the four factor Run_3 model. Conversely, the Phase 1 model did a very poor job of predicting moisture for any of the three drying run's spectra as shown in Figure 4-13 below. The reason for this poor cross prediction performance can easily be seen from the plot shown in Figure 4-14 where the average spectra from the Phase 1 and the Run_3 data sets are plotted. In this plot, one can clearly see that the Phase 1 VIS spectra has strong spectral features from about 600 to 750 nm which is not present in the Run_3 data. The Run_1 and Run_2 spectra are very similar to the Run_3 spectra and also lack the absorbance in this region. Since the first loading vector of the calibration models, shown in Figure 4-15 for the Run_3 and Phase 1 models, incorporate information from this spectral region into the PLS models, the presence of the absorption in the Phase 1 data which are not present in the drying spectra introduces serious bias in the cross prediction results.

It is not clear what the spectral features in the 600 to 750 nm region of the Phase 1 spectra are due to. The only pure component with a strong adsorption in this region is the ferric nitrate and its spectral features in the second derivative spectra do not match that seen in the Phase 1 spectra (i.e. it has a smooth uni-modal peak whereas the Phase 1 spectra show features with three lobes). It should be noted that the

same spectral features in the 600 - 700 nm region were also observed in the VIS spectra from the particle size study. The absence of these features in the VIS spectra from the drying experiments is unexplained. Unfortunately, the instrument used to collect this data has since been re-configured with a fiber optic probe sampling system and so it would be difficult to go back and repeat these experiments and try to track down the source of this variation.

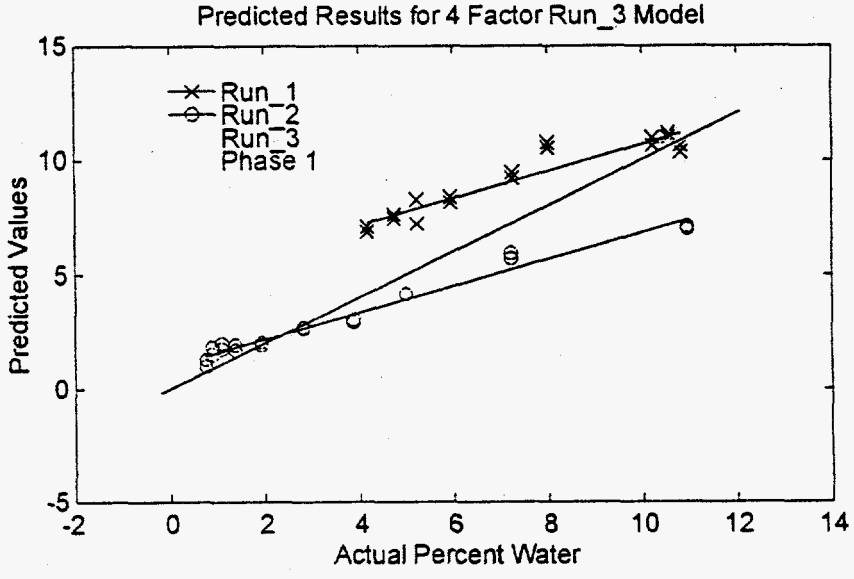


Figure 4-12. Prediction results for Run_1, Run_2, Run_3, and Phase 1 using the 4 factor Run_3 PLS model.

Conclusions

The main goal of this study was to evaluate two different sample preparation methods; water addition and successive drying. One of the issues was whether sample drying introduced changes in the spectral response that would effect the calibration for moisture. For the NIR results, we clearly saw that drying the sample

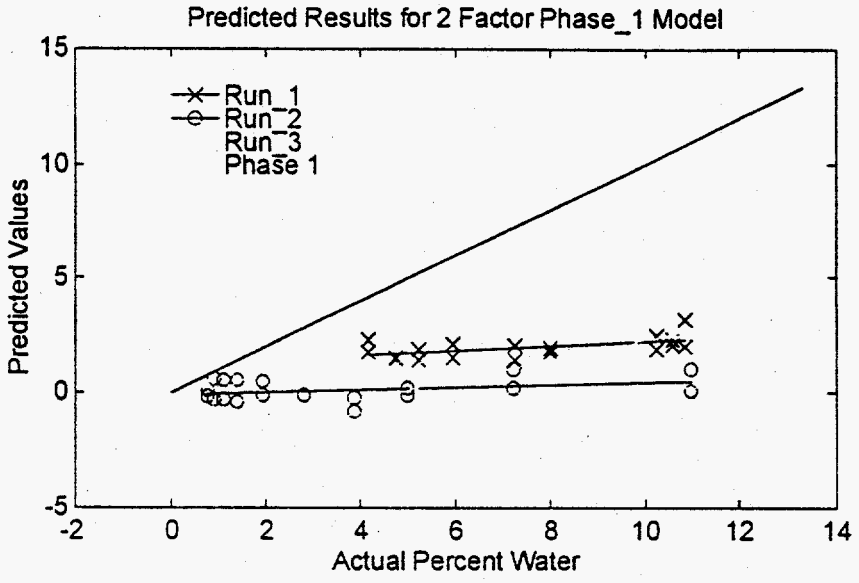


Figure 4-13. Prediction results for Run_1, Run_2, Run_3, and Phase 1 using the 2 factor Phase 1 PLS model.

below approximately two weight percent water induced a change in the spectral response. Based on our analysis we attributed this change to the surface enrichment in sodium nitrate and a depletion of sodium hydroxide as the sample became fully dehydrated. For the VIS spectral region, we were unable to satisfactorily answer this question since the recorded spectra appeared different from our earlier observations, even without the oven drying. This difference was very obvious in the 600 to 750 nm region of the spectra and is currently unexplained.

The second issue addressed by this study was the relative performance of the calibration models obtained with the different sample treatments. For the individual calibration models, there does not seem to be any advantage to either sample preparation technique. For the NIR spectra, both sample treatments gave very similar results in terms of the model fit to the calibration data. The drying models were able to predict moisture from the water addition spectra than the better reverse case (using the water addition model to predict moisture from the drying data).

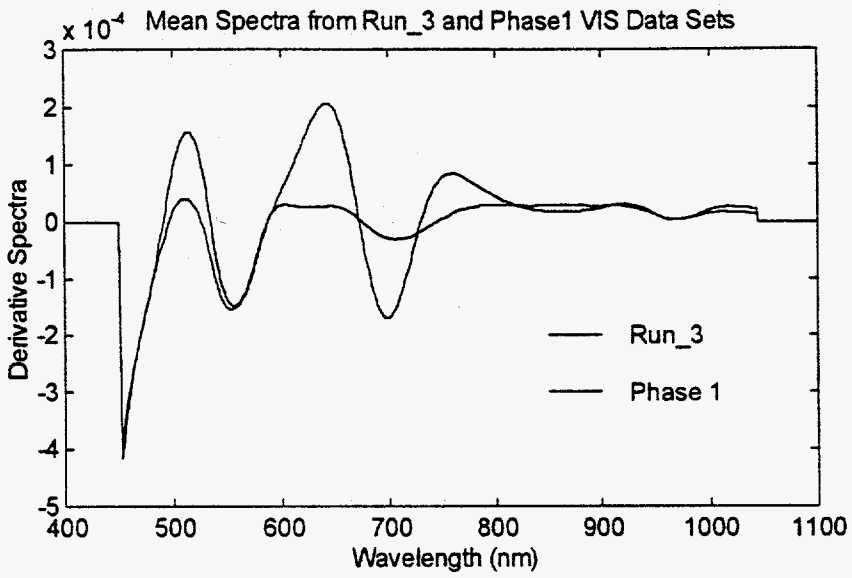


Figure 4-14. Average spectra from the Run_3 and Phase 1 data sets.

However, much of that difference can be attributed to the poor predictions for the low moisture samples from the drying experiments. Since those samples exhibited unique spectral features, related to the nitrate migration, not present in the water addition data, it is difficult to say one method is better than the other.

For the VIS data, the individual models derived from the three drying runs were able to fit the data quite well. However, they were not very good at predicting moisture from the spectral responses of a different drying run. Part of the poor cross prediction performance can be attributed to local overfitting of nonlinear response present in Run_1 and Run_3. The two sample preparation methods yielded models which were unable to satisfactorily predict the moisture content from the spectra of the sample prepared using the other method. However, most of the prediction errors can be explained as arising from the spectral differences discussed above which are mostly independent of the sample treatment.

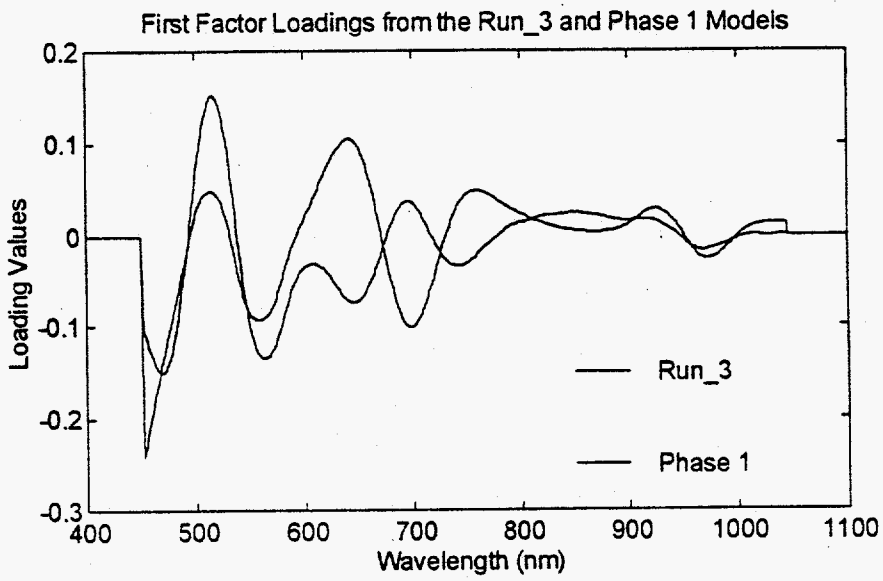


Figure 4-15. First PLS factor loadings from the Run_3 and Phase 1 models.

Based on the results from this study, we recommend that the standard additions of water to dried simulant be the preferred method for calibration sample preparation. There are several reasons why this method appears to be the best choice. First, it is faster and more flexible. Secondly, it does not introduce the artifacts in the very low moisture samples that the oven drying method does. Thirdly, while both methods produce suitable samples for developing good calibration models, the water addition models are simpler and slightly more robust when applied to other data sets.

Appendix A: Pure Component Spectra

Visible Spectra of Pure BY-104 Components

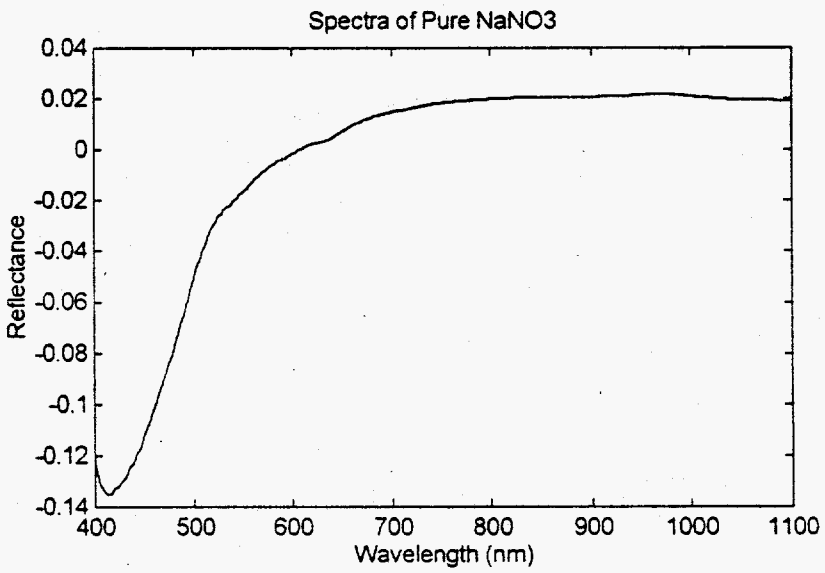


Figure A-1. VIS spectrum of $NaNO_3$

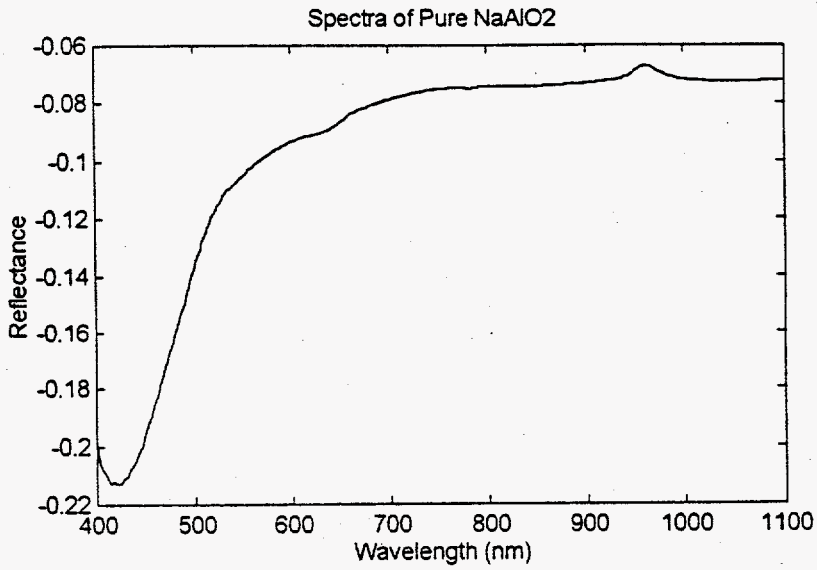


Figure A-2. VIS spectrum of $NaAlO_2$

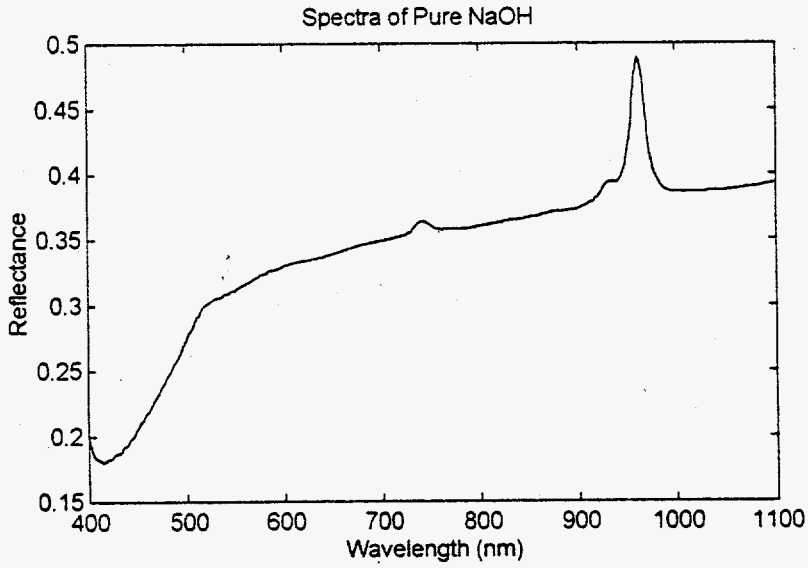


Figure A-3. VIS spectrum of NaOH

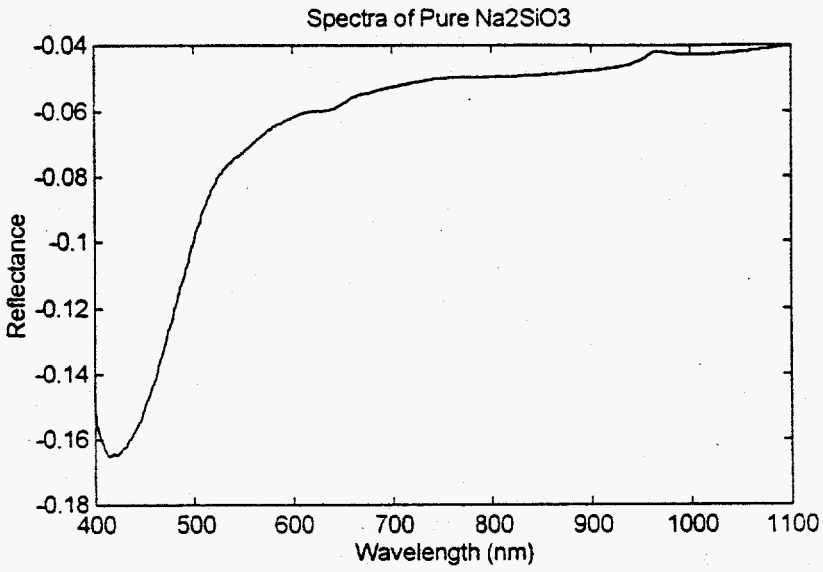


Figure A-4. VIS spectrum of Na_2SiO_3

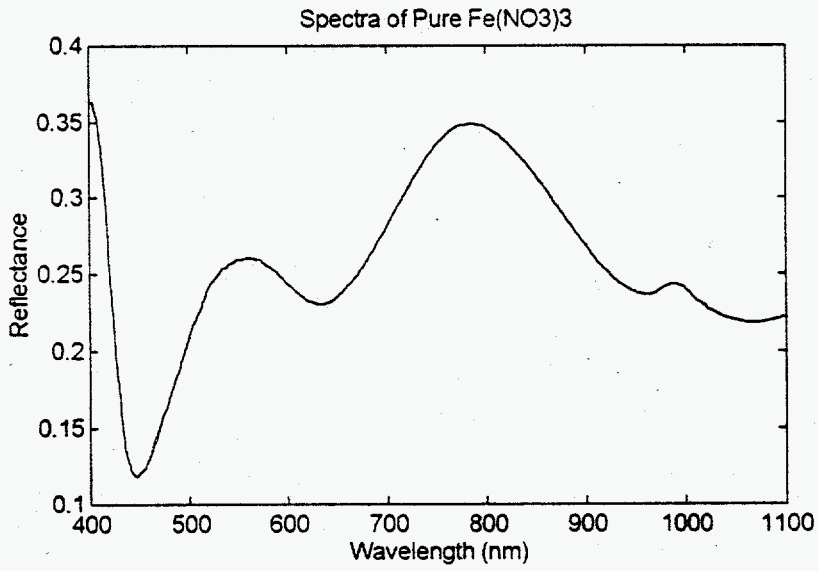


Figure A-5. VIS spectrum of $Fe(NO_3)_3$

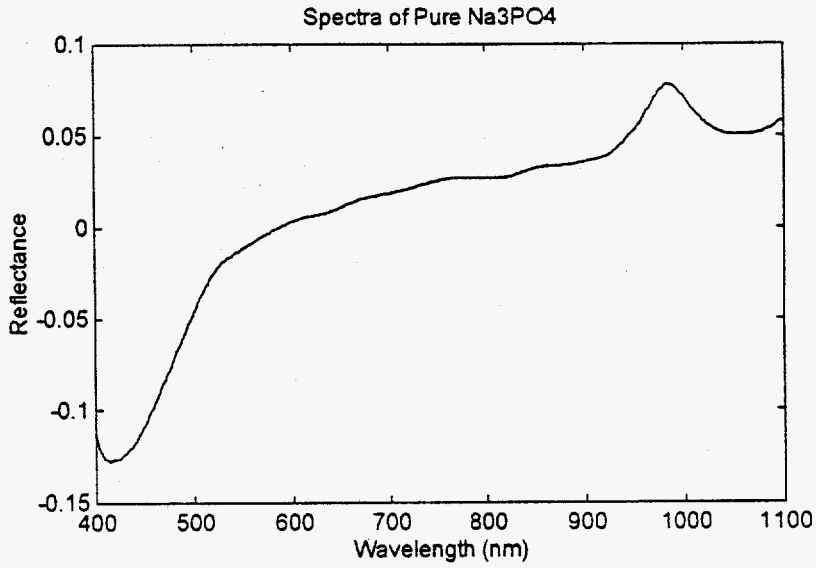


Figure A-6. VIS spectrum of Na_3PO_4

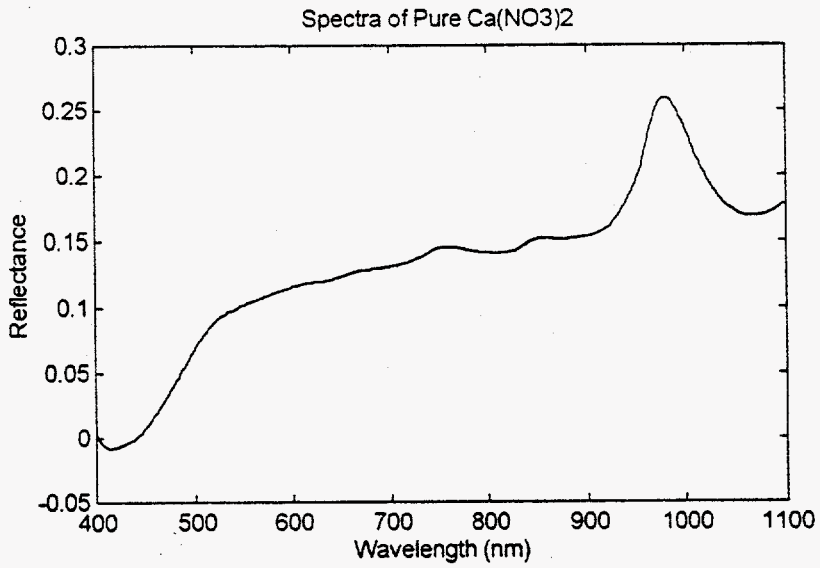


Figure A-7. VIS spectrum of $Ca(NO_3)_2$

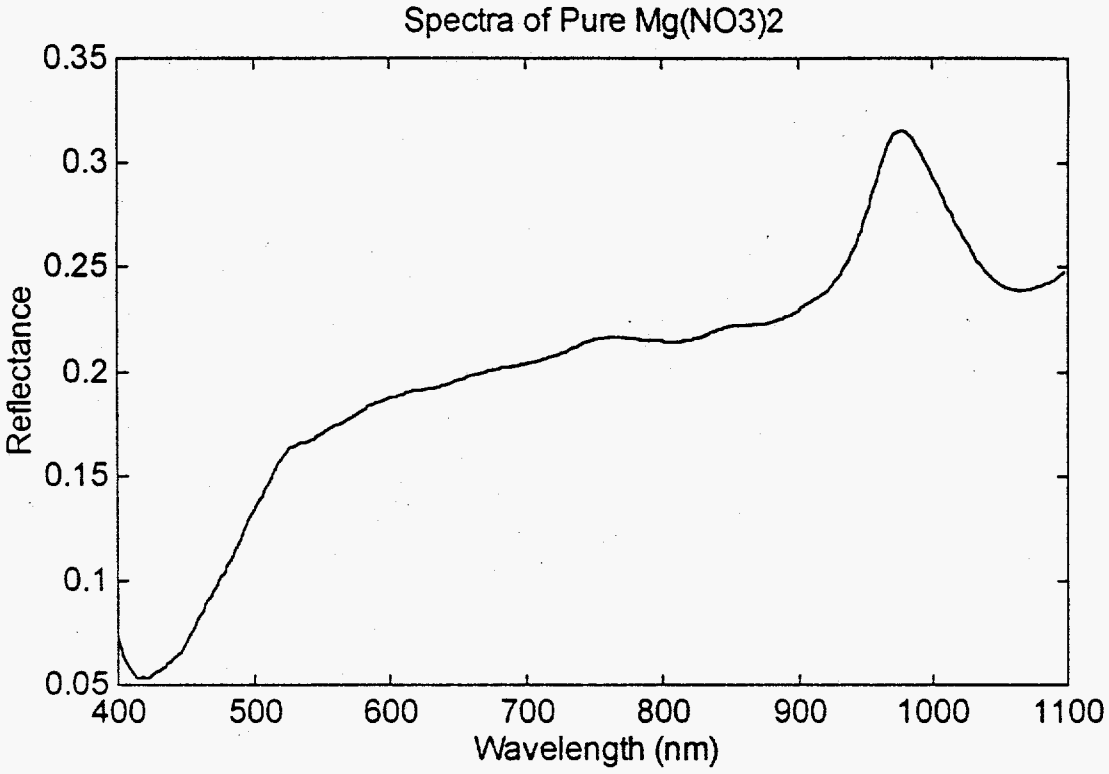


Figure A-8. VIS spectrum of $Mg(NO_3)_2$

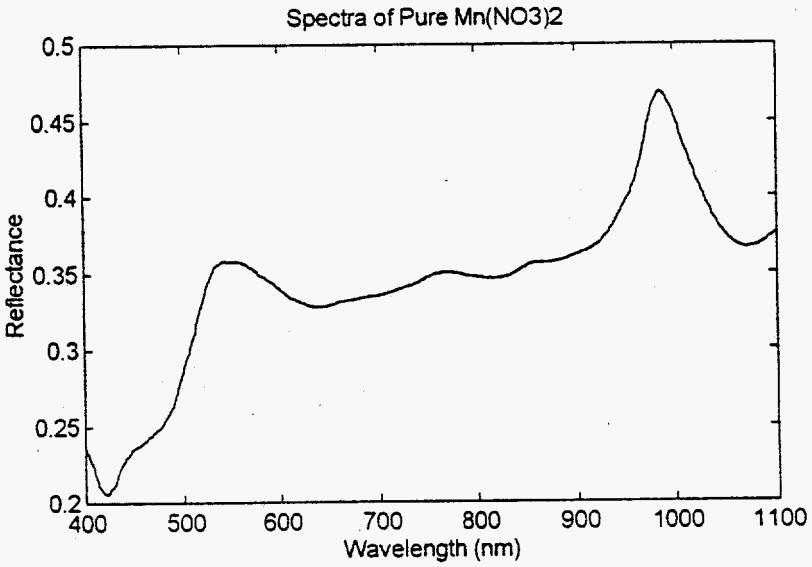


Figure A-9. VIS spectrum of $Mn(NO_3)_2$

Visible Derivative Spectra of Pure BY-104 Components

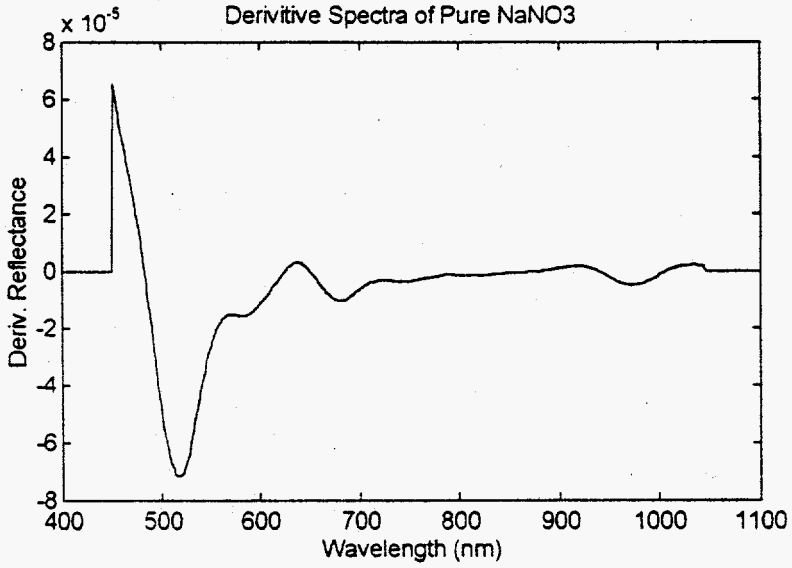


Figure A-10. Second derivative VIS spectrum of $NaNO_3$

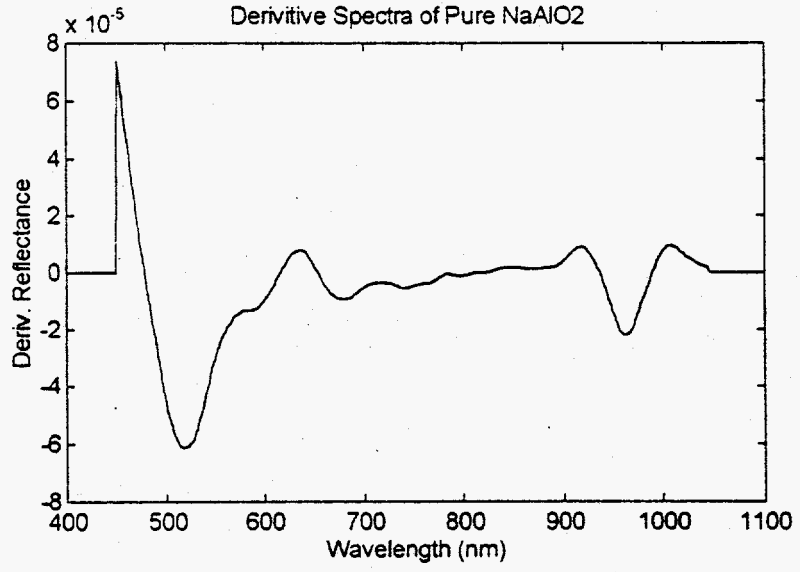


Figure A-11. Second derivative VIS spectrum of $NaAlO_2$

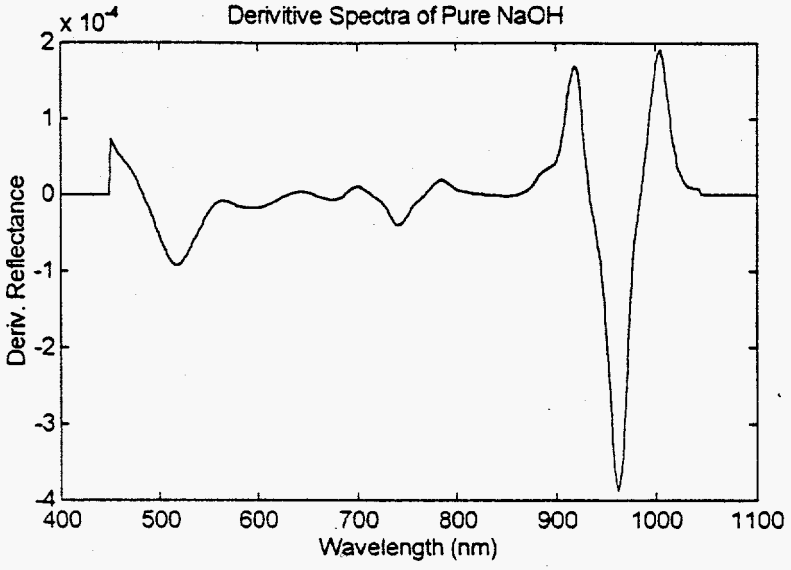


Figure A-12. Second derivative VIS spectrum of NaOH

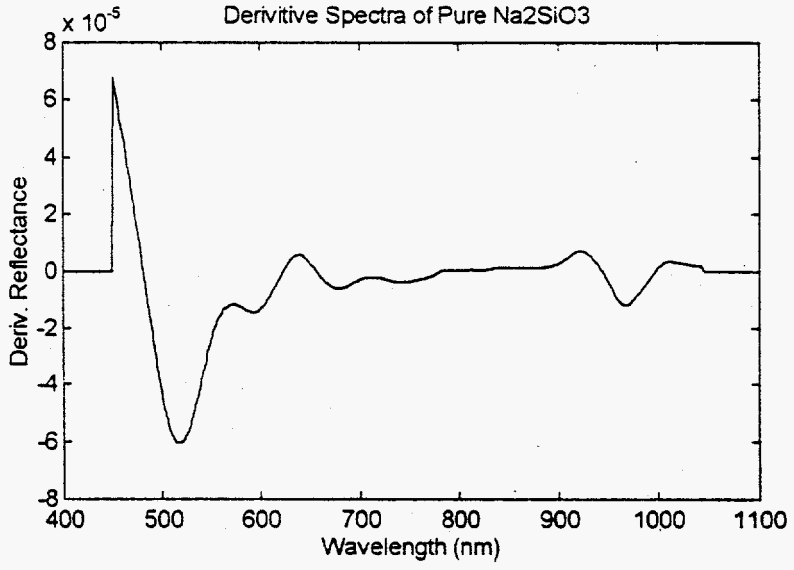


Figure A-13. Second derivative VIS spectrum of Na_2SiO_3

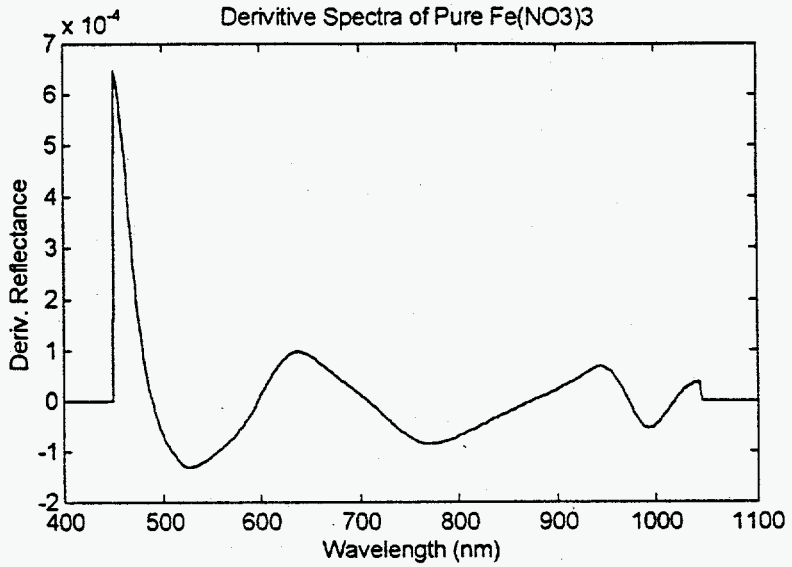


Figure A-14. Second derivative VIS spectrum of $Fe(NO_3)_3$

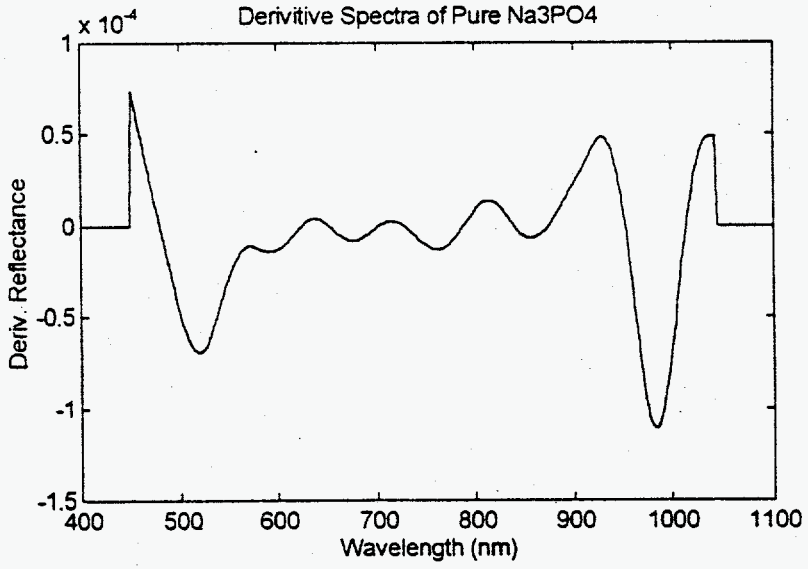


Figure A-15. Second derivative VIS spectrum of Na_2PO_4

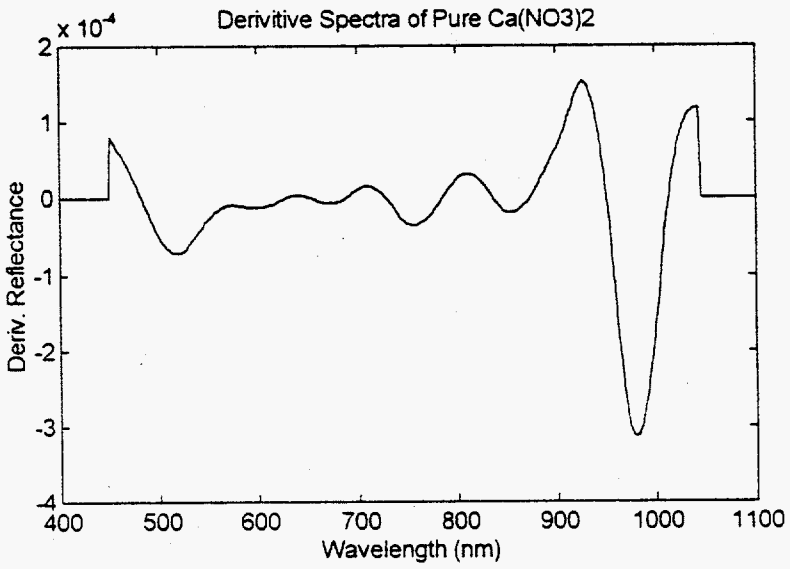


Figure A-16. Second derivative VIS spectrum of $Ca(NO_3)_2$

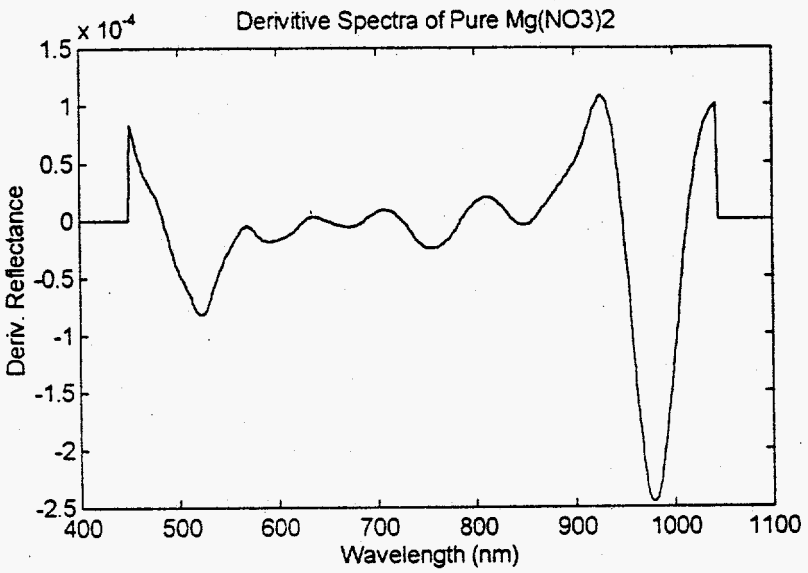


Figure A-17. Second derivative VIS spectrum of $Mg(NO_3)_2$

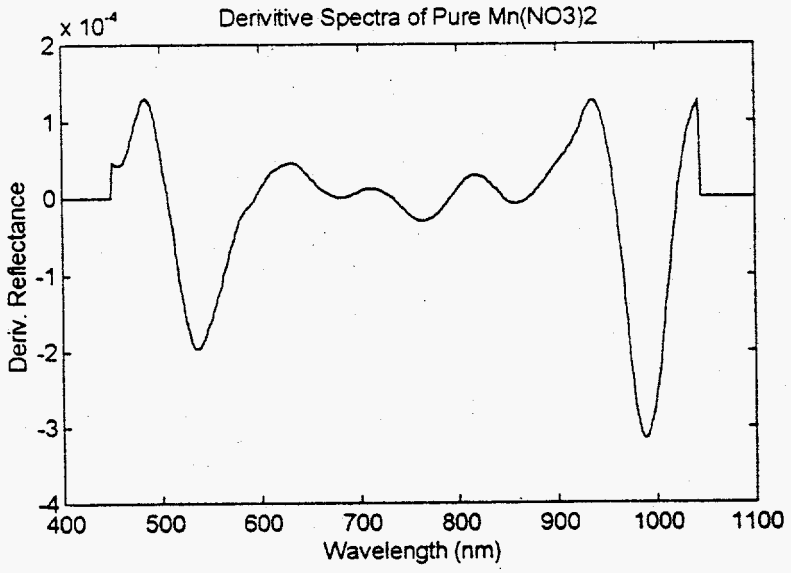


Figure A-18. Second derivative VIS spectrum of $Mn(NO_3)_2$

NIR Spectra of Pure BY-104 Components

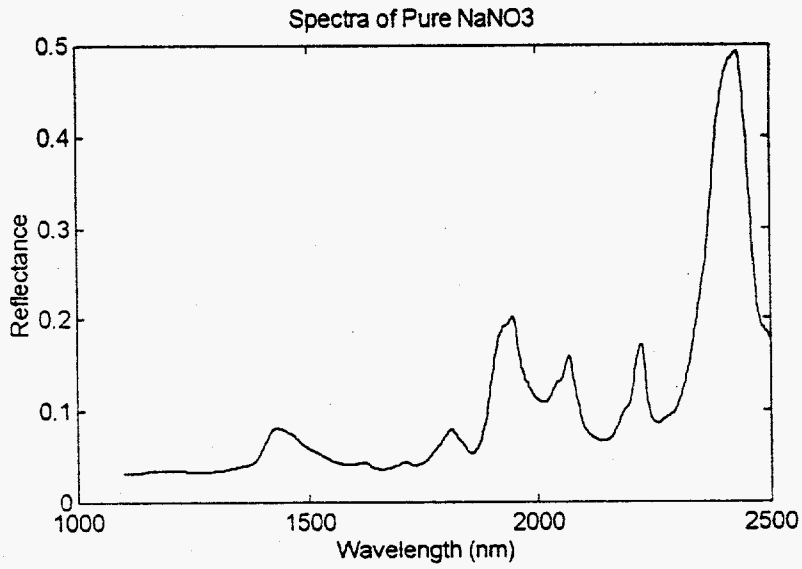


Figure A-19. NIR spectrum of $NaNO_3$

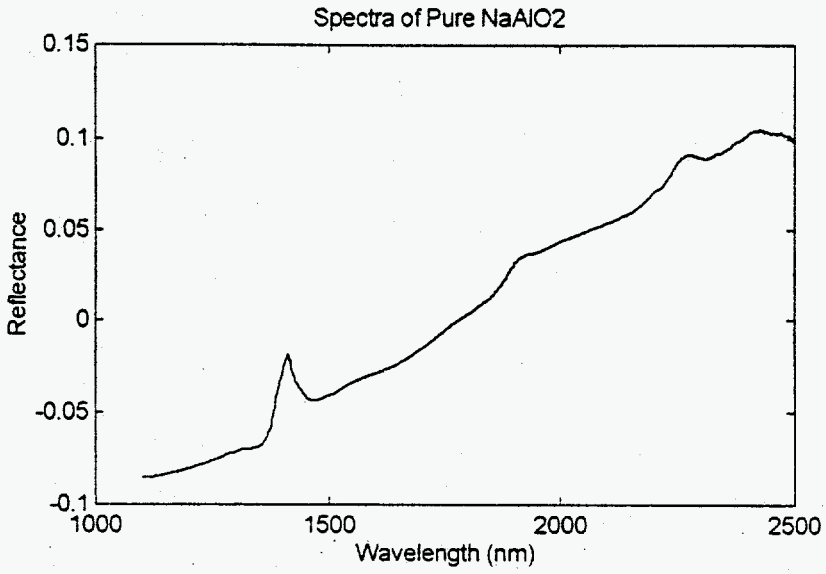


Figure A-20. NIR spectrum of $NaAlO_2$

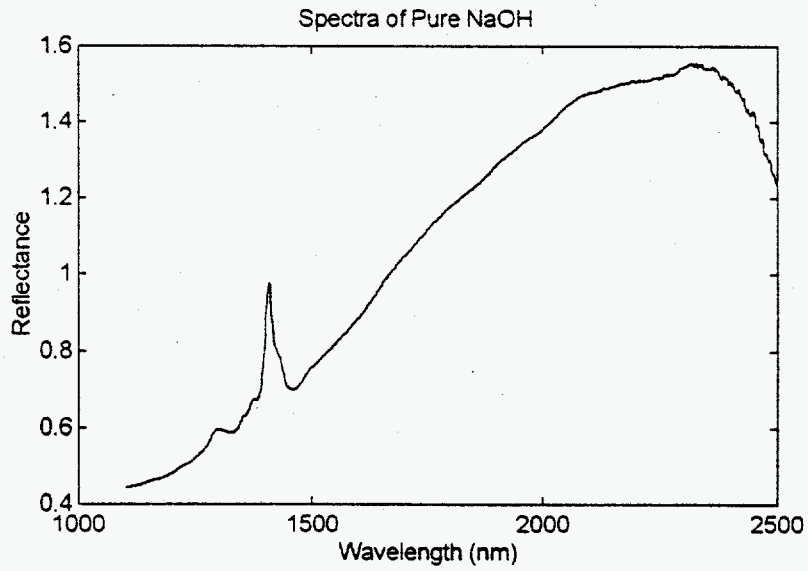


Figure A-21. NIR spectrum of NaOH

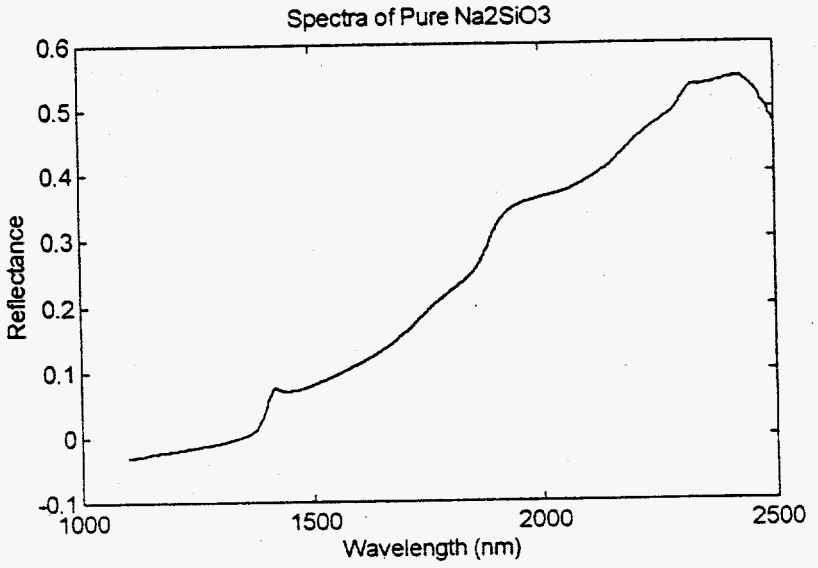


Figure A-22. NIR spectrum of Na_2SiO_3

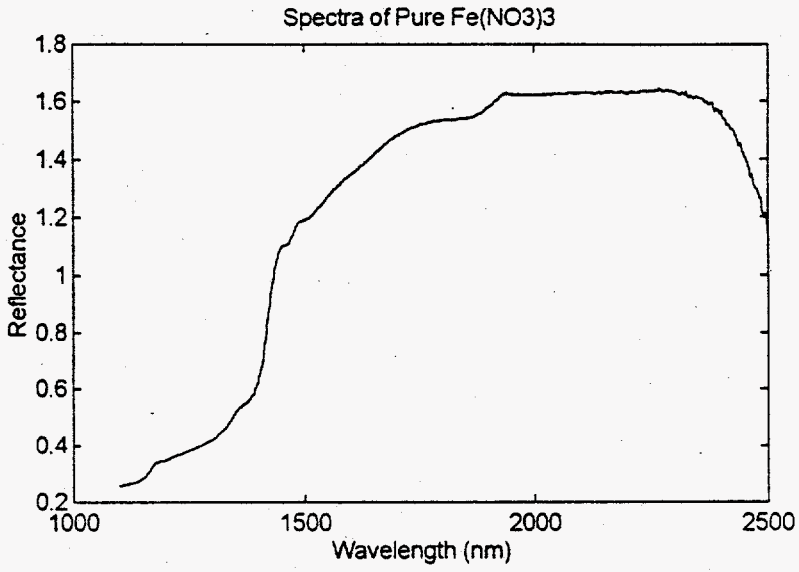


Figure A-23. NIR spectrum of $Fe(NO_3)_3$

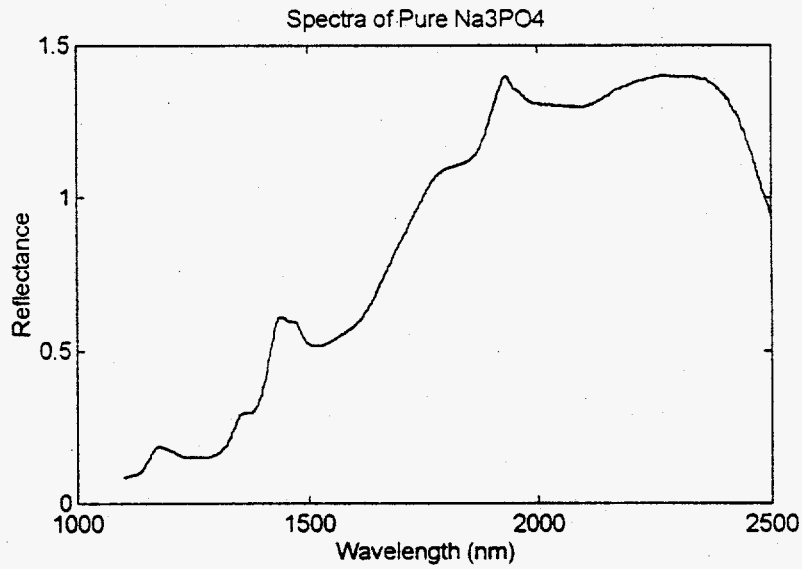


Figure A-24. NIR spectrum of Na_2PO_3

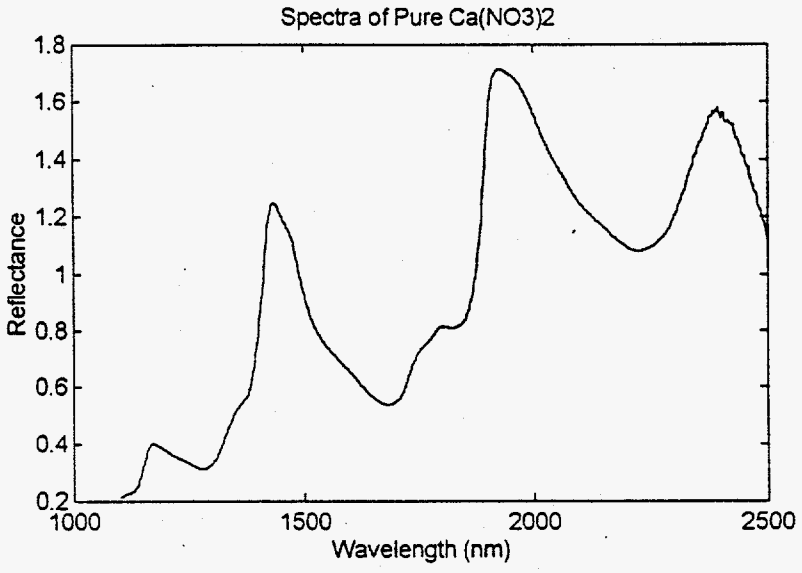


Figure A-25. NIR spectrum of $Ca(NO_3)_2$

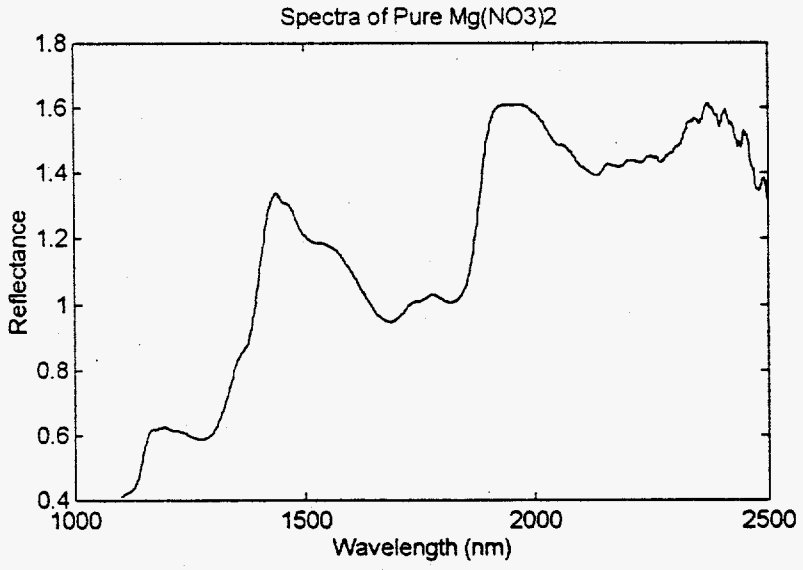


Figure A-26. NIR spectrum of $Mg(NO_3)_2$

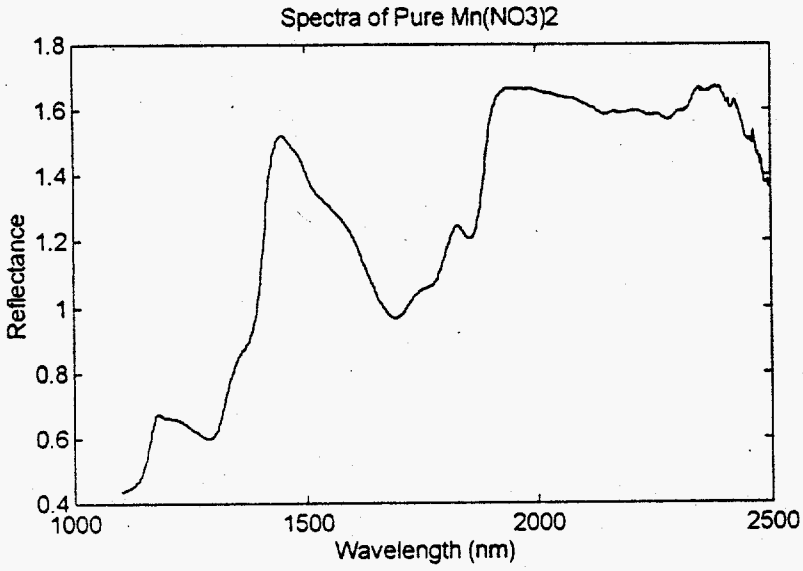


Figure A-27. NIR spectrum of $Mn(NO_3)_2$

Second Derivative NIR Spectra of BY-104 Components

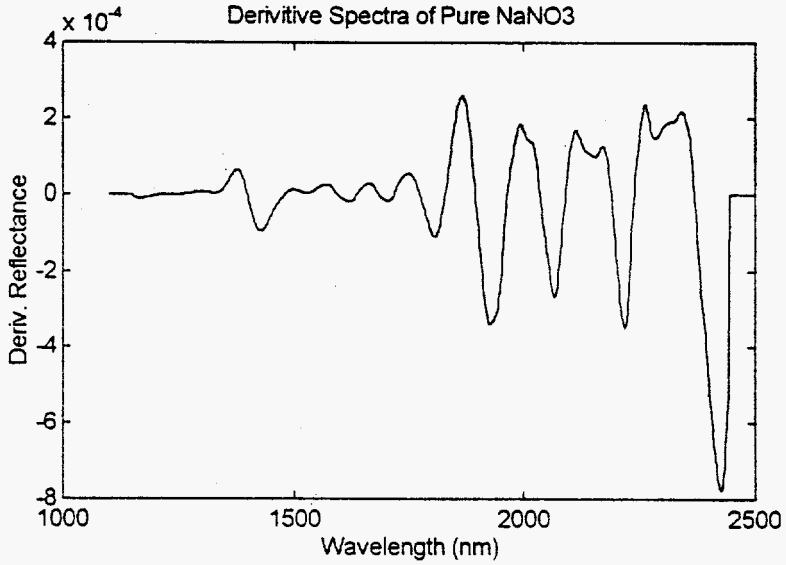


Figure A-28. Second derivative NIR spectrum of $NaNO_3$

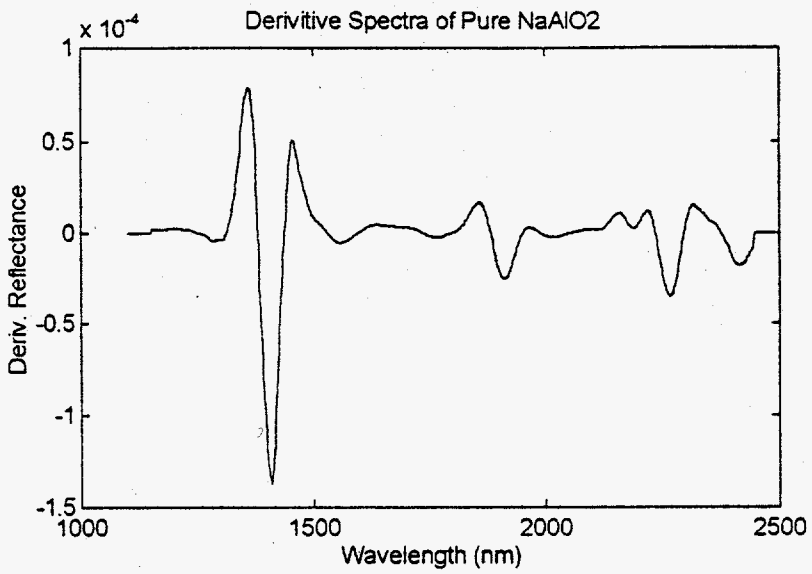


Figure A-29. Second derivative NIR spectrum of $NaAlO_2$

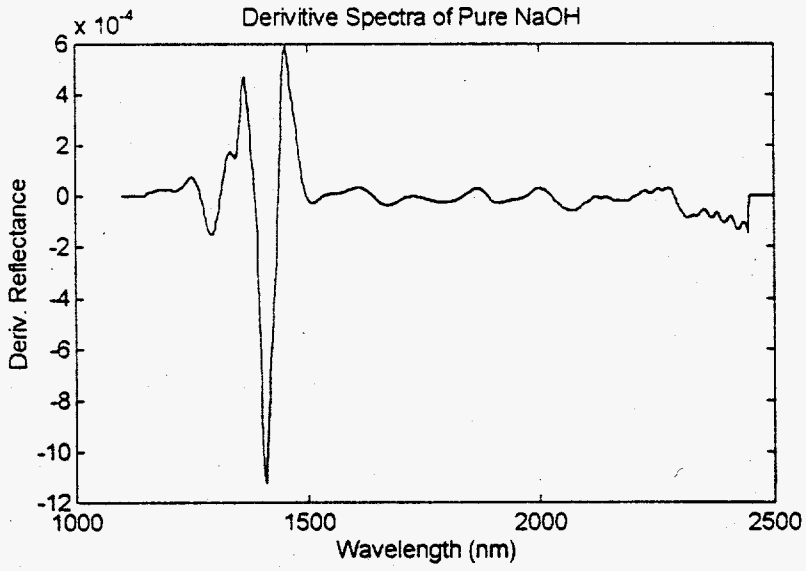


Figure A-30. Second derivative NIR spectrum of NaOH

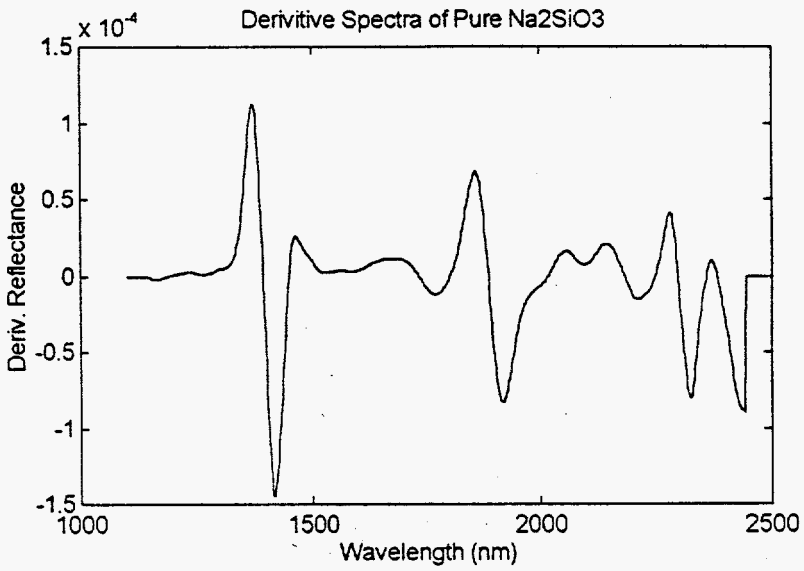


Figure A-31. Second derivative NIR spectrum of Na_2SiO_3

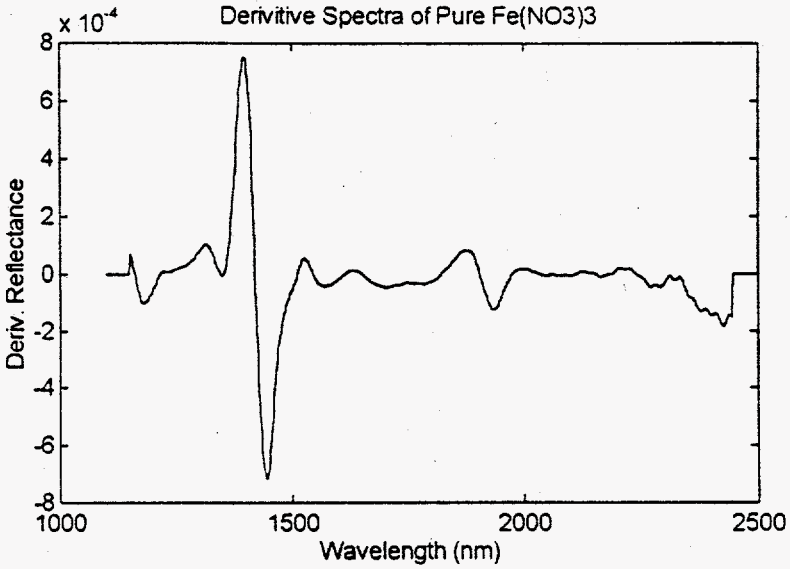


Figure A-32. Second derivative NIR spectrum of $Fe(NO_3)_3$

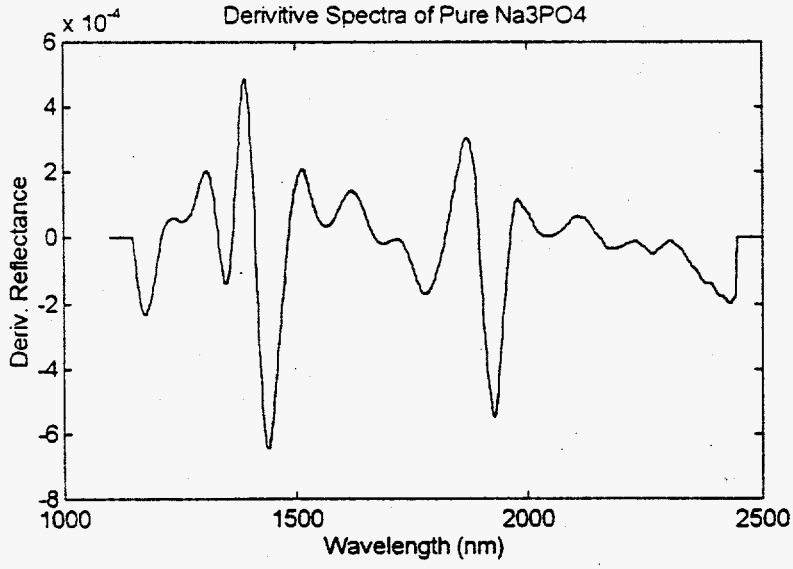


Figure A-33. Second derivative NIR spectrum of Na_2PO_4

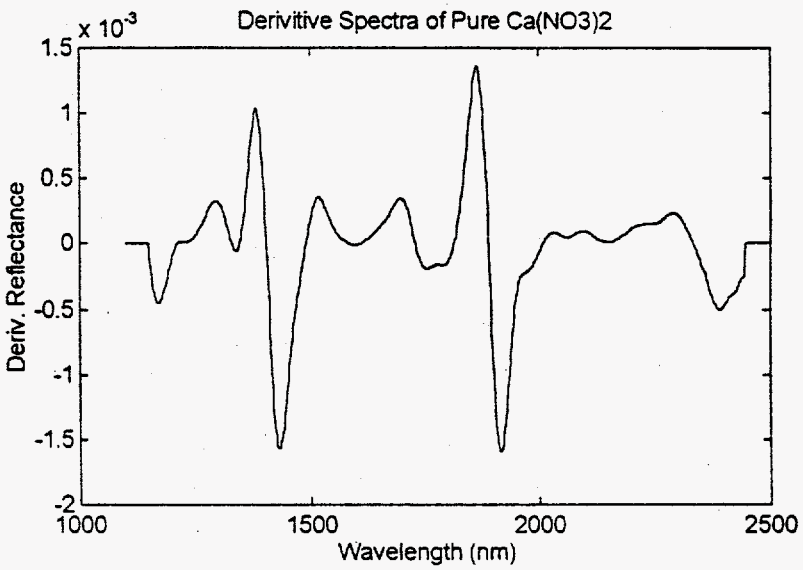


Figure A-34. Second derivative NIR spectrum of $Ca(NO_3)_2$

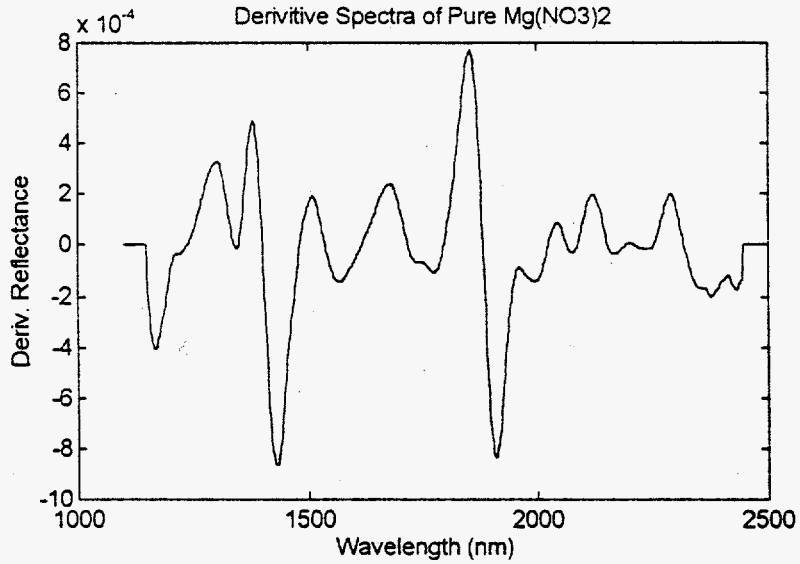


Figure A-35. Second derivative NIR spectrum of $Mg(NO_3)_2$

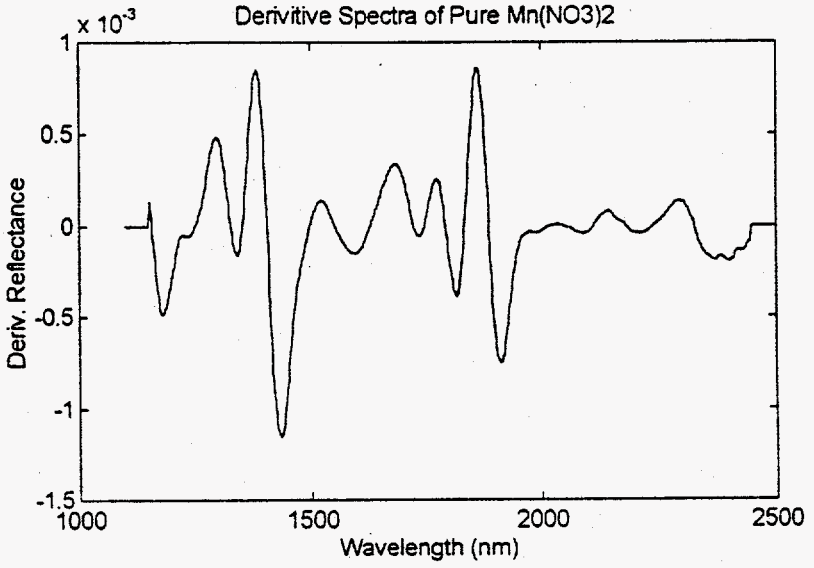


Figure A-36. Second derivative NIR spectrum of $Mn(NO_3)_2$