DATA FUSION WITH PARAFAC AND TRANSFER OF STACKED LOCAL CLASSIFIERS

by

Stephen Kaster

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Chemistry and Biochemistry

Spring 2013

DATA FUSION WITH PARAFAC AND

TRANSFER OF STACKED LOCAL CLASSIFIERS

by

Stephen Kaster

Approved: _____

Steven D. Brown, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Murray V. Johnston, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____

George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____

James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Some data analysis methods yield poor or only adequate information on their own but with data fusion, multiple datasets can be merged to possibly yield more information than when used alone. Data fusion can even be used to merge reduced representations of different parts of the same dataset. Data fusion yields improved results in situations where each set of data to be merged contains information unique from each other.

Stacked Partial Least Squares Discriminant-Based Classification (SPLSDA) transfer attempts to use data fusion to aid in applying previous analysis to new data. Data transfer or model transfer allows for use of datasets taken under different conditions or on different instruments, if this variation can be accounted for.

SPLSDA transfer is based upon a previously developed classification transfer approach which uses a reduced dimensional representation for each different section of the data, in order to classify new samples taken under different conditions. A similar method is Interval Partial Least Squares (IPLS) with the exception that these intervals collectively cover all of the data. Only some of the data in each section is used as much of the data contains very redundant information or is uninformative

which can hinder the classification model. The purpose of SPLSDA transfer is for transferring new infrared samples into an existent database, which were collected on a different instrument. Classification models can be fused from the infrared spectra of new samples and converted to allow classification using an existing database.

Another method involving data fusion is Parallel Factor Analysis (PARAFAC), which can be used to analyze multiple datasets simultaneously to find the causes of underlying variation in the sample. PARAFAC is used here to determine the concentration of three specific compounds found in a growth medium along with other unknown compounds.

Chapter 1

INTRODUCTION

There are instances where many datasets are provided, on a single sample or set of samples, and rather than analyze each dataset individually, the datasets can be combined and analyzed as a single multidimensional dataset. This process, called data fusion, can yield more information than analyzing datasets individually. The benefit of this is that information unique to each of the composite datasets is contained in the fused dataset. Data fusion is a broadly defined process comprised of multiple levels, which each contain many methods, some of which are covered in detail in the following pages.

Chapter 2

DATA FUSION

Data fusion is the technique where multiple datasets, representation of datasets, or resulting analysis from the datasets, can be combined to yield more information than either dataset analyzed individually[1-3]. One such method involves the stacking of classifiers[4].

There are three main levels of data fusion named for the type of representation of the data that is fused. Data level, feature level, and decision level data fusion, are convenient ways to summarize the different methods of data fusion, but methods typically can be classified into multiple levels. Data level fusion usually involves appending datasets together directly after some form of scaling and recentering. Appending features (simplified representations) of datasets is strictly feature level fusion but can also be thought of as data level. Decision level fusion involves classification of the data, or some segment of it, and combination of the results via a voting or other method. SPLSDA transfer involves the feature level of data fusion.

Chapter 3

STACKED CLASSIFIERS

The stacking used here is a form of data fusion where a dataset is broken up into multiple parts, each part is classified individually, then the weighed based on the results[5,6], similarly to IPLS[7,8], except that all parts are used in the analysis. The stacking algorithm was used to merge the information obtained by individual classifiers to yield better or comparable results to each individual classifier. Stacking was done for these classifiers taken over subsets of the dataset, allowing for regions with higher classification rates to improve the overall classification rate of the model. The classification results from this analysis were used as a weight for the stacking of all of the classifiers. The theoretical advantage of using stacking is that it should be approximately no worse than the same individual classifier done on the entire spectrum. If there is a region of very high classification rate or all areas are of similar classification rate, stacking should yield the same results as the individual classifier. In the case where there are multiple regions of relative importance, stacking can place more focus on these than the worse regions, resulting in an overall improvement in the classification.

Stacked Partial Least Squares

The recently developed classification technique Stacked Partial Least Squares Discriminant-Based Classification (SPLSDA)[9] was used to create the model for the classification transfer. Partial Least Squares (PLS) was developed as a regression technique but has been used as a classification technique as well[10-12]. In the closely related technique, Stacked Partial Least Squares regression (SPLS)[13], small intervals of the data matrix comprising the X-block (data) are each regressed on the Y-block (response) values separately. The simple regression models are then combined, giving a simpler and often better regression model, and can even have new regression models transferred to it[14]. For SPLSDA, a discriminant analysis-based classifier is used on each of the small intervals to classify samples. One main reason for using SPLSDA over some other classification techniques is the inherent dimension reduction obtained for each PLS model. Most PLS models are simple, with only a few latent variables needed to describe the class-related information in the data, as much of the data contains redundant information or is uninformative[15-17].

Each of these intervals must be optimized by cross-validation to determine an average misclassification rate. This average misclassification rate is used in the formulation of weights for each interval in the final, stacked, model. For the k^{th}

interval with s_k as the reciprocal of the number of misclassifications, the weight w_k is calculated as shown in equation 1 below. The summation normalizes all of the weights to a unit sum. If there are zero misclassifications for an individual interval, an appropriate weight is used instead (s_k^2 = 10). The purpose of calculating the weights this way is to ensure that a high weight is assigned to any interval with very few or no misclassifications.

$$w_k = \frac{s_k^2}{\sum_{k=1}^{n} s_k^2} \tag{1}$$

The calculated weight matrix is then used to effectively scale each interval's regression coefficients, which are then all summed together to obtain a single regression coefficient matrix. The regression coefficient matrix, $B_{k,SPLS}$, defines the discriminant distinguishing the target class from all other classes and is used to obtain the predicted Y values, $\hat{y}_u$, as seen in equation 2 below. Threshold values are then determined and used to classify samples as in or out of the target class.

$$\hat{y}_u = \sum_{k=1}^{n} w_k X_{k,u} B_{k,SPLS} \tag{2}$$

Establishment of a classification model with SPLSDA, using a previously generated dataset, was completed allowing the classification transfer process[1] to begin. This classification transfer is based on the calibration transfer technique[18]. Another dataset collected from a different instrument uses the previously calculated regression

coefficient matrix to find the predicted Y values as if they had been collected on the original instrument. The same set of discriminants used in the final classification in SPLSDA are also used to classify the dataset collected from the different instrument.

Analysis of Forensic Paint Data

A dataset involving infrared spectra from paint samples and analyzed on a Thermo-Nicolet instrument, was provided by Oklahoma State University (OSU). The intent is to compare this data to the infrared spectra of paint samples collected in the International Forensic Automotive Paint Data Query-Canada (PDQ) database, most of which were analyzed on a Bio-Rad instrument. Only the samples from the PDQ database that were measured on the Bio-Rad instrument were used in the analysis, as any instrumental differences within the (PDQ) database would affect the transfer. The spectra from the PDQ database and from OSU are shown below in Figures 1 and 2 respectively. The forensic paint data included 19 classes where the classes assigned were the identity of the manufacturing plant in which the car paint was produced.

Another assignment of classes, based on clusters of manufacturing plants instead of individual plants, was also attempted. This class assignment yielded improved results because of the simpler discriminants that could be used. By clustering manufacturing plants with similar features, the class discriminants only

need to separate clusters of similar plants from other clustered plants. Because the plant clusters differed much more than the individual manufacturing plants, the discrimination is greatly simplified. The remainder of the analysis of the forensic paint data will use individual manufacturing plants as classes.

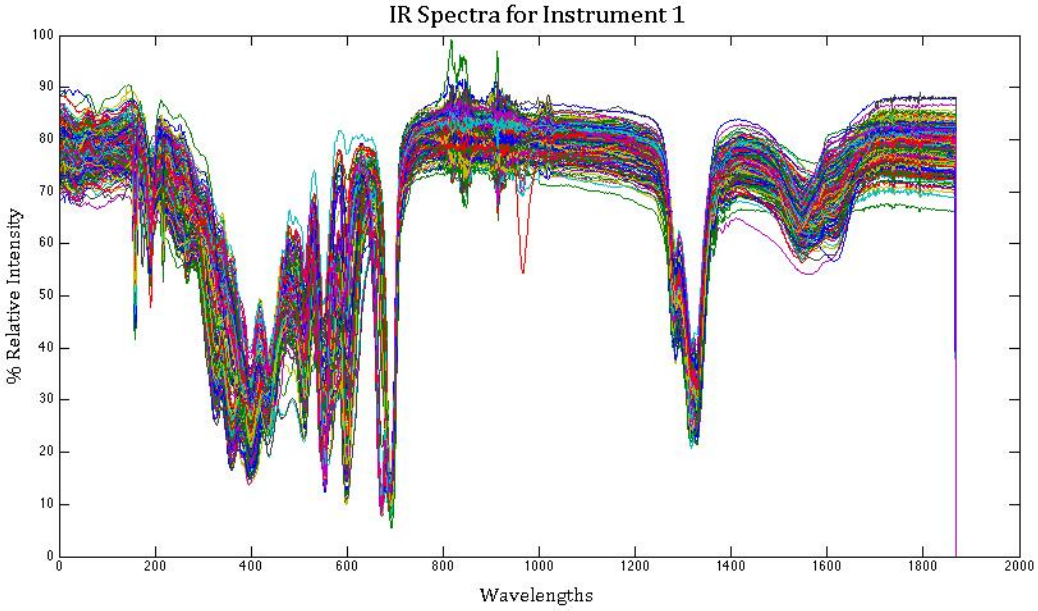

Figure 1 IR spectra collected with the Bio-Rad instrument (PDQ database).

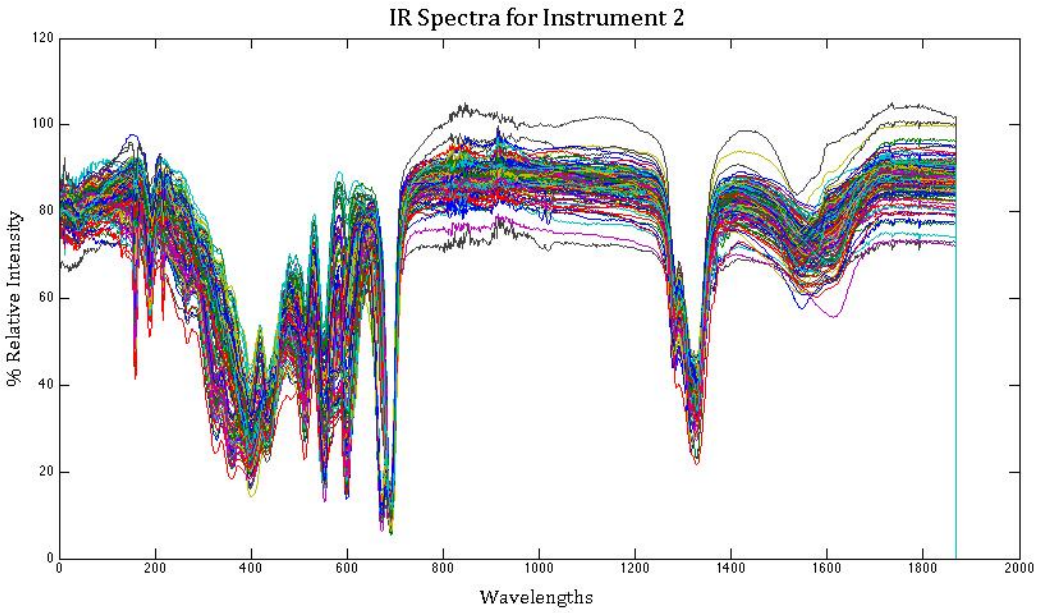

Figure 2 IR spectra collected with the Thermo-Nicolet instrument (OSU dataset).

Both the Bio-Rad and Thermo-Nicolet infrared datasets have the same general shape, due to the analysis being done on the same type of compound, car paint. There are regions that are quite distinct between the two instruments and the Thermo-Nicolet instrument seems to have, in general, more variation in the intensities for the spectra.

Some sample removal and preprocessing must be done to allow the classification and transfer using SPLSDA. Any sample that was the only sample in its class was removed because it could not be cross-validated, which disallows the calculation of the weights. Also, any sample in the second (Thermo-Nicolet, OSU database) instrument's dataset in a class that was not part of the first instrument's

(Bio-Rad, PDQ database) dataset could not be involved in the transfer. If the first instrument did not have any samples in a particular class, the transfer will always misclassify this sample, as the transferred model cannot account for classes that were not in the training set. All datasets were then preprocessed using Savitzky-Golay first derivative smoothing using the window size of 15 wavelengths followed by mean-centering.

SPLSDA transfer was used on the first instrument samples to aid in the classification of the second instrument samples without using their class identities. Because a one-vs.-all discriminant analysis is used within SPLSDA, there is little benefit in analyzing only a few, select classes. Clustering of classes generally improved results, as clusters made from similar classes are more easily distinguished from each other than the classes taken separately.

Chapter 4

MODEL TRANSFER

There are many cases where analysis of standards and samples is performed and data is generated under certain conditions. However, at a later time, when additional data is generated, conditions have changed or standards are no longer available. Model transfer is a method that attempts to correct for these issues, such as instrument failure and subsequent replacement, instrumental variation due to too long of a time span between runs, sample potency decrease, or other reasons. In model transfer, a classification is done using any desired algorithm (here SPLSDA). The weights used in the model along with the regression information are used to regress the new spectra to the old classes. Usually before a direct transfer can be done, the new data needs to be centered and scaled to match the old classes. As with any standards used, the new samples must have no new classes in them otherwise there is no chance of a correct classification.

Classification Transfer of Paint Spectra

The dataset provided by OSU (Thermo-Nicolet instrument) involved infrared spectra of clearcoat paint samples from vehicles made at different manufacturing plants. This dataset was analyzed with stacked PLSDA using a maximum of 50

intervals, each with 10 latent variables, and 10-fold cross-validation on the spectrum. Cross-validation was done by selecting half of the samples from each class for the test set with the remainder in the training set. A total of 1869 wavelengths were available in the spectra as provided. Each interval and latent variable combination seen in Figures 3, 4, 5, and 6 represent a different SPLSDA model with its overall classification errors.

For the paint dataset, there was no region in the infrared spectrum that gave a very high stacking weight for the stacked discriminants. This result is peculiar in that high weights are normally associated with informative regions of the data. Ideally, there will be a region or a series of particular regions that are more informative for the classification than the other regions, and these more informative regions will have good accuracy in the cross-validation, leading to very high stacking weights for these regions. These heavily weighted regions allow the stacked classification to focus more on these informative regions. The fact that no particular interval in the data leads to a much better classification implies that the entire dataset is equally informative or equally uninformative. Because the stacking weights are very similar across all intervals in the paint dataset, the stacking classifier algorithm demonstrates less of an improvement over that obtained with conventional PLSDA (seen as the 1-interval case) when compared to datasets that result in intervals with high stacking weights.

Table 1 Classification errors for SPLSDA as compared to SPLSDA Transfer.

Class	SPLSDA Classification Error (%)			SPLSDA Transfer Classification Error (%)		
	LV 1	LV 2	LV 3	LV 1	LV 2	LV 3
1	26.97	16.87	8.28	14.86	18.45	15.34
2	5.48	1.12	1.26	3.26	2.94	3.06
3	21.90	14.98	13.98	12.95	12.67	13.40
4	26.40	22.05	17.14	15.91	12.70	9.66
5	27.55	17.06	13.73	19.25	21.38	17.14
6	8.25	5.54	5.26	6.21	4.49	5.86
7	0.24	0.26	0.35	2.37	2.55	2.71
8	29.56	14.46	13.46	19.87	32.04	23.66
9	9.54	7.51	8.15	3.82	3.74	4.50
10	3.93	4.22	4.25	9.75	8.55	6.68
11	30.54	9.64	7.83	9.77	12.83	9.15
12	3.21	0.95	0.62	2.73	2.49	2.61
13	29.67	7.10	6.62	23.64	14.98	11.00
14	8.50	9.22	8.94	7.15	10.47	10.37
15	7.00	7.09	9.48	7.08	7.53	6.55
16	12.71	11.41	7.85	5.56	8.84	9.26
17	9.55	9.43	8.39	9.10	11.04	14.04
18	9.76	8.09	6.11	3.88	3.78	5.73
19	7.04	7.30	6.05	7.10	6.74	6.47

The classification errors shown in Table 1 are means calculated for all 19 classes over the number of intervals, as the number of intervals used in stacking has a smaller effect on the classification than the number of latent variables. As an example, class 1 and 2 are representative of high and low classification errors, respectively. The accuracy of classification for SPLSDA transfer was often approximately the same or lower than that obtained from building a classification

model on the target data, when the samples measured on the second instrument are directly regressed on their own class identities using SPLSDA. This result is to be expected, as a classification model built on the training data should yield better classification results on the training dataset than one built on different data. The corrections involved in using the model from the Bio-Rad data, which uses predetermined threshold values on the response matrix adjusted by using the weighted regression coefficient matrix, reduce the increase in classification error for the Thermo-Nicolet data. Because the classification error change between most of the stacking intervals is less than that of the latent variables, a mean classification accuracy can be calculated from the results of each number of stacking intervals used to better compare the classification accuracy of the first few latent variables for SPLSDA and SPLSDA transfer. Only classification results from use of the stacked models using combinations of up to the first three latent variables are reported here, for the same reason as that given above in the discussion of the synthetically-generated IR-spectral data.

Figures 3 and 4, and Figures 5 and 6, depict the performance of SPLSDA transfer for both poorly defined (3 and 4) and well-defined (5 and 6) classes. Figures 3 and 5 show SPLSDA classification errors for the Thermo-Nicolet dataset while Figures 4 and 6 the SPLSDA transfer classification errors from the Bio-Rad dataset, after the model transfer.

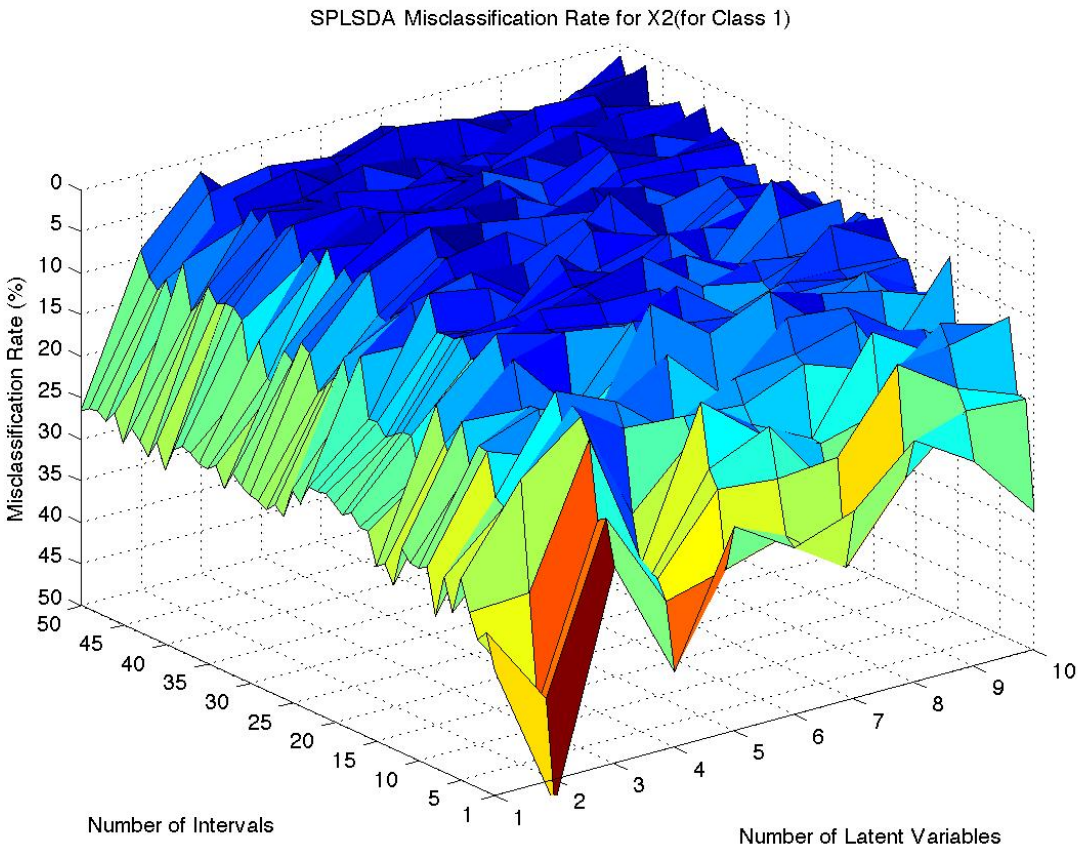

Figure 3 Classification error for class 1 on the Thermo-Nicolet instrument.

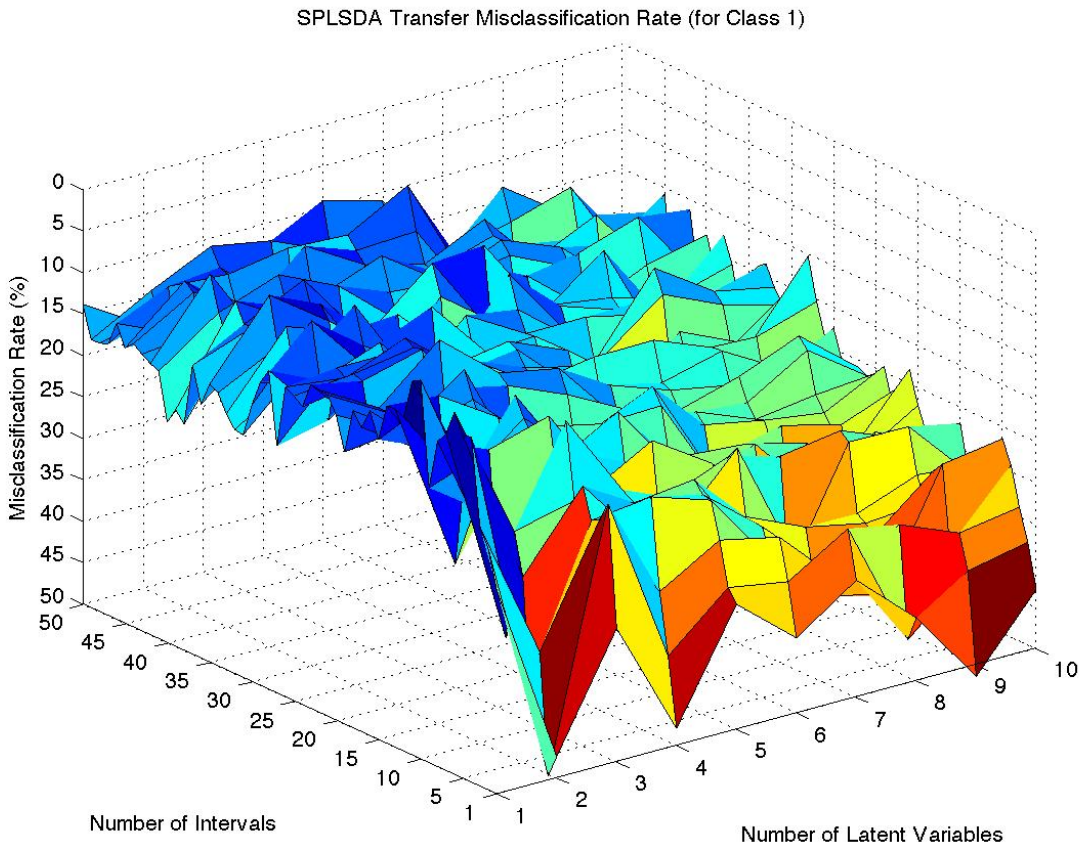

Figure 4 Classification error for class 1 from the trasfer of the Thermo-Nicolet instrument to the Bio-Rad dataset.

There are a few particularly well-defined classes that result from both the SPLSDA classification and SPLSDA transfer. Certain classes, like class 2, had very low classification errors across many combinations of number of intervals and maximum number of latent variables used, as seen in Figures 5 and 6. Classes such as class 2 are very distinct from the other classes in their spectra, allowing for the

improved classification. As only one-vs.-all discriminants were used, classes that have very low classification error (such as class 2) must have an easily established discriminant boundary. Unfortunately, there are some classes that have much similarity in the spectra and it is difficult to correctly separate these classes, like class 1.

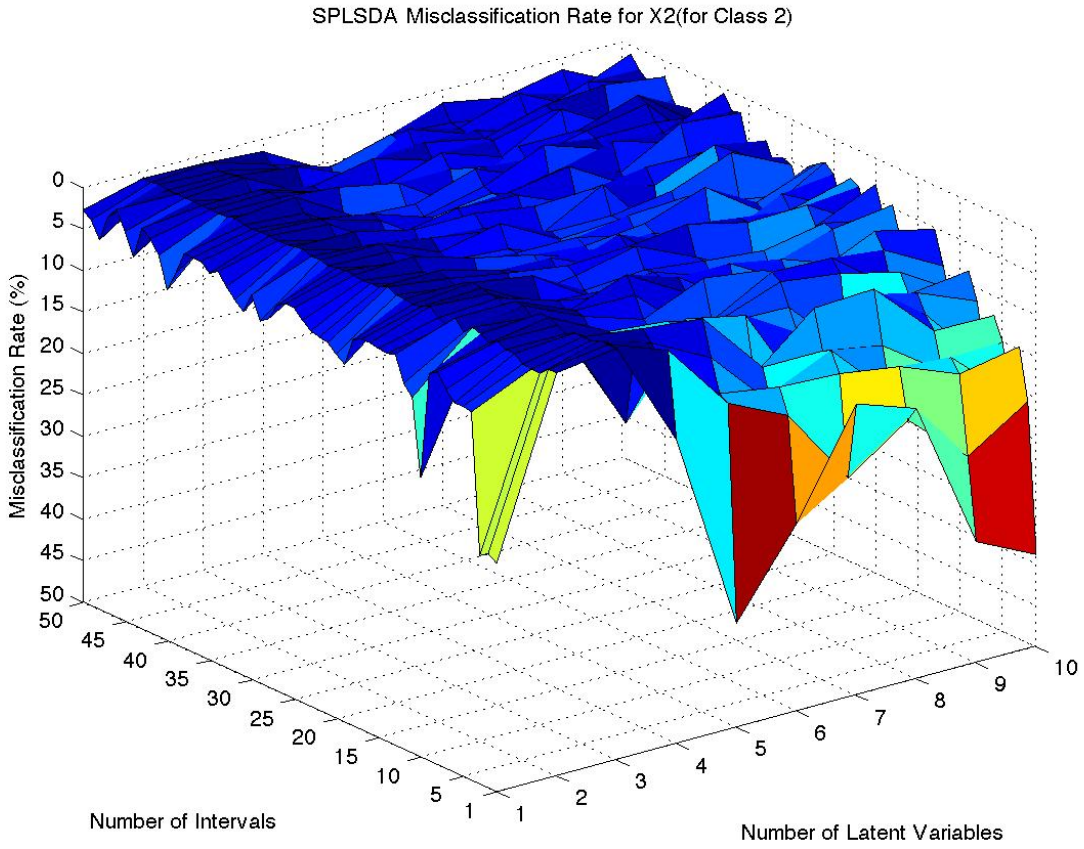

Figure 5 Classification error for class 2 on the Thermo-Nicolet instrument.

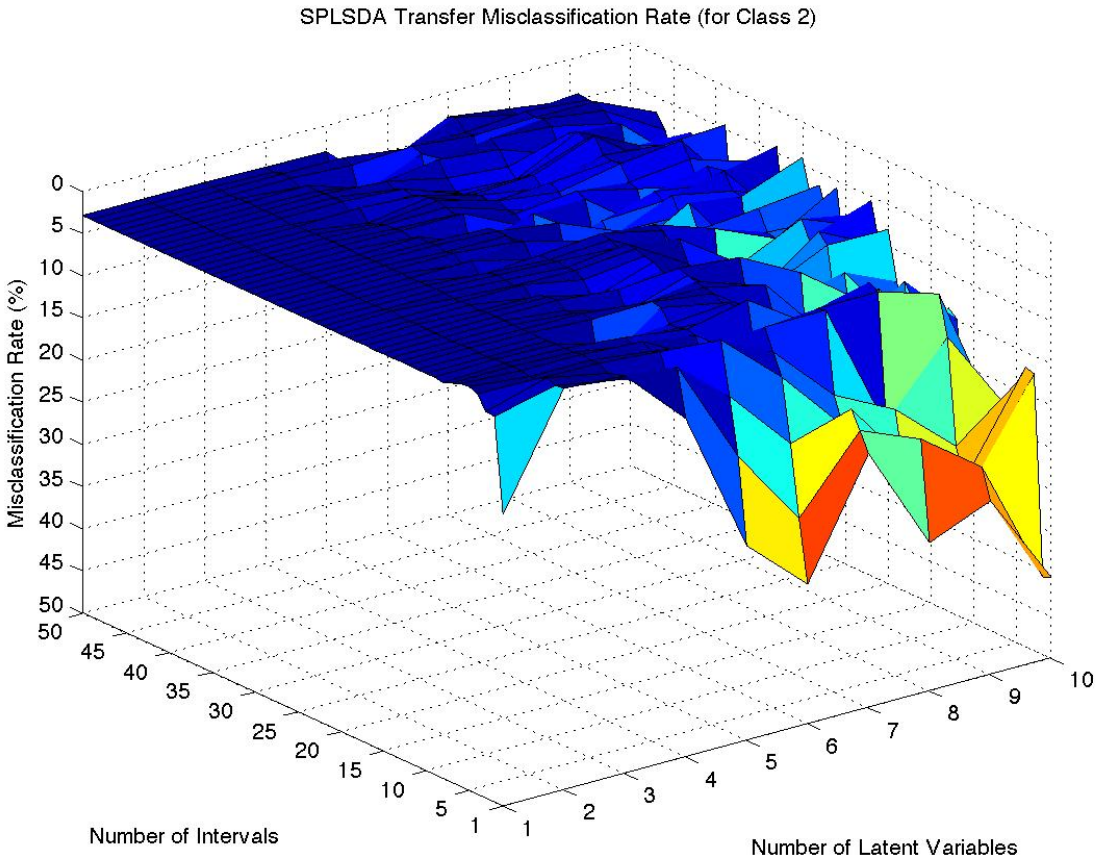

Figure 6 Classification error for class 2 from the transfer from the Thermo-Nicolet instrument to the Bio-Rad dataset.

Figures 5 and 6 above show that SPLSDA transfer has higher classification errors than a direct SPLSDA classification for some combinations of stacking intervals and number of latent variables used. The classification error of SPLSDA transfer did not rise considerably over those obtained from a direct classification, especially considering no class information was provided in the transfer. In the cases where the

error in classification after transfer does not differ much from the error obtained from a direct SPLSDA classification, it appears that both spectral datasets have the same underlying distributions for each class. Even though only the first few latent variables are similar in many classes, these latent variables are the most important to the classification. The remainder of the computed latent variables seems to be mostly contributions from spectral noise and instrumental differences.

Chapter 5

PARAFAC

Parallel Factor Analysis (PARAFAC)[19] is a method that can be used to analyze a three or more dimensional matrix of data. Two datasets of the same samples can be analyzed alongside each other by first merging them into an outer product (OP) matrix. This is simply the outer product of both matrices taken for each sample. This new matrix will resemble a three dimensional dataset that can be used in PARAFAC. If given standards for each desired component, PARAFAC can attempt to find the concentrations of these standards within an unknown mixture. This method should work given that there are no other compounds that have a high response in regions similar to the standards in both spectra. If at least one spectrum differs between the standard and an unknown, PARAFC should be able to correctly calculate the value of the known compound with no actual input of what the unknowns are.

The goal in using PARAFAC was to use the design to find the spectral variations due to the differing concentrations of the compounds used in a process. These spectral differences found in the design near-infrared (NIR) and nuclear magnetic resonance (NMR) datasets could then be used to find the concentrations of the compounds in an unknown mixture later in the production. The design datasets are akin to creating standards for multiple concentrations for each of the compounds, so that the concentrations of all 3 can be determined simultaneously.

As PARAFAC is very computationally draining for large outer product matrices, the original data must have its dimensionality reduced. PLS was originally used to regress each sample to the corresponding Y-block(s). Different preprocessing of both Nuclear Magnetic Resonance Spectroscopy (NMR) and Near-Infrared Spectroscopy (NIR) were tried, using the Root-Mean-Square Error of Cross-Validation (RMSECV) from PLS to determine which preprocessing was most appropriate. As the number of variables in both datasets needed to be reduced, any regions found to have low importance in the PLS were removed. The final preprocessing after variable selection is shown in the table below for each dataset. Note that Savitzky-Golay smoothing was done before variable selection.

Table 2 Preprocessing methods for OP used in PARAFAC.

Dataset	Preprocessing Methods (in order of use)		
Design NIR	Savitzky-Golay 1st Derivative Smoothing	Mean-centering	
Design NMR	Variable Alignment	Length Normalization	Mean-centering
Sample NIR	Mean-centering		
Sample NMR	Variable Alignment	Length Normalization	Mean-centering

The reduced datasets are both over an order of magnitude smaller than the original datasets, making the OP matrix over 100-fold smaller (NMR 64836 → 4776, NIR 4407 → 364). This led to a great reduction in computation time for PARAFAC.

Although mean-centering can be done on the OP matrix, it was done to the datasets prior to OP creation. The process of mean-centering is more complicated on tensors than it is on the matrices used to create them. After creation of the fully preprocessed OP matrix was completed, PARAFAC could be run.

Two differently coded PARAFAC algorithms were compared, one from Rasmus Bro, Ph. D. and the other from Eigenvector Research, Inc. Both produced the same result (within error) but the algorithm from Dr. Bro allowed for finding the optimal number of principle components (PC) and had a cleaner graphical output. Also, the only non-graphical data output from the Eigenvector model was the model itself while Dr. Bro's algorithm also gives correlation coefficients, residuals for the fit of the model, and the iterations needed to complete each run of PARAFAC. From this point on, the code was tailored to use Dr. Bro's PARAFAC.

PARAFAC was then run with many different combinations of preprocessing, constraints, and number of extracted components (sequentially largest sources of variation in the data, ideally corresponding to the growth medium concentrations). PARAFAC takes exponentially longer after passing the optimal number of

components as it attempts to find the least correlated next component. For the design set, all PARAFAC runs had very few iterations for the first three components compared to the fourth and beyond. This should be expected, as there are only 3 components that were varied in the design.

The design was most apparent in the spectra with all preprocessing done except for mean-centering. Mean-centering seemed to amplify the differences between samples of similar composition. The scores from this model were regressed to the concentrations using least squares fitting. Constraining the slope of the fit line to nonnegative seemed to worsen the fit, meaning that some of the extracted components may be anti-correlated to the concentrations or PARAFAC simply generated a poor model.

The loadings for each component should correspond to the spectra associated with the compound the component is correlated with. The loadings do not seem to match as they are missing key regions that are specific to the compounds they are correlated with. It seems that the components being extracted do not follow the design in that any adjustment to the overall mixture changes the concentration of all 3 compounds, not just the ones corresponding to the compound concentrations that were altered. As each component seems to correlate to all three compounds, even if in differing intensities, it is not surprising that the mixture that regressed the best was the one with equal concentrations of all 3 compounds.

For the analysis of the sample dataset, there appears to only be one or two components for the combinations attempted. This is problematic as the loading need to be matched to the design to determine which component corresponds with which compound.

Using least squares regression to fit the PARAFAC scores to the concentrations yields Root Mean Square Error (RMSE) errors on the order of 0.2 to 0.6. This is much worse than the RMSECV of 0.02 or less found using PLS of the NMR or NIR individually. A better comparison would be to cross-validate the PARAFAC analysis of the OP matrix, but PARAFAC is too time consuming a method for this to be practical. Note that all unconstrained runs of PARAFAC had come at least close to the same minimum so that the scores and loadings were barely distinguishable if at all.

Design Set and Transfer of PARAFAC Model

The goal in this analysis is to create a calibration for multiple compounds in a growth medium. Two sets of NIR and NMR datasets were provided. The first is spectra generated from mixtures of 3 compounds of various ratios. The second is spectra taken on a growth medium comprised of amino acids, sugars, etc., including the 3 target compounds. The purpose is to find the concentrations of these 3 particular

constituents of the growth medium and set up a calibration for these concentrations so that only the growth medium need be analyzed directly at a later point.

At first Lorentzian peaks were used to model the 'pure' spectra for each compound based on the samples containing only 1 component. This method may decrease error in regression due to the removal of instrumental error but the process of modeling each peak in the NMR is too time consuming.

Another method for creating the 'pure' component loadings was used by making an extra assumption. Assuming the samples containing only the 1 molar pure compound have insignificant instrumental error and contamination, their spectra can be used as the basis for the pure compound loadings. By generating 1-component PARAFAC models using the 3 samples corresponding to each pure component, the pure loadings can be created. These loadings are then used as fixed loadings in the PARAFAC using the entire OP matrix. Results have lowered errors compared to all previous PARAFAC models created. The result of the regression can be seen in Figure 7, in which the relative concentrations of the compounds are shown. The design was constructed according to relative concentrations of 3 dilute solutions of these compounds. There are triplicate of each combination with the fraction of each compound summing to 1. The combinations shown below are approximately 1:0:0, 0.5:0.5:0, 0.67:0.33:0, and 0.33:0.33:0.33.

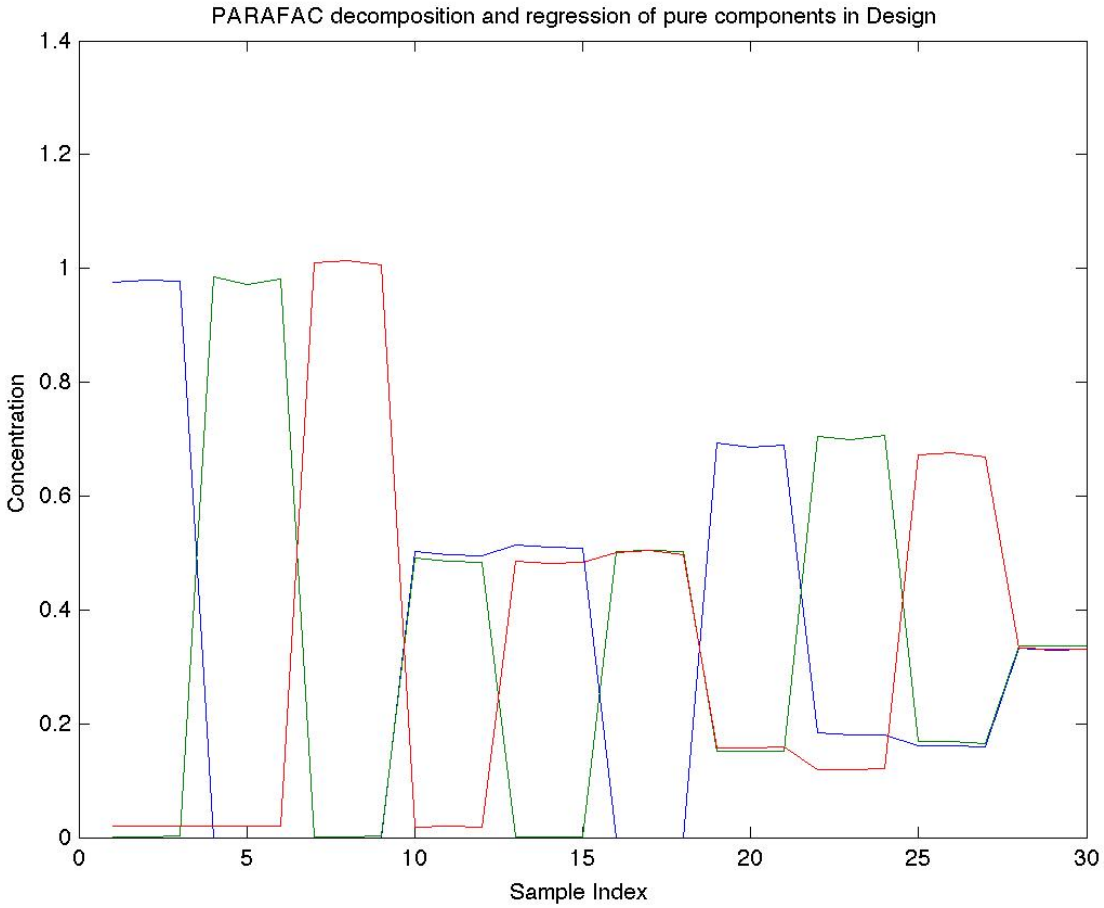

Figure 7 Predicted relative concentrations for each compounds after PARAFAC analysis and original least squares regression (slope and intercept).

Table 3 RMSE of PARAFAC regression compared to PLS regressions.

	RMSE		
PARAFAC	0.0122	0.0149	0.0197
NIR PLS	0.009549	0.010051	0.010250
NMR PLS	0.012955	0.015432	0.022643

As seen in Table 3 above, the RMSE of this PARAFAC model is worse than that of PLS done on the NIR individually but shows a slight improvement over using just the NMR. This is expected as PARAFAC has taken into account both datasets and being essentially a PLS fit of its own, cannot improve results over the lower error dataset (NIR) but uses its information to lower the error compared to the worse error dataset (NMR).

One other major disadvantage is that the PARAFAC model is highly dependent on the pure compound samples, so any contamination or significant noise can degrade the model. The advantage to PARAFAC is that with this method so cross-validation need be done, as there are as many components as pure compounds used. As only 1-component PARAFAC models are used followed by the fixed loadings PARAFAC, computations are very fast.

After the regression using the design PARAFAC results, the loadings were used along with the sample dataset OP matrix to generate corresponding scores (relative concentrations assuming PARAFAC worked as intended), as seen in Figure 8 below. Regression of these scores leads to concentrations more than 5 orders of magnitude below the design matrix with approximately equal concentrations across all samples.

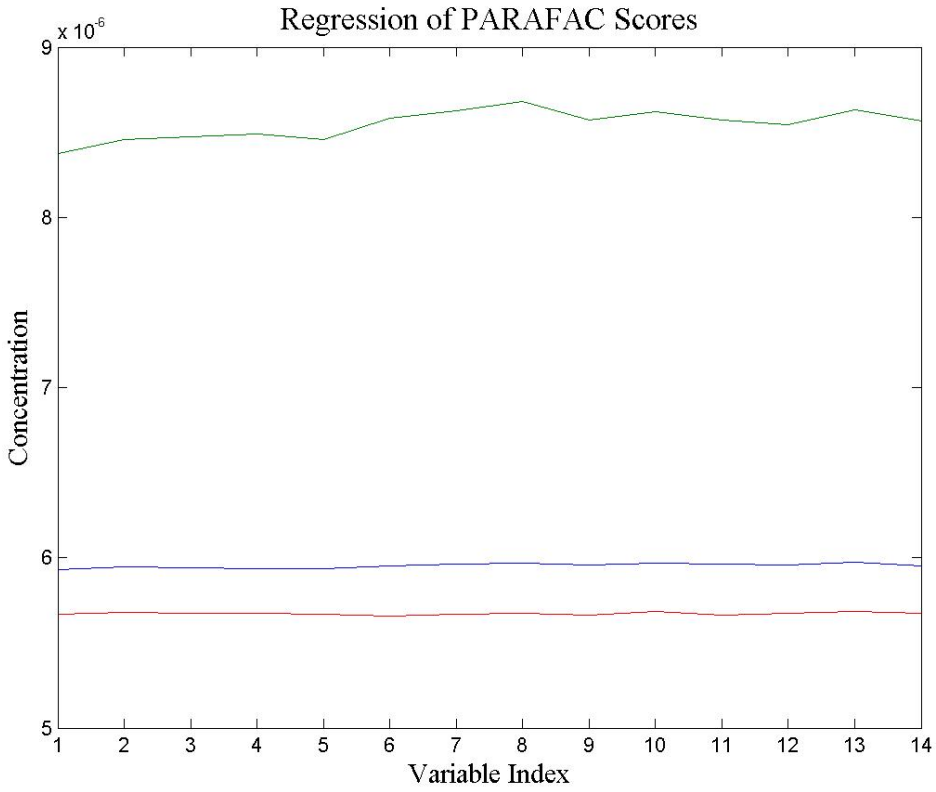

Figure 8 Regression of the 3-component PARAFAC model generated using the fixed loadings.

The discrepancy between the predicted and actual concentrations in the sample dataset is likely because the PARAFAC model cannot account for the extra sources of variation. Ideally the PARAFAC model would be created with samples that represent the 'pure' version of the growth medium with the design compounds. As it stands now, PARAFAC cannot extract any information regarding the design compounds as

any amount in the growth medium samples have too low of spectral intensities to detect.

One possible solution to this is to have at least one sample for each pure compound be spiked with a known concentration so that PARAFAC will be able to extract the pure loadings for the both the pure component and the remainder of the sample dataset components. Alternatively a single sample spiked with all 3 pure compounds may also work. This would hopefully allow the pure loadings from the design to be used in the sample dataset PARAFAC along with these newly generated sample dataset loadings.

Chapter 6

CONCLUSION

Data fusion can aid in the analysis of complicated or multiple datasets, such as this, but there are limitations. Methods such as PARAFAC are contingent upon the desired components having a large effect on the data and can fail when the data does not vary enough based on these components. Other types of methods involving data fusion, like SPLSDA, take advantage of particular parts of the data that vary with the response. Data fusion methods are limited by the quality of the data itself, and by issues of dimensionality, as it is uncommon to have very large datasets with few variables.

REFERENCES

1. M. Pottmann, B.A. Ogunnaike, and J.S. Schwaber, Ind. Eng. Res. **44**, 2606 (2005)
2. L. Fortuna, S. Graziani, G. Napoli, and M.G. Xibilia, IEEE Industrial Electronics, 32nd Annual Conference**,** 229-234 (2006).
3. M. Hibon and T. Evgeniou, Int. J. of Forecasting **21**, 15 (2005).
4. L. Kuncheva, IEEE Trans. Pattern Anal. Machine Intell. **24,** 281 (2002).
5. L. Breiman, Machine Learning **24**, 49 (1996).
6. L. Breiman, Machine Learning **24**, 123 (1996).
7. L. Nørgaard, A. Saudaland, J. Wagner, J.P. Nielsen, L. Munck, and S.B. Engelsen. Appl. Spectrosc. **54**, 413 (2000).
8. J.H. Jiang, R.J. Berry, H.W. Siesler, and Y. Ozaki, Anal. Chem. **74**, 3555 (2002).
9. W. Ni, S.D. Brown, R. Man, "Stacking local classifiers for multivariate classification." (to be submitted for publication)
10. W. Rayens and M. Barker, J. Chemometrics **17**,166 (2003).
11. G. Musumarra, V. Barresi, D.F. Condorelli, G.C. Gortuna, and S. Scirè, J. Chemometrics **18**,125 (2004).
12. S. Wold and J. Trygg, J. Chemometrics **20**, 341 (2007).

13. W. Ni, S.D. Brown, R. Man, "Stacked partial least squares regression analysis for spectral calibration and prediction," J. Chemometrics, **23**, 505-517 (2009).

14. W. Ni, S.D. Brown, R. Man, Anal. Chim. Acta **661**, 133–142 (2010).

15. L. Xu and I. Schechter, Anal. Chem. **68**, 2392 (1996).

16. A. Höskuldsson, Chemom. Intell. Lab. Syst., Lab. Inf. Manag. **55**, 23 (2001).

17. D. Ballabio, T. Skov, R. Leardi, and R. Bro, J. Chemometrics **22**, 457 (2008).

18. A.J. Myles, T.A. Zimmerman, S.D. Brown, Appl. Spectrosc. **60**, 1198-1203 (2006).

19. R. Bro, Chemom. Intell. Lab. Syst., **38**, 149-171 (1997).