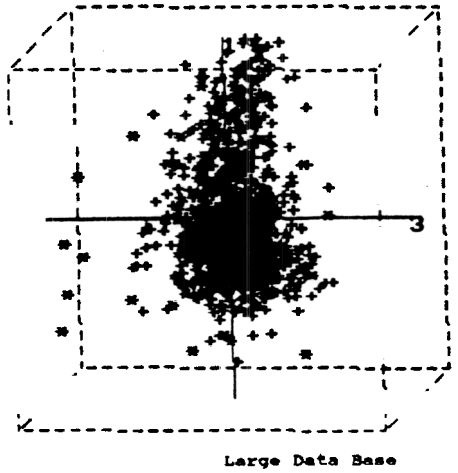
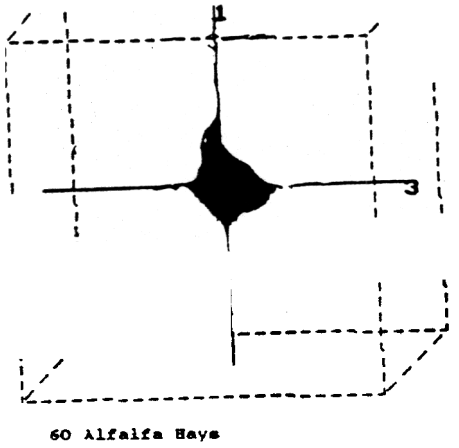
Regionalization of Near Infrared Calibrations

As calibration data bases for near infrared calibrations increase in size, the error associated with these calibrations tend to increase. These broad based calibrations are necessary where the samples for analysis come from a large geographical location and from different sample types and mixtures. An example of this would be a hay calibration that is to be used on hay samples that come from all over the country or the world. The data base used here is hay calibration which consists of 1300 samples selected from samples from all over the world. Thus the errors associated with the prediction of samples from one specific area would be higher with this calibration than the errors associated with a calibration developed from samples collected from that specific area. California alfalfa hay samples are a good example of this. A calibration developed from samples only from California should produce predictions with less error than the large data based calibration. However, the development of regional calibrations is a time consuming and expensive process as hundreds of samples would have to be screened and analyzed. In order to reduce the time and expense related to this regionalization, a method of selecting samples from the large data base that are similar to the local samples is useful.

In this example, a hay calibration based on 1230 samples selected from 1999 samples was used to predict 14 california alfalfa hay test samples. An additional sixty different California alfalfa hay samples were collected and analyzed to use in the regionalization of this "large" hay calibration. The following graphics display in 3 dimensions the PLS scores of the original calibration population and the relative position of the 60 California alfalfa hay samples.



Large Data Base



60 Alfalfa Hays

Author: Thomas R. Boyd, Application Specialist, NIRSystem, Inc., 12101 Tech Road, Maryland, 20904

Note the broadness of the large population and the relative compactness of the 60 California alfalfa hay samples. This indicates that a more narrow calibration might be useful in predicting samples of this type.

When the broad based calibrations was used to predict the 14 test samples the following were the statistical results:

	WET ADF	NIR ADF	WET PROTEIN	NIR PROTEIN
SEP	2.40		0.99	
Means	28.33	30.41	23.51	22.92
SEP(C)	1.24		0.83	
STD	2.90	2.56	1.81	1.36
Slope	1.02		1.20	
RSQ	0.82		0.82	

The sixty regional samples were then used to select samples from the large hay data base for use with the sixty samples for developing the regional calibration. Three samples from the large data base were selected for each of the 60 alfalfa hay samples. The Mahalanobis distance from each of the 60 regionalization samples to the closest "neighbors" in the large data base was used to identify the three matches for each sample. This resulted in about 180 samples from the large data base plus the sixty original samples, to give a total of 240 samples for the regional calibrations. A calibration was developed and the original 14 test samples were predicted with the new calibration. The statistical comparison is as follows:

	WET ADF	NIR ADF	WET PROTEIN	NIR PROTEIN
SEP	0.92		0.66	
Means	28.33	28.56	23.51	23.12
SEP(C)	0.93		0.55	
STD	2.90	2.89	1.81	1.53
Slope	0.95		1.14	
RSQ	0.90		0.92	

The regionalization of the large based hay calibration significantly improved the statistics associated with the prediction of the chemical method. This type of regionalization would be useful in any situation where a large data base exist and the samples for analysis come from a clearly defined, narrower population.