CLASSIFICATION OF DRUGS OF ABUSE USING MASS SPECTRAL DATA FOR THE IDENTIFICATION OF NOVEL PSYCHOACTIVE SUBSTANCES.

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Chemistry.

By

Garion Lucas Schneider

Advisor: Dr. Nuwan Perera
Assistant Professor of Forensic Science
Department of Chemistry & Physics

Committee Members: Dr. Scott Huffman, Department of Chemistry & Physics
Dr. William Kwochka, Department of Chemistry & Physics

November 2022

ACKNOWLEDGEMENTS

I would like to thank my committee members for their advice and assistance. In particular, I would like to thank Dr. Perera for his constant willingness to work and try new approaches with me. While the work has been challenging, these faculty members have always pushed me to try new things.

I would also like to recognize my parents, as they have instilled in me the drive to try difficult paths to build myself as a person.

Finally, I would like to recognize my loving wife, Mary, whose continued support and encouragement has made this work possible. She has, as always, been my rock, and the foundation of my support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NPS	Novel Psychoactive Substance
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
FTIR	Fourier Transformed Infrared spectroscopy
GC-MS	Gas Chromatograph Mass Spectrometry
SWGDrug	Scientific Working Group for the Analysis of Seized Drugs
EI	Electron Impact ionization
GA	Genetic Algorithm
MATLAB	MATrix LABoratory
GUI	Graphic User Interface
DEA	Drug Enforcement Administration
WHO	World Health Organization
FBI	Federal Bureau of Investigation
LV	Loading Vector
TPR	True Positive Rate
TNR	True Negative Rate
FPR	False Positive Rate
FNR	False Negative Rate
CTP	Cathinones Tryptamines and Phenethylamines

ABSTRACT

Classification of Drug of Abuse Using Mass Spectral Data for the Identification of Novel Psychoactive Substances (NPSs).

Garion Schneider

Western Carolina University (November 2022)

Advisor: Dr. Nuwan Perera

Novel psychoactive substances (NPSs) have been increasingly reported in recent years and possess significant risks to public health worldwide. These substances, sometimes known as “legal highs”, are newly designed drugs that mimic the effects of commonly abused drugs and are comprised of several drug classes which include opioids, cannabinoids, stimulants, and benzodiazepines. Many NPSs share similar chemical structures with commonly abused drugs and produce similar psychoactive responses by binding to receptors in the body. These NPSs are designed to circumvent the regulations that limit the use of recreational drugs and to create more potent drugs such as fentanyl derivatives. In a typical forensic laboratory analysis, an analyst uses a panel of known drug standards or reference materials to identify and quantify drugs present in a sample (or evidence) using chromatographic methods such as gas chromatography mass spectrometry (GC-MS) or liquid chromatography mass spectrometry (LC-MS). If a compound present in the sample is not included in the test panel, mass spectral libraries can be used to find the identity of that compound by comparing the mass spectrum of the unknown with the mass spectra of known compounds present in the library. In the case of new NPSs that have not been reported, no reference materials or reference spectra are available. In this scenario, forensic labs have to rely on gathered intelligence data, prior knowledge of these NPSs, and some additional analysis methods, such as nuclear magnetic resonance spectroscopy (NMR) or high-resolution mass spectral data (HRMS), to determine the presence of NPSs. However, structural elucidation of novel compounds is time consuming and costly, thus there is a growing interest to develop methods that can proactively determine the presence of NPSs using chemometric methods. The focus of the current research work is to develop proactive solutions to identify newly designed NPSs when the reference spectra are not present in the spectral libraries used in forensic laboratories. A classification system is developed using existing data of known substances that can be used to determine the presence of NPSs. Herein, we demonstrated a model developed using mass spectral data and chemometric methods, such as principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA), that can accurately discriminate novel fentanyl derivatives from non-fentanyl related drugs. Furthermore, we have developed a sub-model using aforementioned methods to discriminate fentanyl derivatives based on structural modifications. Validation results show that these methods are robust with high accuracy (>95%), true positive rates (>95%), and true negative rates (>95%).

CHAPTER ONE: INTRODUCTION

Background

NPSs became popular worldwide around the late 2000's and over 1150 NPS compounds have been reported to the United Nations Office on Drugs and Crime (UNODC) Early Warning Advisory (EWA) by 137 countries and territories by October 2022.[1,2,3,4] It is reported that there were 24 NPSs identified for the first time in the USA in 2020.[5] There are over 60,000 deaths reported in the US attributed to opioid crisis and the majority of them are related to fentanyl and related substances.[3] In post-mortem analysis related to drug overdoses, synthetic opioids like fentanyl derivatives made up 23% of the cases, second only to sedatives/hypnotics. [4] Currently, this epidemic is taking the form of synthetic fentanyl derivatives with various structural modifications.[5] In a typical forensic laboratory analysis, an analyst uses a panel of known drug standards or reference materials to identify and quantify drugs present in a sample (or evidence) using chromatographic methods such as gas chromatography mass spectrometry (GCMS) or liquid chromatography mass spectrometry (LCMS). If a compound present in the sample is not included in the panel, mass spectral libraries can be used to find the identity of that compound by comparing the mass spectrum of the unknown with the mass spectra of known compounds present in the library. These libraries are continuously updated to include NPSs that are identified by various institutions such as crime laboratories. In the case of new NPSs that are not reported before, since there are no reference materials or reference spectra are available, these labs rely on intelligence data, prior knowledge on NPSs, and some additional analysis methods such as nuclear magnetic resonance spectroscopy (NMR) or high-resolution mass spectral data

(HRMS) to determine the presence of NPSs. However, structural elucidation of novel compounds is time consuming and there is a growing interest to develop methods that can proactively determine the presence of NPSs using chemometric methods.[6,7,8]

Fentanyl Related Compounds

Fentanyl is a synthetic opioid that interacts with the opioid receptors present in the central nervous system. It is used as a strong pain medication in clinical settings and used illicitly as a recreational drug. Different sites of fentanyl core structure can be substituted to create novel fentanyl related substances. These sites include phenethyl group, piperidine ring, aniline ring, and amide group (see Figure 1.) and minute changes on these sites would create compounds that have similar chemical behaviors to fentanyl.[1] Fentanyl and fentanyl related substances are responsible for majority of opioid-related overdose deaths in the USA.[5] In a recent study, Koshute, et al. have shown that mass spectral data can be used to differentiate fentanyl derivatives from other drugs using machine learning techniques as a complimentary technique for mass spectral library search of unknown compounds.[8] However, this study does not extend to the subclasses of fentanyl analogues. In a separate study, Gilbert, et al. have developed a chemometric model to classify fentanyl related substances based on the structural position of a modification using hierarchical classification and principal component analysis (PCA).[6] However, this study does not include a method to classify fentanyl related compounds from other drug classes. Fentanyl subgroups used in this study has only a few samples (3 to 19 samples per class) and total number of samples used in the model calculation was 54.

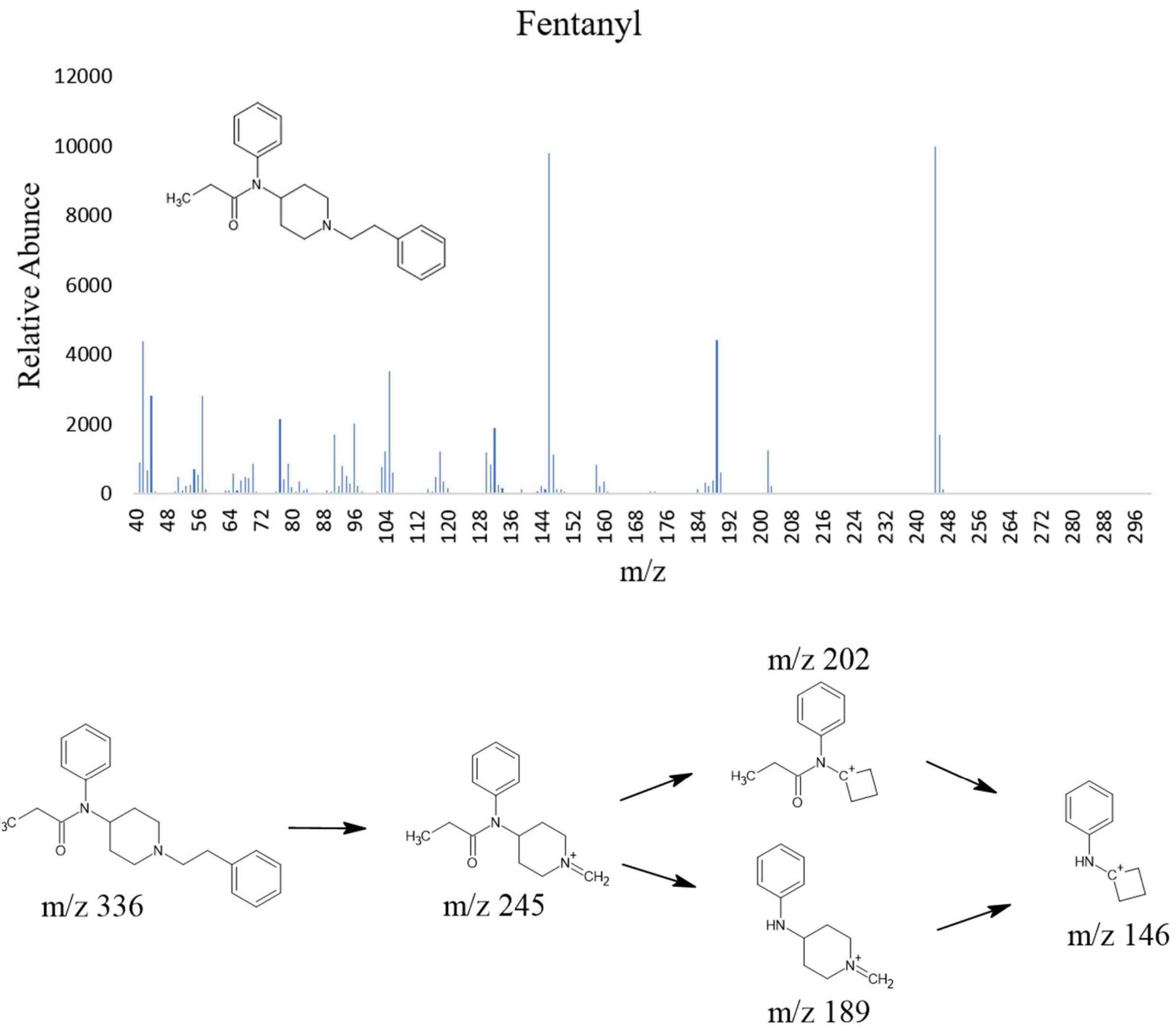

Figure 1: Mass spectrum and common fragmentation pattern of fentanyl[1,10].

Gas Chromatography

Gas chromatography is the most widely used instrument in forensic drug analysis. The gas chromatograph can separate compounds in a mixture based on their polarity and boiling point. When a mixture of chemical compounds is injected into the gas chromatograph, it is constantly transported through a capillary column by an inert gas called the carrier gas. The molecules in the sample interact with a stationary phase either coated or packed in the column. Based on these interactions' the molecules separate and reach a detector located at the end of the capillary column. A tandem instrument like a GCMS will then pass each compound into a mass spectrometer.[9,10]

Mass Spectrometry

A mass spectrometer contains three parts: the ionization source, the mass analyzer, and the detector. When a sample enters the instrument, it is ionized in the ionization source. Ionization sources can be hard, which will generally fragment the analyte molecule into many pieces, or soft, which will generally have less fragmentation. All spectra in this study were obtained using a hard ionization technique called electron ionization. During this process, the analyte is bombarded with high energy electrons that will knock an electron off the analyte. The sample compound will then break a bond and turn into fragments; one with a positive charge, and a neutral radical fragment. The location of this bond breakage is most likely to be the weakest bond on the molecule, but several bonds will break following predictable trends. After being ionized, the charged fragment is propelled into the mass analyzer, where it is subjected to a magnetic field. This field deflects the fragment off its original path based on its mass to charge ratio. The magnetic field can be manipulated so that the fragments interact with the detector in order of their mass to charge ratio. This detector will measure how many of each fragment interacts with it to create a mass spectrum. Most fragments will have a positive charge of one, therefore, the spectrum is generally only based on the mass of each fragment. The most common or intense fragment, which is known as the base peak, is assigned a value for abundance, usually either 100 or 1000, and all the other fragments are given relative abundance values based on their ratio to the most common fragment. Since a molecule will fragment the same way when put under the same ionization conditions, these spectra can be used to identify a single compound.

An unknown compound can be identified using mass spectral data by comparing its mass spectrum with a library of spectra of known compounds.[9,10]

Library

Mass spectral data used in this study was obtained from the Scientific Working Group for Seized Drug Identification (SWGDrug). This library is comprised of mass spectra of the known drugs, metabolites and other drug related compounds. All spectra of this library were collected using electron impact ionization (EI) and it is continuously updated to include novel compounds reported from various sources.[11]

PCA

The PCA is an unsupervised learning method that is used to reduce the dimensionality of complex data. In PCA, the covariance of each possible combination of variables is calculated in a matrix called the covariance matrix. This matrix will have both dimensions equal to the number of variables in a set. In the case of a mass spectrum, each mass to charge ratio would be considered one variable. The eigenvectors of this matrix correspond with the axes in the plot that contain the most variance in the data. This means that a line can be drawn through the data to represent the information contained in it using less variables. This is referred to as a principal component. These lines are orthogonal to each other, so that the information contained in one never overlaps with another. The principal components are ranked by how much information they contain. Ideally, the classes of the resulting graph from PCA should show samples from one class grouping together and separate from the other classes.[12,13,14]

$$cov_{x,y} = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{N-1} \tag{1}$$

PLS-DA

Like PCA, PLS-DA is also a dimensionality reduction tool that can be used as a discriminant analysis to predict class membership of a dataset. A significant difference in PCA and PLS-DA is that PCA is an unsupervised process, and PLS-DA is a supervised process. PLS-DA uses the class of the known compounds when making a model. PLS relates two matrices of information where one is an independent variable (like mass spectra), and one is a dependent variable (class in this case). These two sets of information can be related by

$$M = SX \tag{2}$$

$$C = UY \tag{3}$$

Where *M* is a matrix made up of the independent variable, *C* is a matrix made up of the dependent variable, *S* and *U* are score matrices similar to those calculated in PCA, and *X* and *Y* are matrices of the loading vectors. PLS-DA differs from partial least square (PLS) in that the dependent variable, *C*, is a categorical class defined by the user. Two different forms of PLS were used in this study, known as PLS1 and PLS2. PLS1 is used when *C* has only two values (when only one class is being discriminated), and PLS2 is used to expand *C* to contain more than two values.[15] The loading vectors of these models can be plotted to observe where the model assigns value. The *x* loading vector can be plotted to show which variables (m/z in this study) carry the most weight in determining the discrimination value for whichever class they are used in (see Appendix A). The *y* loading vectors can be plotted to show which *x* loading vectors are used to calculate discrimination values for each class (see Appendix B).

This has made PLS-DA models more accurate, but caution should be exercised when choosing what drug classes should be used to avoid overfitting the data. In the PLS-DA analysis.

This allows variable selection programs to accurately judge their own fitness. Using PLS-DA, DA prediction plots can be generated for each class, and any sample that is identified as that class will show up above the determination threshold is considered correctly classified with the class of interest. Cross validation was performed on each model in this study to ensure their accuracy. This was performed using the venetian blinds method, which divides the model samples into *n* number of windows then takes every n^{th} window for validation. The models in this study were all validated using this method with 20 windows.

The classification models developed using PLS-DA were evaluated by using true positive rate (*TPR*) and true negative rate (*TNR*). TPR is the proportion of positive samples that were correctly identified by the model and TNR is the proportion of negatives samples that were classified correctly as negative samples.

These values were calculated using following equations after building the PLS-DA models,

$$TPR = \frac{TP}{(TP + FN)} \tag{4}$$

$$TNR = \frac{TN}{(TN + FP)} \tag{5}$$

where *TP* is true positives, *FN* is false negatives, *TN* is true negatives, and *FP* is false positives. The true positive rate is a measure of how a test can predict trues positives, and the true negative rate predicts how well a test can predict true negatives. When determining what class, a sample belongs to, PLS-DA makes use of several loading vectors. These are like the principal components in PCA in that they show which variables in each sample are important for determining a specific class. In this study, variables correspond to each mass to charge ratio on a mass spectrum. Models using PLS1 in this study were calculated using 8 loading vectors, while PLS2 models were calculated using 15 loading vectors. Using the *y* loading vectors, one can determine which *x* loading vectors contribute to each class, then use the *x* loading vectors to determine which variables are important.[12,13,14,16]

Variable Selection

Variable selection methods choose specific variables in data based on their importance in the model's fitness. This can be helpful when samples contain substantial amounts of information that may not all be important to the purpose of the model. In this study, every sample is a mass spectrum, and not every mass to charge ratio (m/z) helps identify the class of a drug. In order to narrow the scope of the model to useful information, a genetic algorithm (GA) was used.

Genetic algorithm uses concepts of natural selection to select the most useful variables in a dataset to maximize the performance of the models. The algorithm will first create several models to make up the first generation, then it will give each model an accuracy value for its fitness to the correct classes of the samples. The next generation of models will utilize useful variables for the fitness from the prior generation. Mutation and crossover are used to introduce new variables to the problem and improve the model performance. The algorithm will then assess the fit to see if that generation was an improvement, before moving to the next generation. This process continues until the fitness reaches a set value or a maximum number of generations. When variable selection is finished, the model is calculated with only those peaks in the spectrum that the GA identified as important. Then using a variable selection method like GA, it is important to avoid overfitting the model. A technique like this can easily remove a variable from the study that may contain useful information for the problem at hand. Using less generations, and performing replicate runs of GA can alleviate this problem. Each generation of GA was performed using a widow width of one, population size of 30% of the initial terms (m/z), and a mutation rate of 0.005. Each PLS-DA model used 15 loading vectors and the initial population included 64 PLS-DA models. This GA routine was performed on 80 maximum generations and three replicate runs were completed for each model. Using GA in this study will force the model to only use the m/z in the spectra that are important for classification of a drug.

Hierarchical Clustering

Ward's method is a hierarchal clustering can be used to generate groups from data. This can be applied to large amount of data to create a classification without prior knowledge about the samples. The method is an agglomerative method, meaning that each sample starts the process as its own cluster. These clusters are then combined into larger clusters in such a way that the variance within the cluster is minimized. In this study, Ward's method was performed on the PCA scores of the data using four principal components. This can be used to create new classes for use in PLS-DA without having to manually define them.[8]

Research Objectives

This research work focuses on providing a proactive solution to identify newly designed NPSs when the reference spectra are not present in the spectral libraries used in forensic laboratories. A classification system is developed using existing data of known substances that can be used to determine the presence of NPSs. The initial work is designed to classify major drug classes in order to identify the class that an unknown belongs to, for example if an unknown drug is a fentanyl derivative or not. The second half of this work if focused on extending this classification system to identify the position of the derivatization on the fentanyl molecule. Fentanyl related compounds are divided to three groups based on the structural position of the modifications.

Starting with a library of known compounds with known classes, one can calculate an initial model in order to divide the mass spectral library into different drug classes. These classes can then be further divided into subclasses by use of another calculated sub-model. Use of sub-models allows one a great amount of liberty to decide how specific of a structural difference is being classified. Once this model framework is calculated with the known compounds, an unknown structure can be inserted into this framework to place it into one of the predetermined classes (see Figure 2).

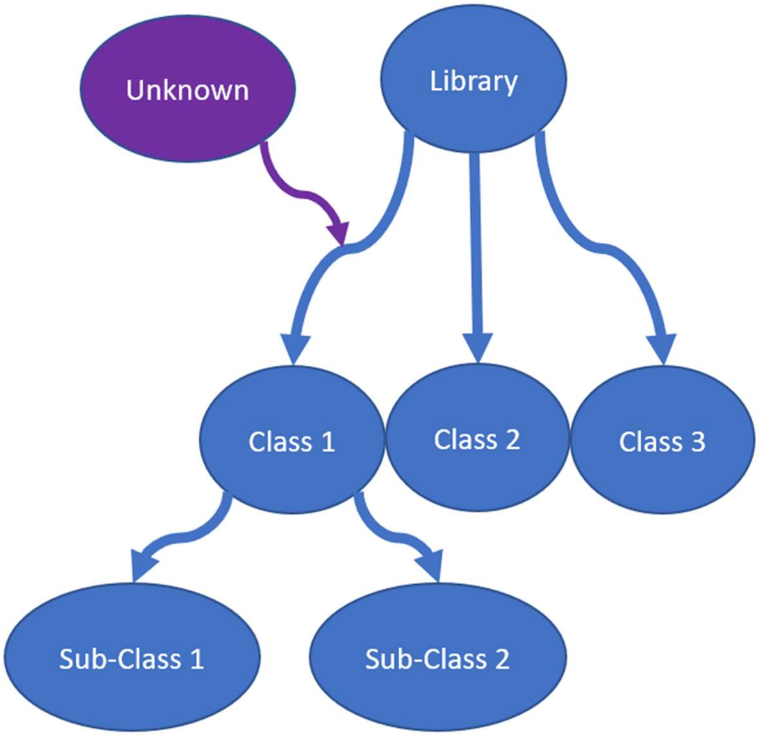

Figure 2: General scope of hierarchal model.

CHAPTER 2: EXPERIMENTAL

Mass Spectral Data

Data for this study was acquired from the scientific working group for the analysis of seized drugs (SWGDrug). A total of 102 mass spectra were used in the feasibility study to determine if there is sufficient variability among different drug groups present to build a classification system. A total of 474 mass spectra were used in the next step, and these spectra represent eight drug classes. (See Table 1). These mass spectra are comprised of mass to charge ratios (m/z) and the relative abundance.

Table 1: Number of samples in each model and class in each study.

Model	Initial PCA	Second PCA/ Initial PLS-DA	Second PLS-DA/ Binary Problem	Fentanyl Sub-Model	
Number of Samples	102	474	498	173	
Fentanyl Derivatives	15	194	194	Class 1	92
Cannabinoids	15	156	156		
Opioids	9	27	27	Class 2	48
Cathinones	15	49	121		
Tryptamines	7	7			
Phenethylamines	14	14		Class 3	33
Steroids	15	15	0		
Barbiturates	12	12	0		

Data Pre-Processing

The data from the library first had to be made into a uniform shape. This meant adding zeroes to the shorter spectra until their dimensions matched the longest spectrum. Once the initial analysis on this was done, it was found that there was a specific range of mass to charge ratios that were important in modeling, and other peaks were lowering the model's accuracy. This was remedied by cropping the spectra to the helpful range before further analysis. Originally, the spectra were measured from 11 m/z to 662 m/z, but this was found to be too many variables to make an effective model. The useful information in a spectrum for class identification in this setting is a smaller range of mass to charge ratios. The data was cropped from 40 m/z to 300 m/z to remove the less helpful variables from model calculation and reduce the variance in the model to a more manageable size.

Software

MATLAB is a software commonly used in science for data analysis. Short for MATrix LABoratory, MATLAB works well for analyzing large amounts of data in a matrix. MATLAB can also be used to visualize data in figures with its inbuilt plotting software.

In MATLAB, a toolbox called PLS toolbox was used for all PCA, PLS-DA, and GA analysis. This toolbox allows one to perform several analysis methods on a single data set while simplifying the process through a GUI and saving each model as it is created.

Data Analysis Methods

Two machine learning techniques are used in this study to create models to differentiate drugs based on their mass spectra: principal component analysis (PCA), and partial least squares discriminate analysis (PLS-DA). The goal of this study is to build a classification model to be used to identify the class of an unknown drug using mass spectral data. Both PCA and PLS-DA decompose substantial amounts of data into more manageable pieces by assigning them variance values to contain the same information in less variables. The resulting graph from PCA should show the different classes of drug grouping together separate from the other classes. In this study, an initial analysis was performed to determine if discrimination was possible using this technique.

Using PLS-DA, the program gives different graphs for each class, and any sample that is identified as that class will show up above the determination threshold. PLS-DA was used after initial analysis to more accurately separate classes in the study. Cross validation was performed on each model in this study to ensure their accuracy. This was performed using the venetian blinds method, which divides the model samples into *n* number of windows then takes every *n*th window for validation. The models in this study were all validated using this method with 20 windows.

CHAPTER THREE: RESULTS AND DISCUSSION

Initial PCA

The mass spectra from the SWGDrug library were converted to a comma separated value (.csv) file and imported into MATLAB workspace for analysis. To determine if there was enough variance in the data to differentiate these drugs, a small number of compounds (7 to 15) were identified from eight drug classes including fentanyl derivatives, cannabinoids, phenethylamines, cathinones, tryptamines, barbiturates, steroids, and opioids with no fentanyl derivatives and an initial PCA was performed. The drugs representing each group were chosen based on the classifications used by Drug Enforcement Administration (DEA), World Health Organization (WHO) and Federal Bureau of investigation (FBI) literature. PC score plots are usually constructed using the first two principal components, since they contain the most information, or variance of the original data. More dimensions can be used, but it was decided that the result of two PCs provides necessary information for this study. While it was not expected that this initial PCA be able to classify each compound, some initial grouping should be present to support further work.

The results of this analysis show some separation of data suggesting that there was enough variance between classes of drugs for further analysis. It was hypothesized that there might not be enough representatives or samples in each class with enough spectral information to effectively separate the classes. First, more samples were identified from the SWGDRUG library and another round of PCA was performed to see if the added samples improved the variance in the study.

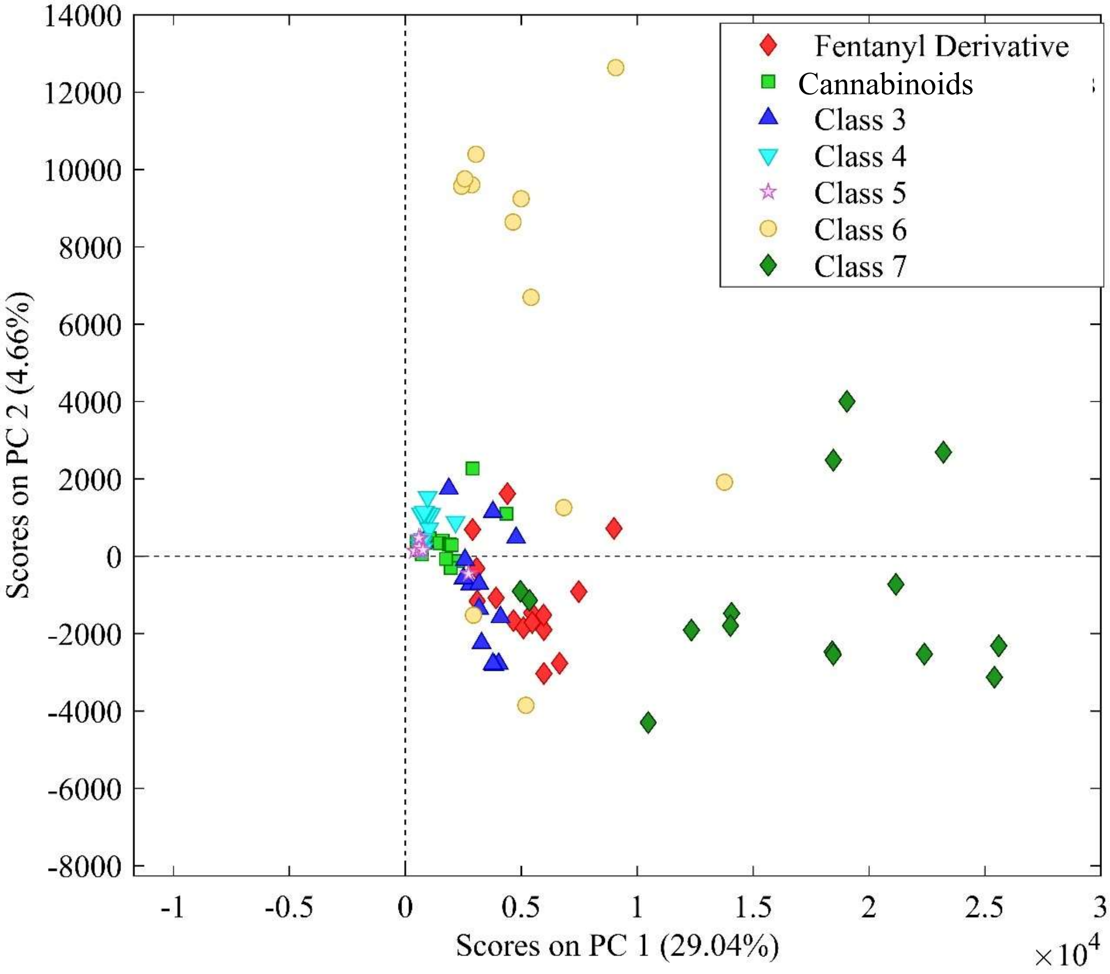

Figure 3: PCA score plot of 102 compounds on PC 1 vs. scores on PC 2. Evidence of grouping and limited separation implies that further work can be done to classify samples.

Second PCA

More samples were added for each drug class except phenethylamines, steroids, tryptamines, and barbiturates to construct second PCA scores plot (see Table 1.). More samples increase the variance within each group, causing them to spread out on the plot, but it will also increase variance between the classes as more representative compounds are added to each class. Special attention was given to synthetic cannabinoids and derivatives of fentanyl, as they are currently the most common NPSs encountered thus, more samples were included in this study. This PCA showed a greater correlation within each class, but not enough variance to cluster them in different areas of the PCA scores plot. Although, some clustering of data is visible, significant overlap between classes is present.

Fentanyl derivatives, cannabinoids, steroids, and opioids show significant grouping while the drug classes barbiturates, cathinones, phenethylamines, and tryptamines do not show a separation. This can be due to two reasons. First, the drug classes barbiturates, cathinones, phenethylamines, and tryptamines share common chemical groups such as phenyl, amine, short alkyl chains, and contain smaller drug molecules compared to fentanyl derivatives, opioids, cannabinoids, and steroids. Second, drug classes barbiturates, cathinones, phenethylamines, and tryptamines contain small number of samples compared to fentanyl derivatives and cannabinoids, and therefore, the spectral information related to fentanyl derivatives and cannabinoids are well represented in the data space.

Initial PLS-DA

Initial PLS-DA was performed for all the samples that are used in the second PCA study. Since each class is defined, and every sample must be put in a class before a model can be made in PLS-DA, the model can specifically use the variance between classes to discriminate them.

Each compound in the study was assigned a class for the model based on their structure, then the first PLS-DA analysis was performed using seven classes of drugs. In a PLS-DA model, DA prediction values are plotted per sample, and a threshold is set for a positive identification. If a prediction value falls above that threshold (above the red line in Figure 4), then it is considered a part of the class for which that plot was generated. In the case of a sample scoring above the threshold of multiple classes, the class that had the higher prediction value is chosen for that sample.

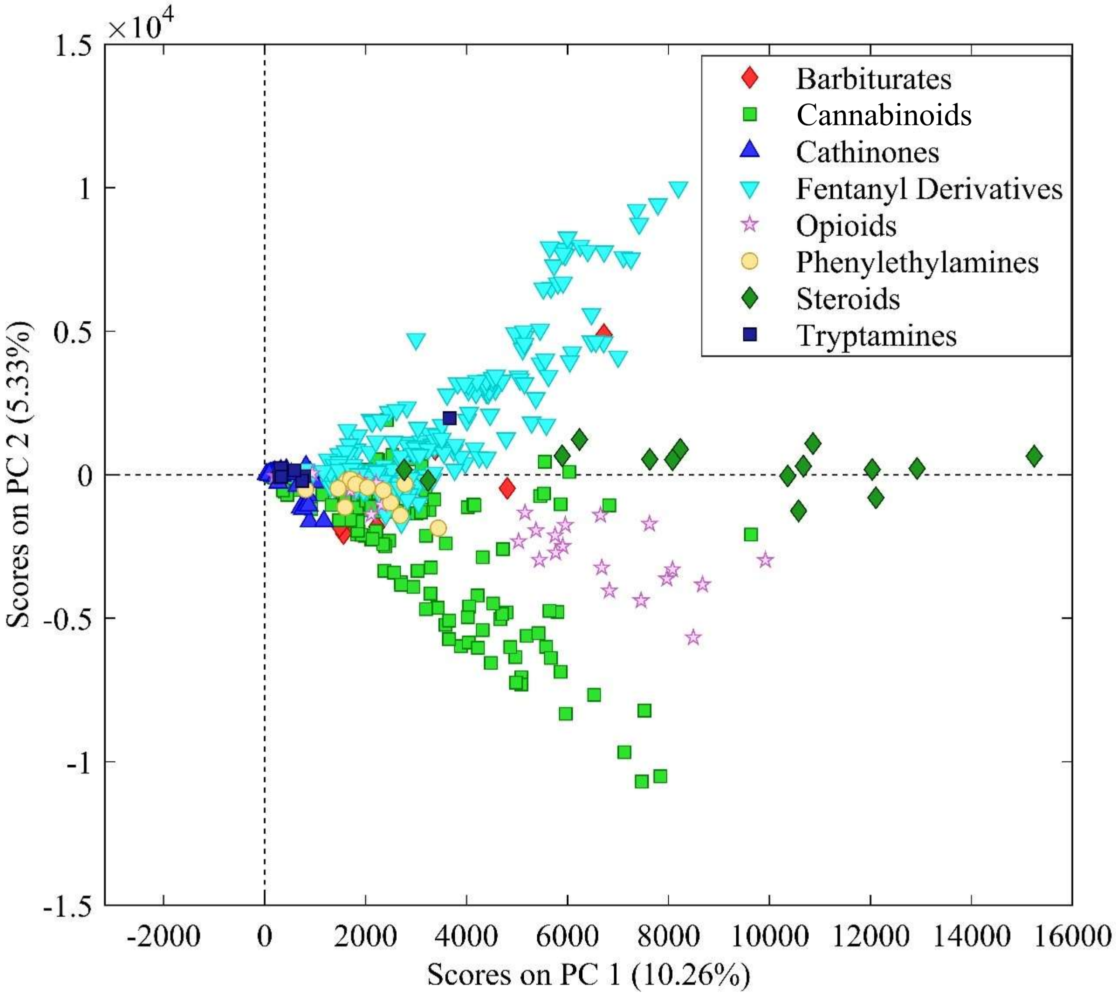

Figure 4: PCA score plot of 474 compounds. More group correlation and significant overlap is evident.

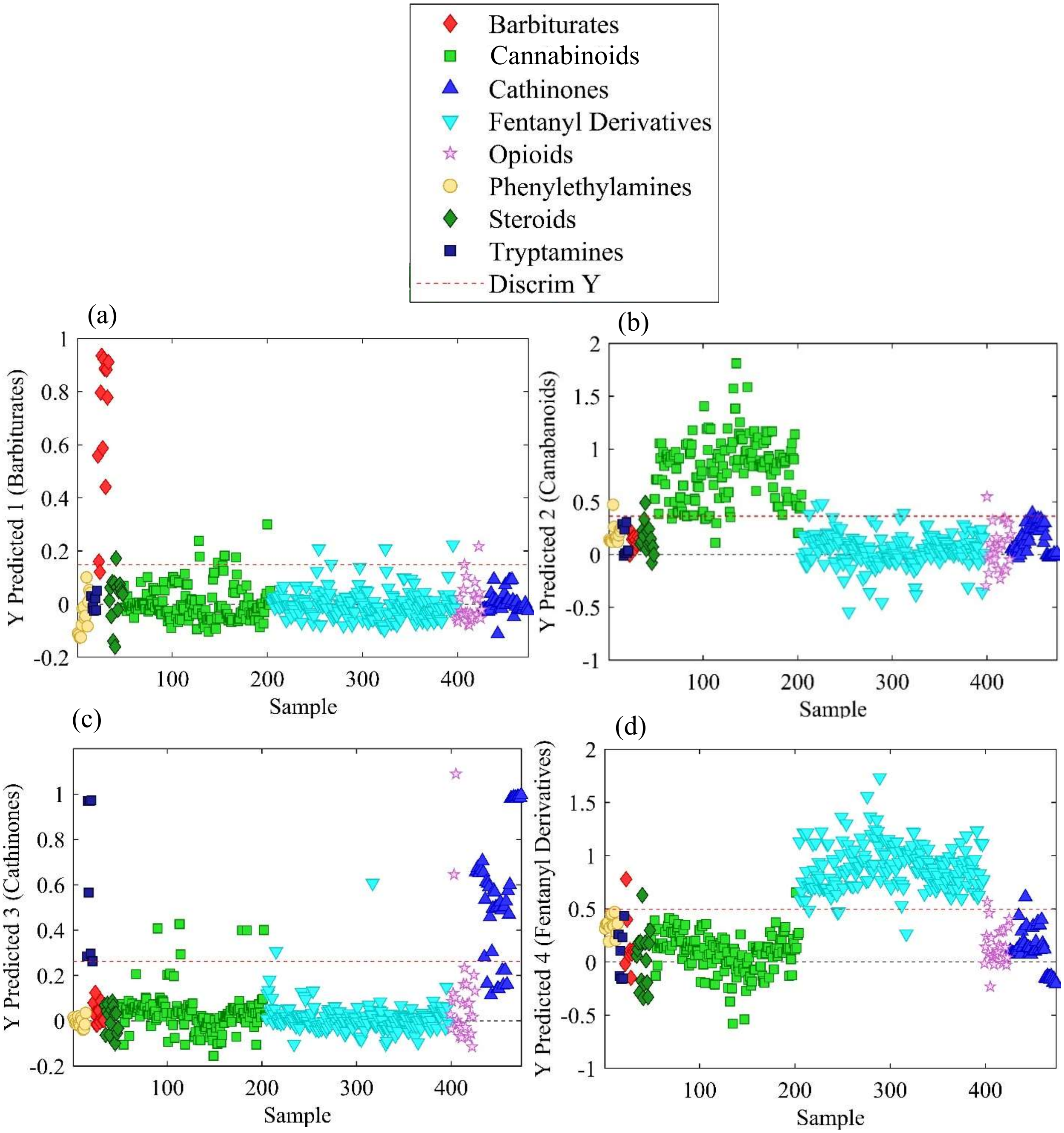

Figure 5: Initial PLS-DA results plotted prediction values vs. sample number for 474 compounds. Discrimination of Cannabinoids and Fentanyl Derivatives is observed, more work is required to separate smaller groups including Phenylethylamines and Tryptamines

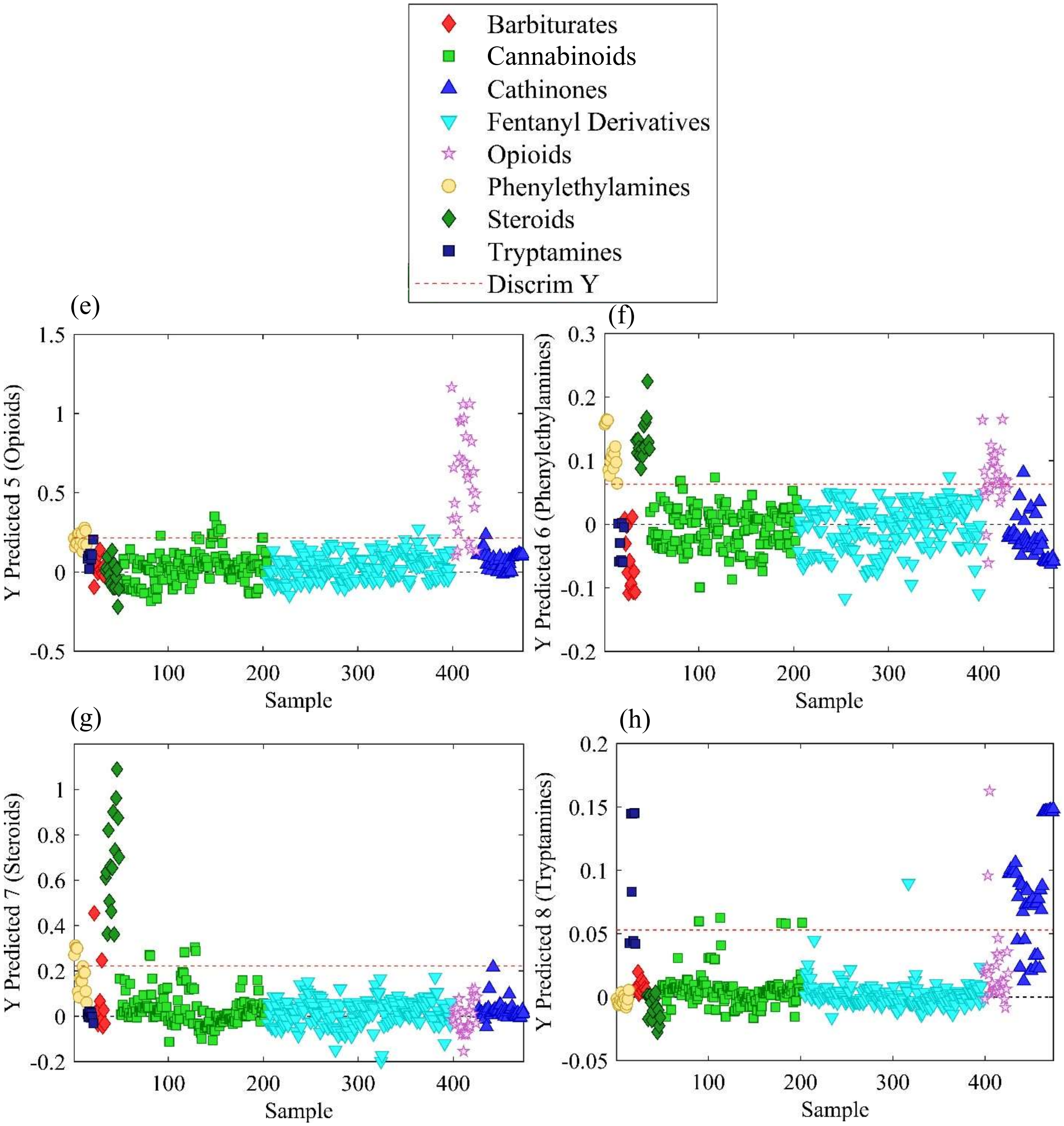

Figure 5 Cont.: Initial PLS-DA results plotted prediction values vs. sample number for 474 compounds. Discrimination of Cannabinoids and Fentanyl Derivatives is observed, more work is required to separate smaller groups including Phenylethylamines and Tryptamines

The initial PLS-DA model classified many samples incorrectly (See Figure 5). Most cannabinoids (green) and fentanyl derivatives (blue) are correctly classified or projected above the threshold line in their corresponding PLS-DA plots (see Figure 5. (b) and (d)) with only a few false positives and false negatives. All other drug classes show a significant number of false positives or false negatives in corresponding DA plots. This may have been due to unnecessary peaks and extra m/z ratios that had no discriminatory value that were included in the mass spectral data. The original mass spectra used in this initial PLS-DA included m/z ratios from 11 to 605, and all sample spectra were cropped to include only the m/z values from 40 to 300 that were found to contain the most useful information of each class. This eliminates the unnecessary peaks such as peaks for small fragments and large peaks such as the molecular ion peak that have no discriminatory value. Since the molecular mass of a given compound plays no or little role related to the drug class, the molecular ion peak is not helpful in this analysis.

To improve the model further, more samples were added to each class to increase the performance of the model (see table 1.). Barbiturate and steroid drug classes were problematic in the classification problem. Since the compounds belong to the steroid class tend to be larger than the other compounds, their mass spectra contain many peaks that can be present in other classes but are not representative of that drug class. There were a large number of extraneous peaks, that the variable selection (GA) was not able to remove before reaching its maximum number of generations (data not shown). This was causing the model to include these peaks in the data space and misclassify large number of compounds.

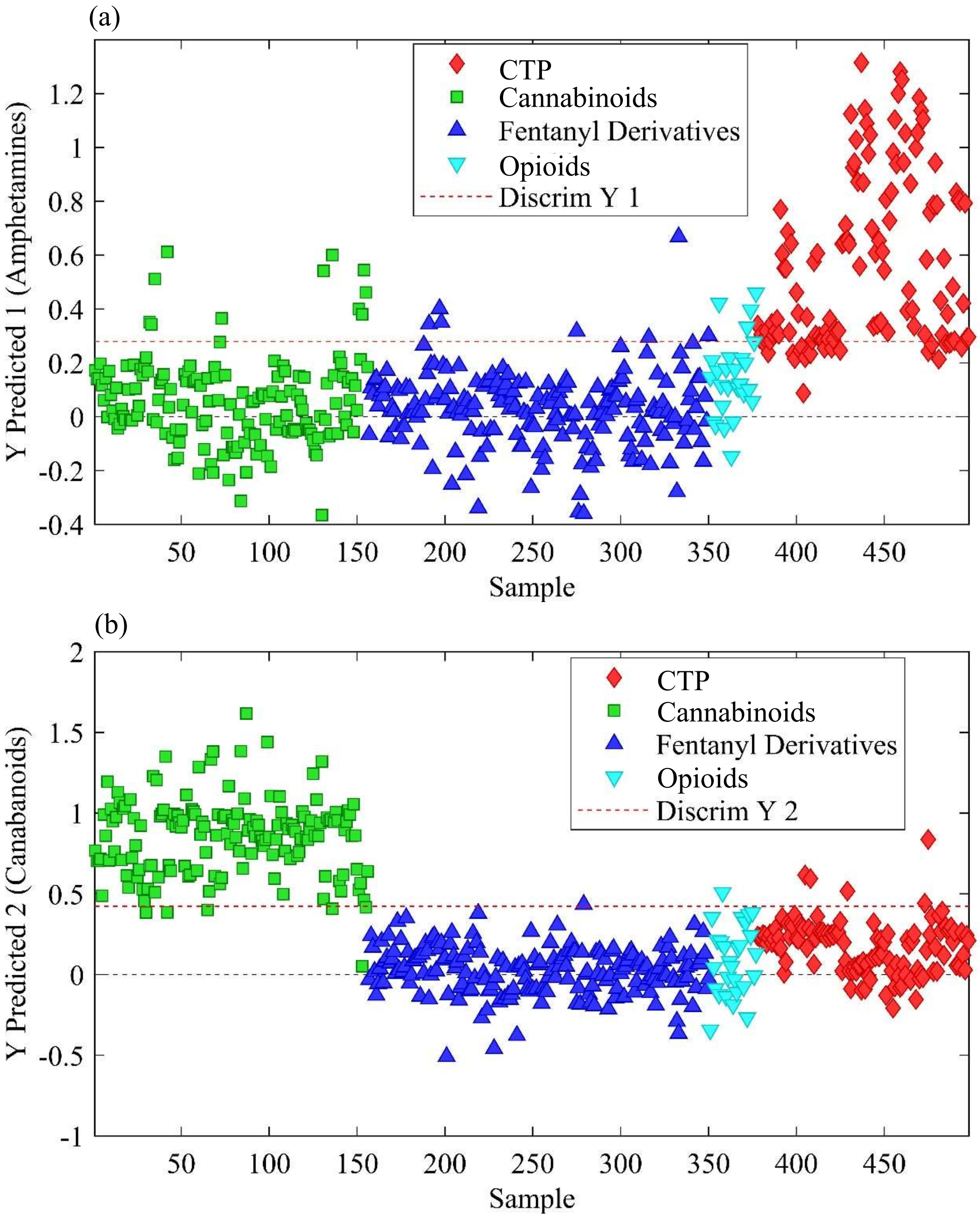

Figure 6: PLS-DA of 498 compounds using reduced number of classes. Large amount of spreading in prediction values can be seen in CTP class and opioid class.

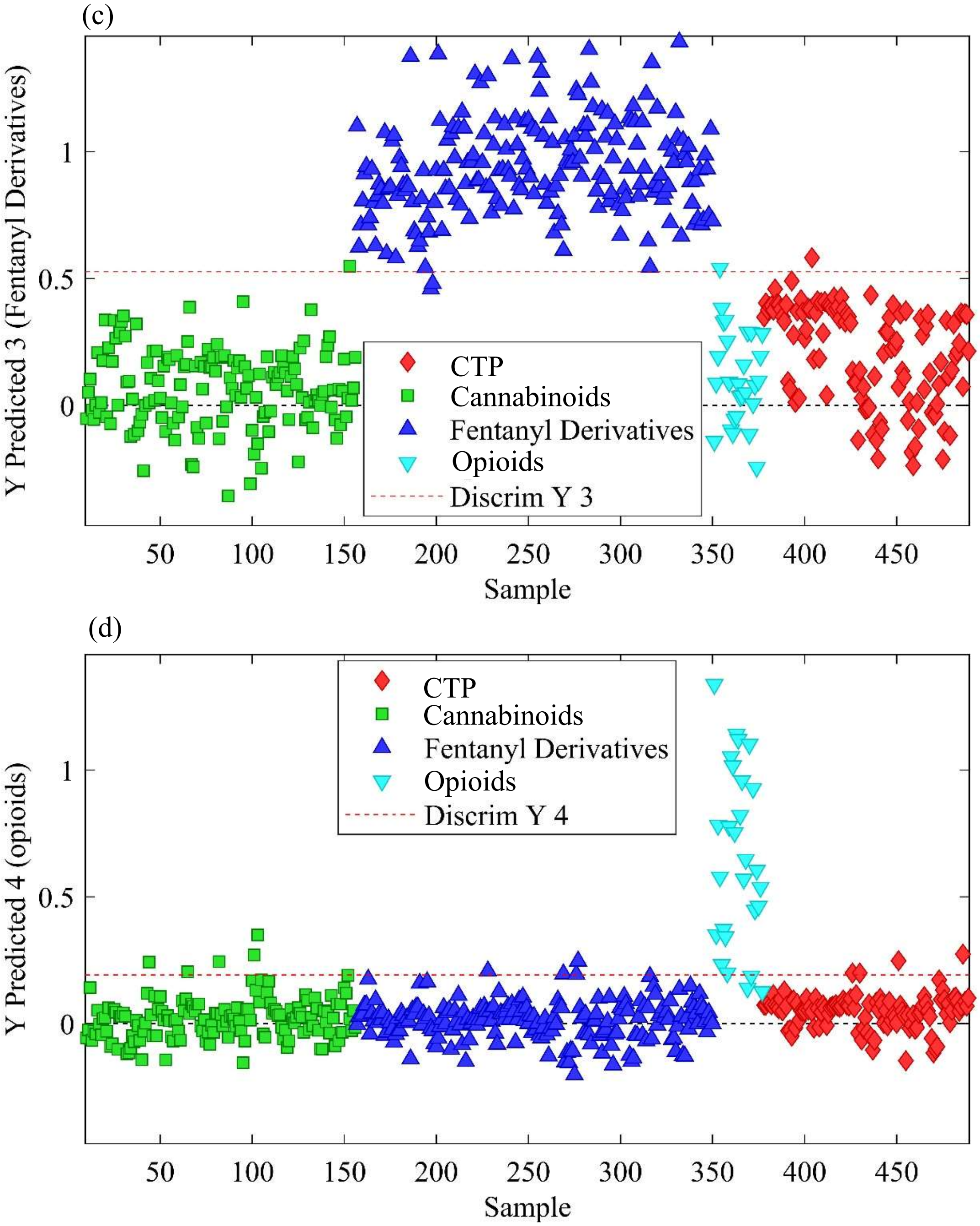

Figure 6 Cont.: PLS-DA of 498 compounds using reduced number of classes. Large amount of spreading in prediction values can be seen in CTP class and opioid class.

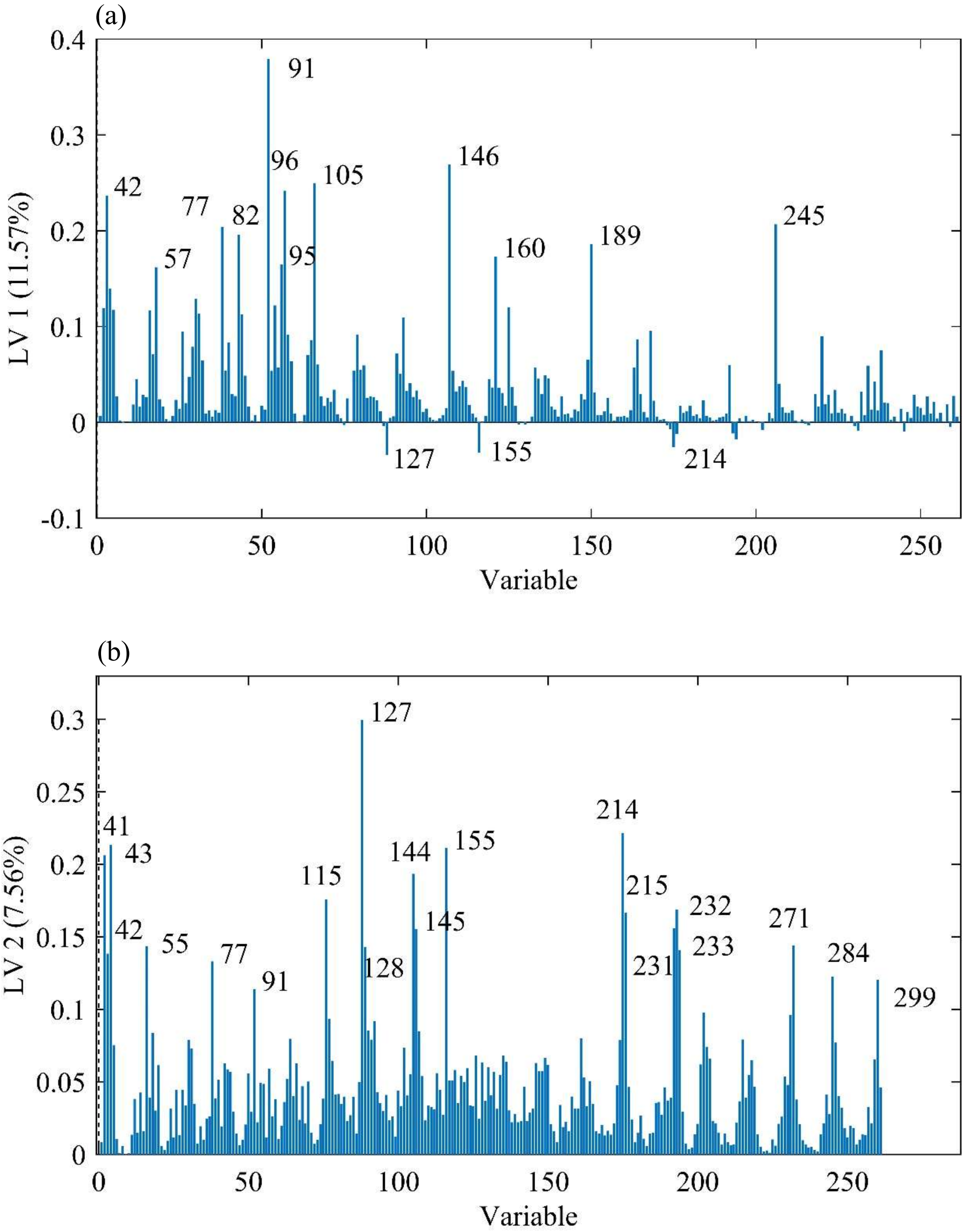

Figure 7: First two loading vectors for the DA plots of Figure 6. (a) LV 1, positive peaks correspond to m/z values that are responsible for classification in the fentanyl derivative class. (b) LV 2, positive peaks correspond to m/z values that are responsible for classification in the cannabinoid class.

Additionally, during a typical drug analysis, acidic drugs and basic drugs are extracted using two separate extraction procedures using acidic and basic pH buffers. Therefore, it is not possible to encounter an acidic drug in a basic drug analysis process and thus, it is not required to retain them in this classification problem. Due to this reason, steroids and barbiturates were removed from further studies. Although cannabinoids share the same chemical properties, they were used in further models to understand the performance of this method. Additionally, the cathinone, tryptamine, and phenethylamine drug classes from the previous model were combined to form the cathinones, tryptamines, and phenethylamines (CTP) class.[1]

The results of this PLS-DA analysis show improved discrimination between classes (see Figure 6.). Reducing the number of classes helped the model differentiate fentanyl derivatives from other drugs included in the study while small number of samples are misclassified compared to the previous model (see Figure 6). To understand the mass spectral features that are important for this discrimination, loading vectors (LVs) were analyzed. Each LV shows the loadings of features (variables) that on individual spectra will cause a sample to be placed within a drug class. The fentanyl derivative class (Figure 6 (c)) is dependent on positive peaks on LV 1 (Figure 7. (a)), and negative peaks on LV 2 (Figure 7. (b)) meaning that a presence of peaks that have higher values in LV 1 and the absence of peaks that have higher values of LV 2 in a mass spectrum of a drug sample will be resulted in classifying that sample as a fentanyl derivative. In LV 1, the positive peaks include common fragments found in fentanyl related compounds such as m/z = 91, m/z = 146, and m/z = 245, the base peak of fentanyl (see Figure 2). Positive peaks on LV 2 serve to remove samples from the fentanyl derivative class include m/z = 214, 215, 231, 232 and these peaks are found to be present in mass spectra of common cannabinoids.[1]

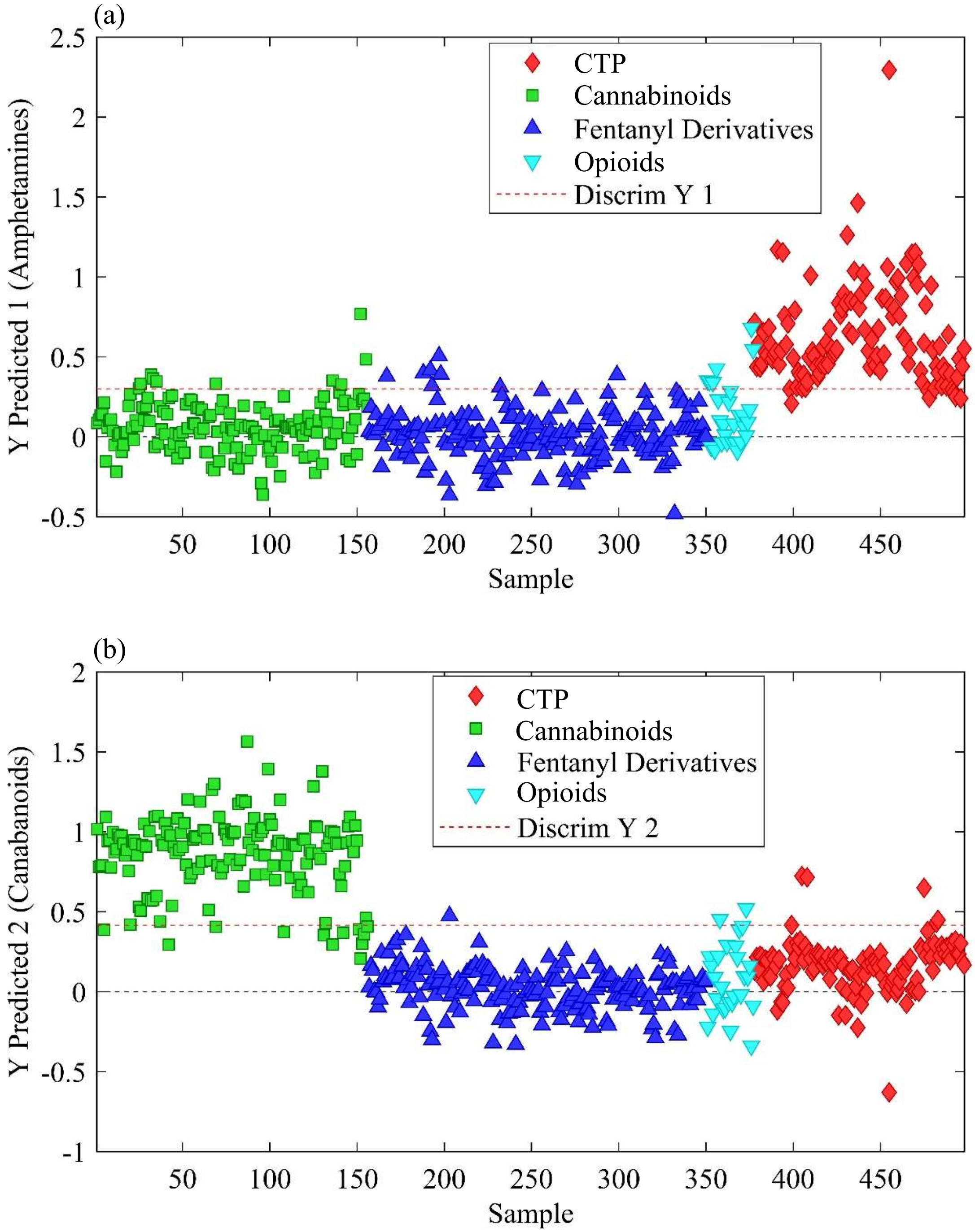

Figure 8: PLS-DA performed with reduced number of classes using GA as a variable selection method. Reduced prediction value spreading observed, but two outliers appear.

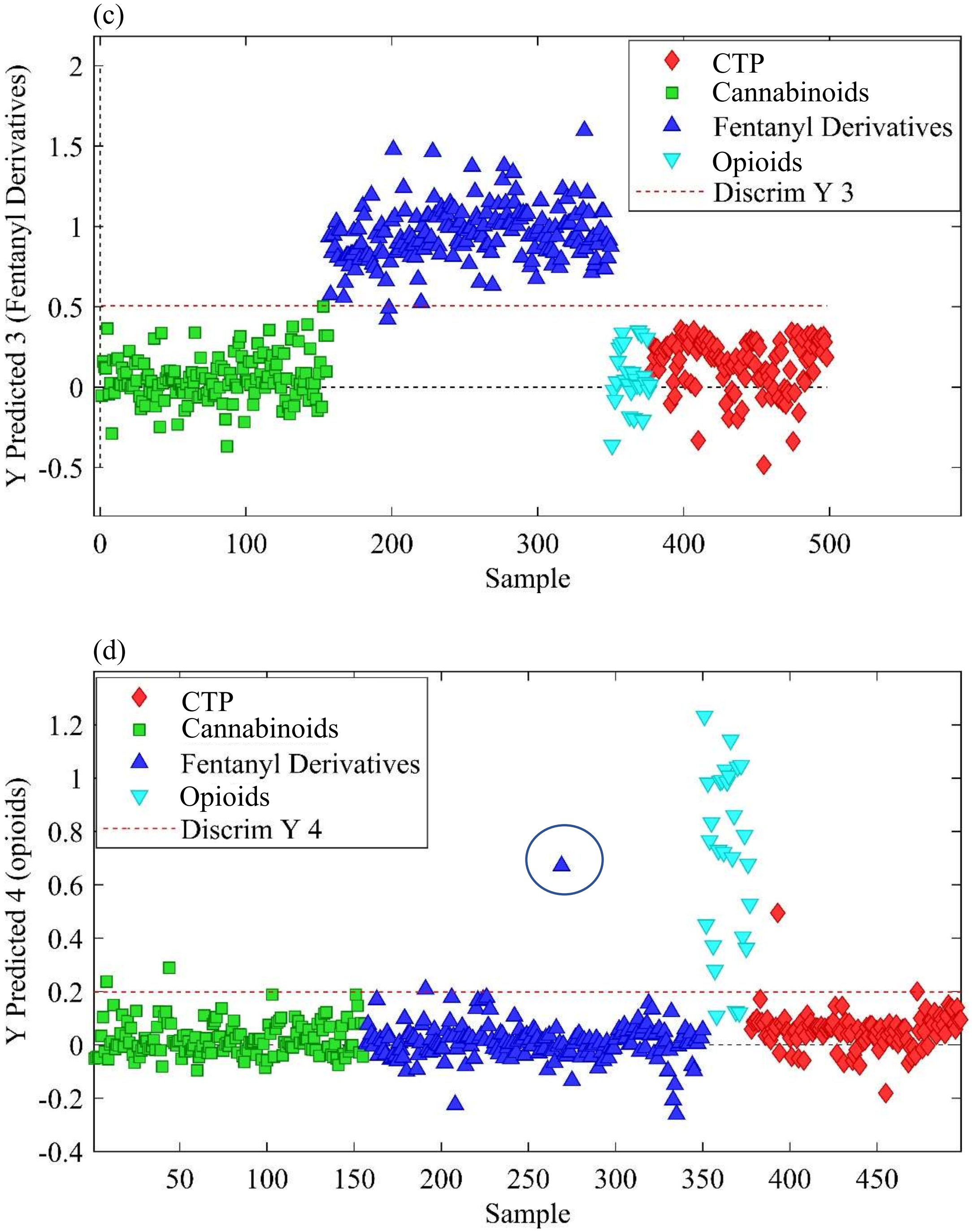

Figure 8 Cont.: PLS-DA performed with reduced number of classes using GA as a variable selection method. Reduced prediction value spreading observed, but two outliers appear.

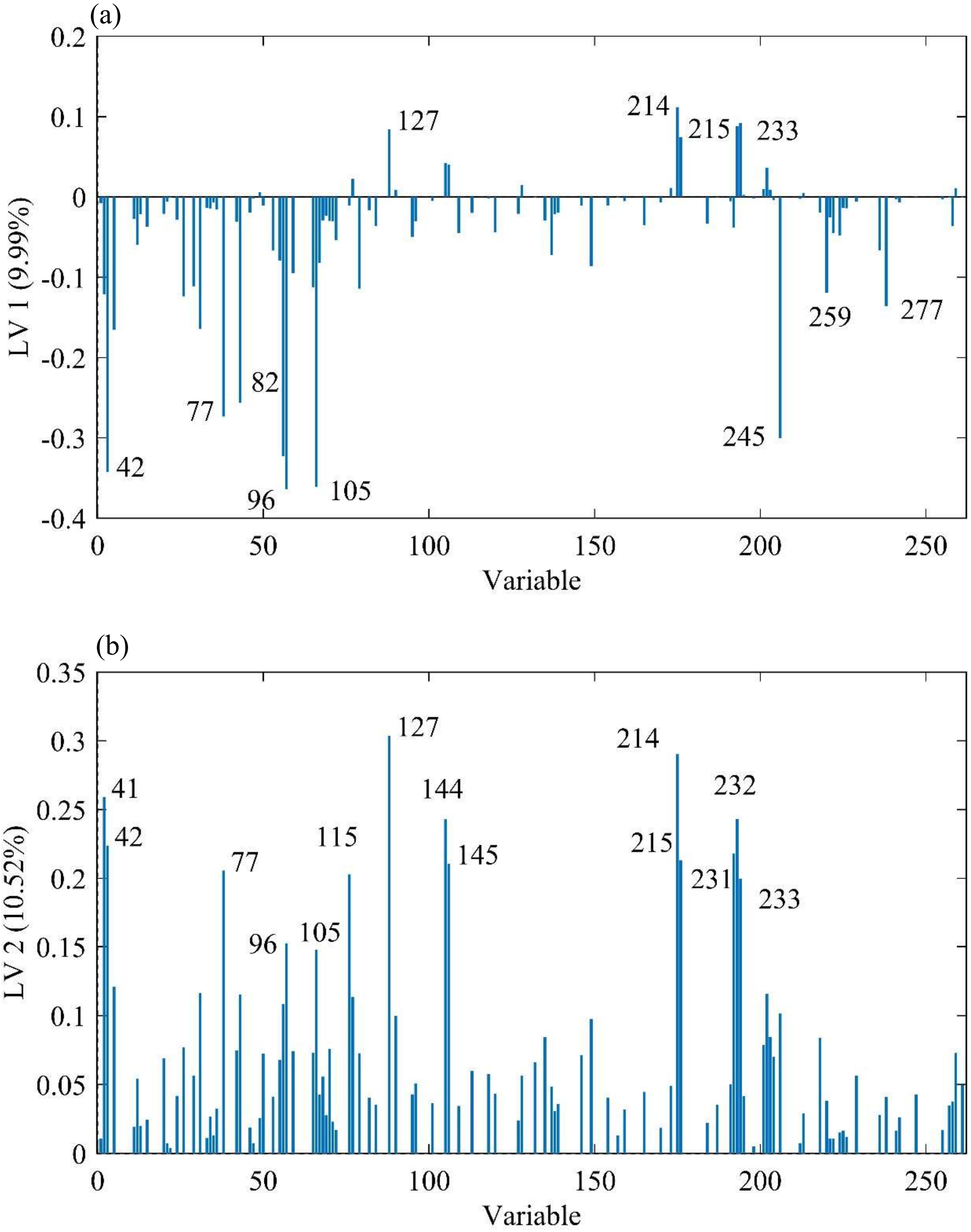

Figure 9. First two loading vectors for the DA plots of Figure 8. (a) LV 1, positive peaks correspond to m/z values that are responsible for classification in the cannabinoid class, while negative peaks correspond to the fentanyl derivative class. (b) LV 2, positive peaks correspond to m/z values that are responsible for classification in the cannabinoid class.

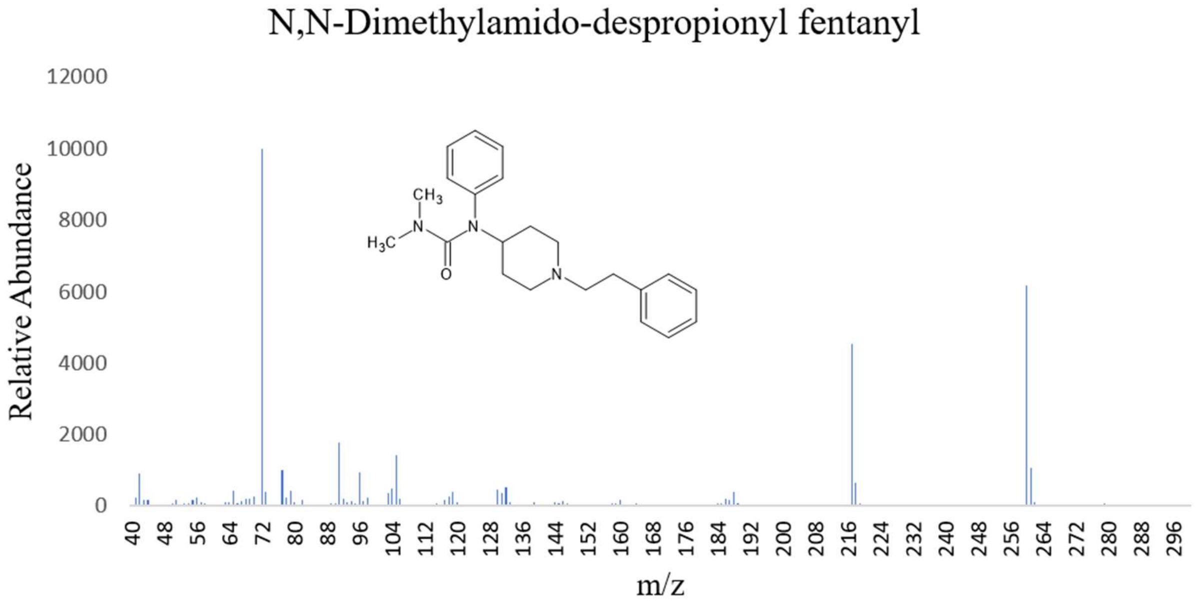

Figure 10: Mass spectra of misclassified fentanyl derivative in Figure 8. (d).

To improve the class separation, GA was used as a variable selection method. Special care was taken to ensure that the GA was not causing the model to overfit the data. The number of generations was reduced to where the fitness no longer improved significantly, and three replicate runs were performed for each model. GA improves the study by removing peaks in the mass spectrum that are not contributing to meaningful variance (see Figure 9.). Some peaks are common to specific classes, while other peaks are simply randomly distributed between the classes. These randomly distributed peaks reduce the effectiveness of the model by causing it to associate them with a class erroneously and assign them a higher prediction value in calculation.

The peak at 245 is retained for the fentanyl derivative class in LV 1 and LV 3, while the peak at 91 m/z is removed from the model, as many compounds can form fragments at that mass since toluene and tropylium are common fragments for many compounds not just fentanyl derivatives. The misclassified fentanyl derivative (circled in Figure 8. (d)) is found to be *N,N*-

Dimethylamido-despropionyl fentanyl and the mass spectrum of this compound (see Figure 10.) shows that the common mass fragments found in the fentanyl related compounds are not present. This causes the model to classify this compound as a non-fentanyl related compound. Table 2. shows that this model has a very high TPR and TNR for fentanyl derivatives (more than 0.95) and can predict other drug classes reasonably well.

Table 2: Model prediction results and cross validation prediction results before and after variable selection for PLS-DA using 4 classes.

Model Results					
Before Genetic Algorithm					
	Accuracy	TPR	TNR	FPR	FNR
CTP	0.962	0.950	0.966	0.034	0.050
Cannabinoids	0.970	0.936	0.985	0.015	0.064
Fentanyl Der.	0.992	0.985	0.997	0.003	0.015
Opioids	0.988	0.889	0.994	0.006	0.111
	After Genetic Algorithm				
	Accuracy	TPR	TNR	FPR	FNR
CTP	0.970	0.958	0.973	0.027	0.041
Cannabinoids	0.964	0.930	0.980	0.020	0.071
Fentanyl Der.	0.992	0.985	0.997	0.003	0.015
Opioids	0.986	0.889	0.992	0.008	0.111
Cross Validation Results					
Before Genetic Algorithm					
	Accuracy	TPR	TNR	FPR	FNR
CTP	0.946	0.917	0.955	0.045	0.083
Cannabinoids	0.956	0.917	0.974	0.026	0.083
Fentanyl Der.	0.984	0.969	0.993	0.007	0.031
Opioids	0.978	0.815	0.987	0.013	0.185
After Genetic Algorithm					
	Accuracy	TPR	TNR	FPR	FNR
CTP	0.950	0.934	0.955	0.045	0.066
Cannabinoids	0.950	0.904	0.971	0.029	0.096
Fentanyl Der.	0.984	0.959	1.000	0	0.041
Opioids	0.976	0.852	0.983	0.017	0.148

To simplify the problem, the analysis was modified as a binary problem, meaning the number of classes were reduced to two. All non-fentanyl drugs are grouped (as one class to in this two-class problem). The goal of this study is to investigate if fentanyl derivatives can be separated from all possible drug classes to mimic a real-life scenario. In a hypothetical analysis, if an unknown compound found during a GCMS analysis will allow an analyst to obtain a compound classification for that mass spectrum of that unknown without having to collect additional data about individual m/z peaks. This two-class model will allow the analyst to determine if that unknown is a fentanyl derivative. Several methods of validation were implemented throughout the study to estimate the robustness of the models. In this binary problem, all data was divided into a training set and a test set.

The training set was used to develop the model and the test set was used to test the model by predicting these samples with the model. This test set comprised of about ten percent of all mass spectra used in the study (5). The PLS-DA plots were first generated without applying GA and only two samples were misclassified as shown in Figure 11. All samples were correctly classified after using GA for feature selection (Figure 12). TPR and TNR for this model is more than 0.99 (see Table 3.) meaning that this can predict if an unknown drug is a fentanyl derivative or not with more than 99% accuracy. The one missed compound is shown in Figure 14; its major feature in the mass spectrum is its base peak at m/z = 278. This peak does not coincide with any peaks on the model's loading vector, so a low prediction value is given to this compound.

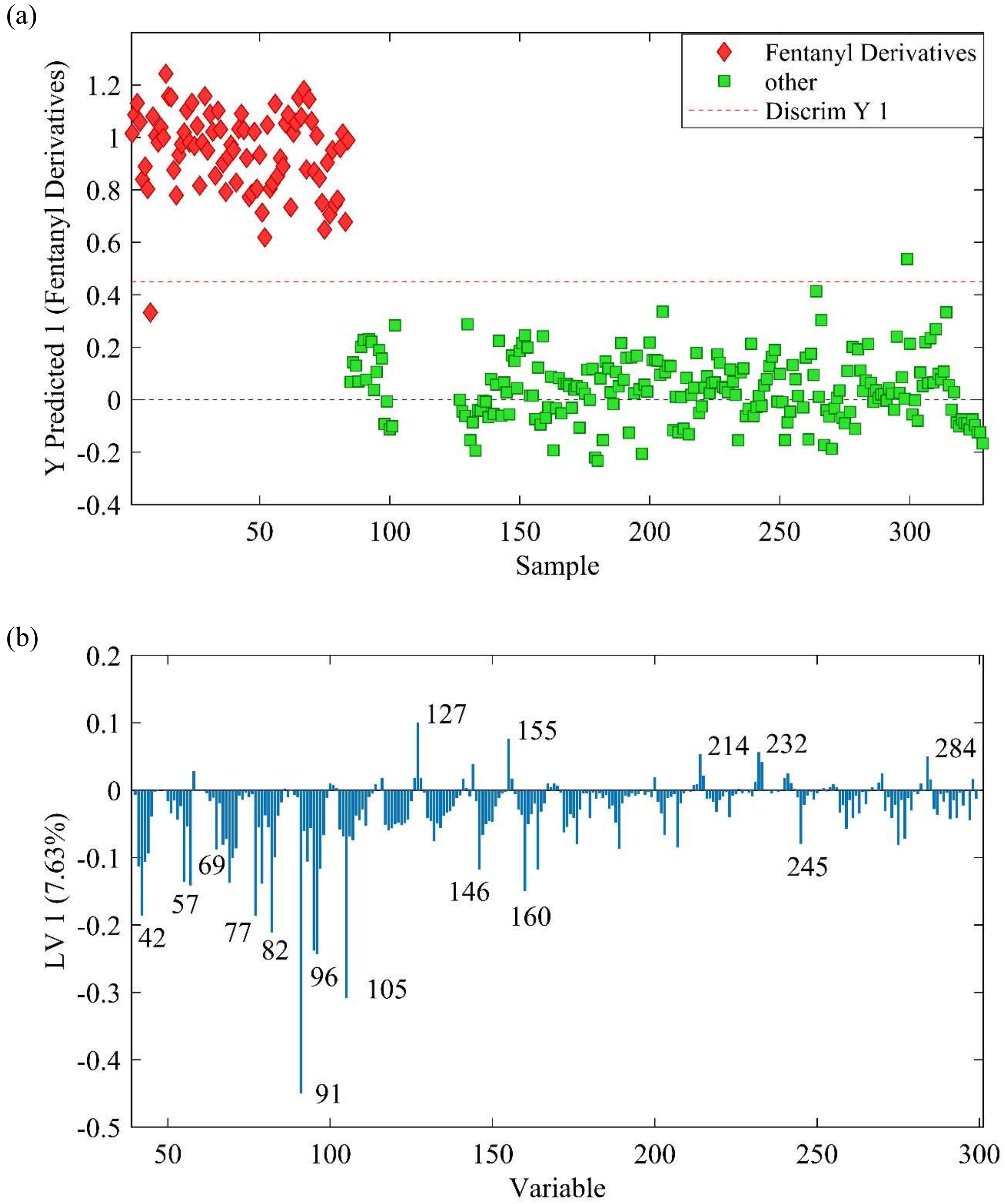

Figure 11. (a) DA plot of training set before GA for fentanyl derivatives. (b) The first LV for the model was included showing negative peaks present in fentanyl derivative class.

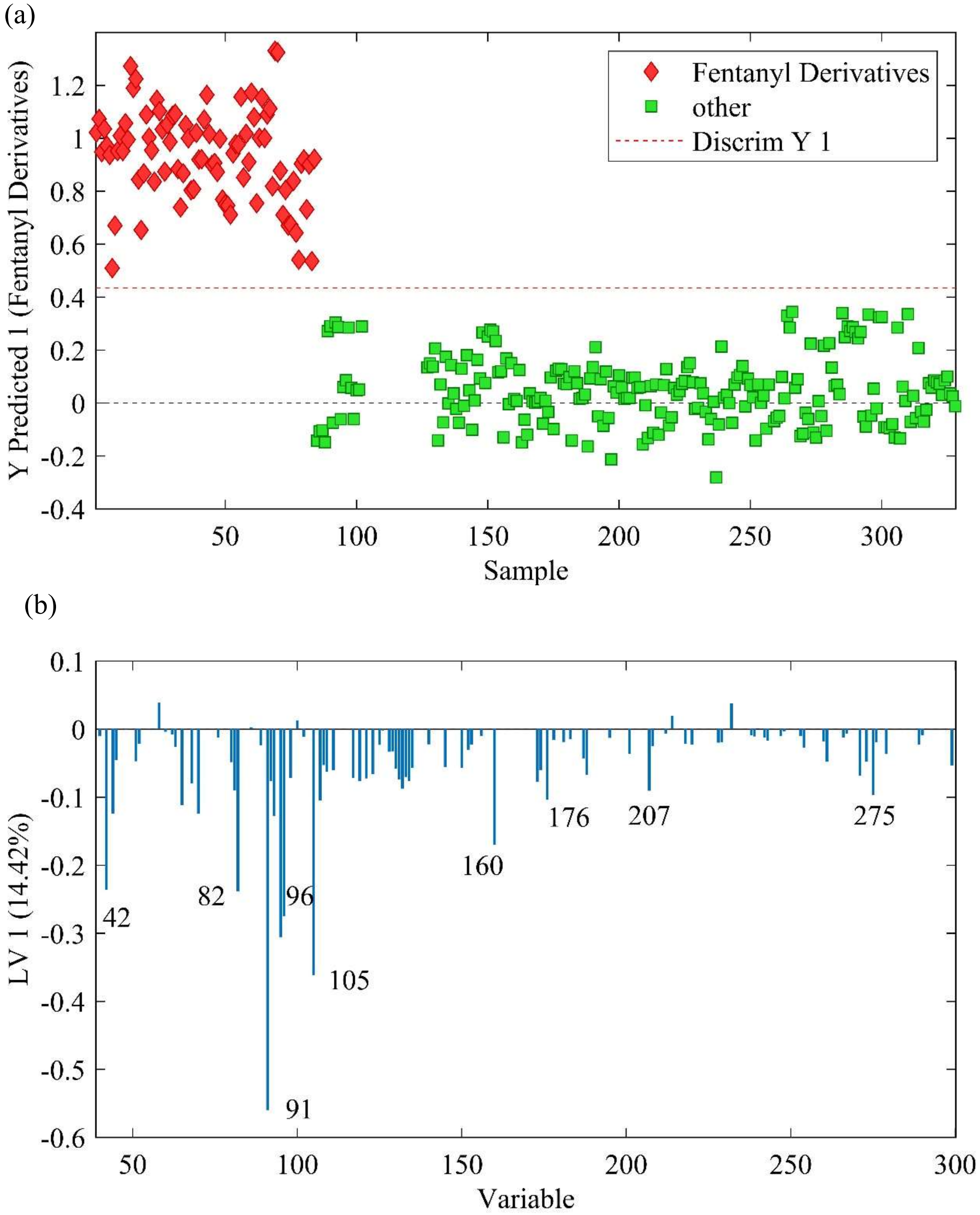

Figure 12. (a) DA plot of training set and after GA for fentanyl derivatives. The first LV for the model was included showing negative peaks present in fentanyl derivative class.

Table 3: Model prediction and cross validation prediction results for PLS-DA of the binary problem

Model Results					
	Accuracy	TPR	TNR	FPR	FNR
Fentanyl	1.000	1.000	1.000	0	0
Other	1.000	1.000	1.000	0	0
	Cross Validation Results				
	Accuracy	TPR	TNR	FPR	FNR
Fentanyl	0.997	1.000	0.995	0.005	0
Other	0.997	0.995	1.000	0	0.005

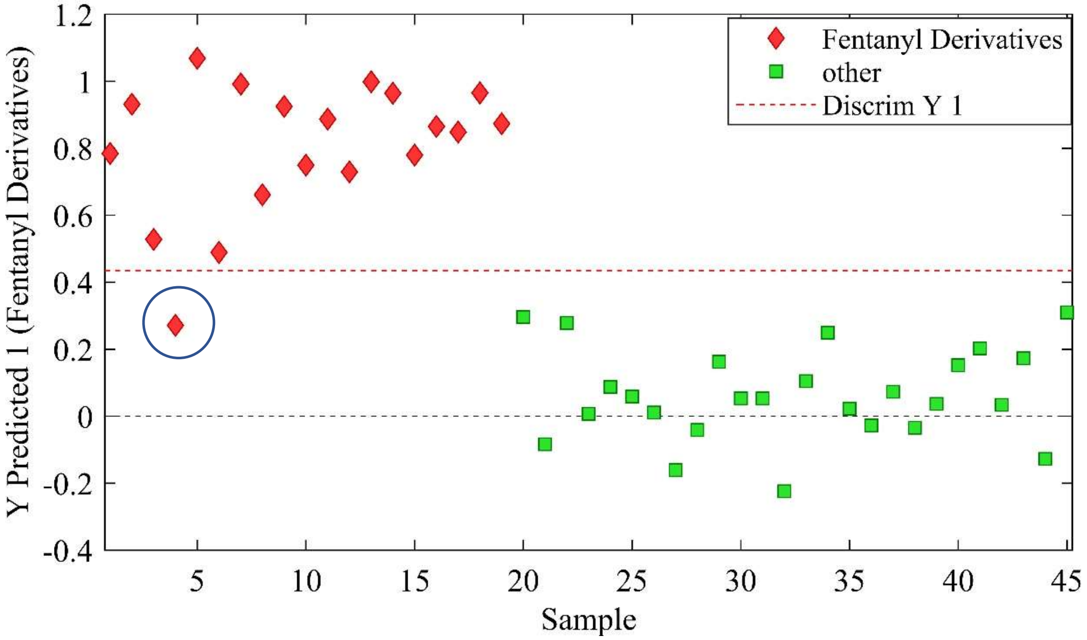

Figure 13. Prediction results from separate validation. Correct identification of all compounds except one.

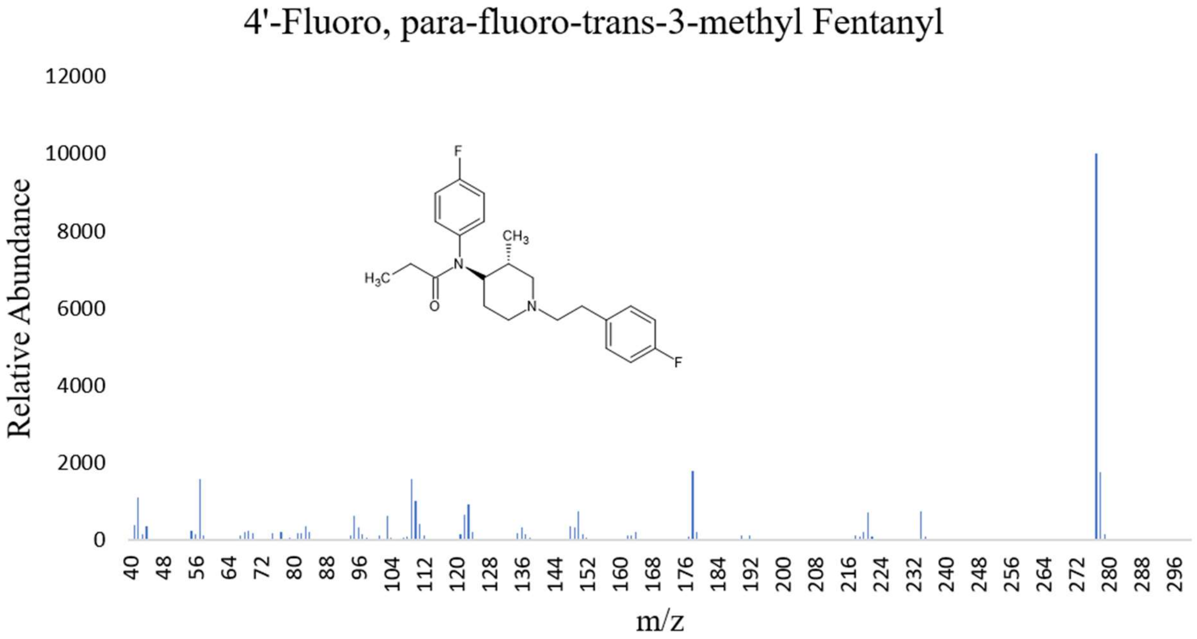

Figure 14. Mass spectrum for the misclassified compound from the test set.

Fentanyl Derivative Sub-Model

Once an acceptable model to discriminated fentanyl derivatives from other drugs, a sub-model was developed to separate a number of substituted fentanyl derivative compounds into classes based on their structural modifications. In the preliminary studies, fentanyl derivatives were grouped based on a method described in a previous study,[6] and a corresponding model was developed using PLS-DA. This model did not provide satisfactory results as the classes used did not seem to correlate with the structural modifications, and not all types of fentanyl derivative were represented in this study. A new way to classify these compounds was needed and to accomplish this, hierarchal clustering Ward's method was performed for 173 fentanyl derivatives. This produced three subgroups based on the PCA data (see Figure 15) (see Appendix C). These groups were analyzed to determine the structural modifications on the fentanyl core structure.

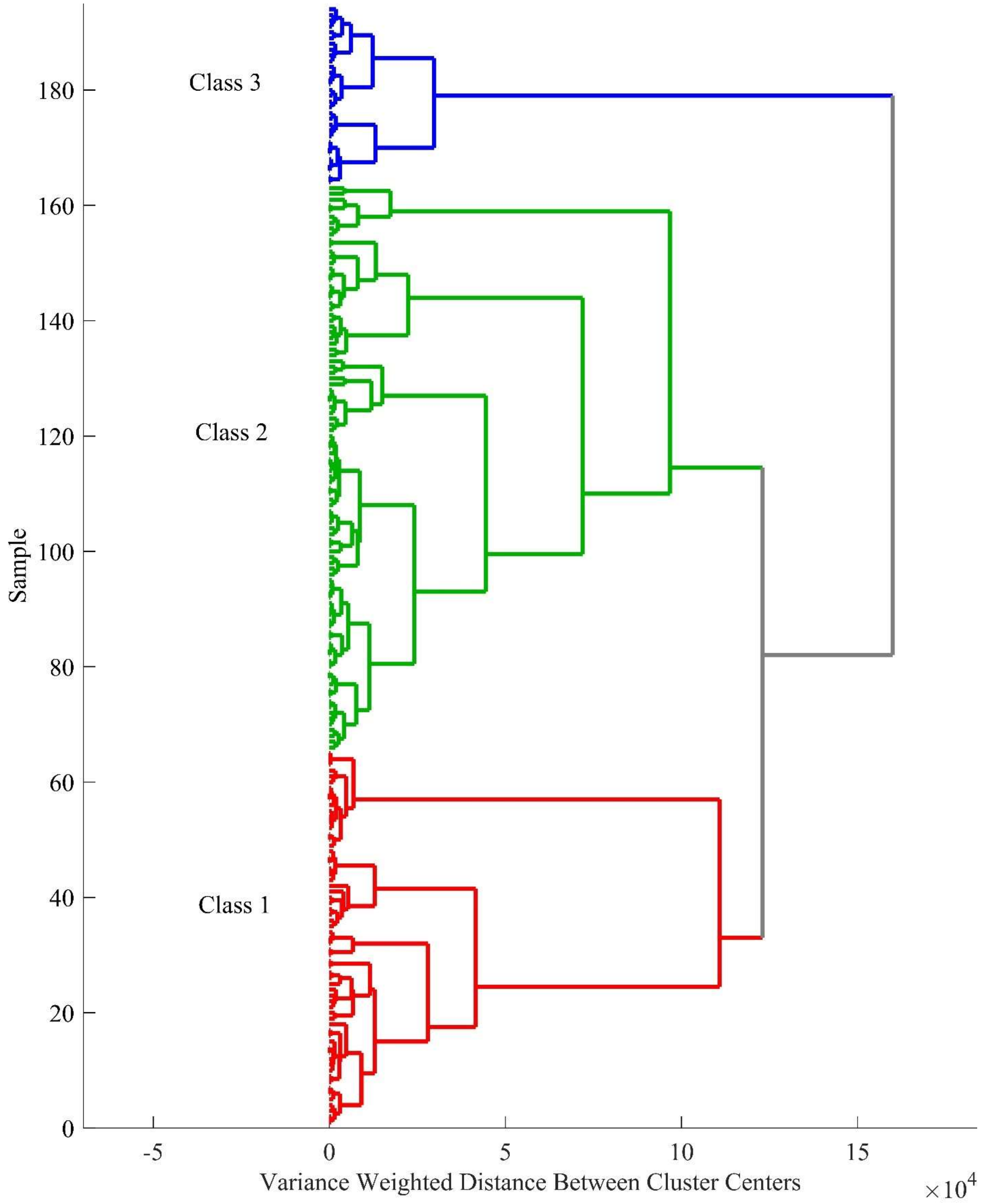

Figure 15. Dendrogram of the three derived classes resulting from Ward's method on the sub-model data.

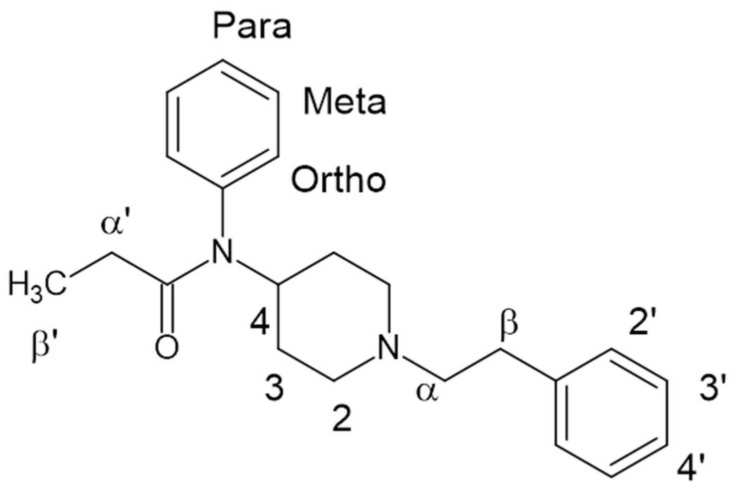

Figure 16: Labeled structure of fentanyl.

Class three compounds appeared to be mainly halogen containing compounds. The compounds in this class contained mostly a fluorine atom, or in some compounds a chlorine atom, on the alpha side of the molecule (see Figure 16). Class two compounds primarily contained modifications on the alpha prime and beta prime carbons. This would cause class two compounds to have a similar fragmentation pattern to fentanyl (see Figure 2). Class one was comprised of compounds that do not belong to Class 2 or Class 3. Common structural modifications in this group were changes to the ortho, meta, and para carbons, or groups at alpha prime that were able to form a fragment stable enough to change the mass spectral fragmentation pattern of the entire molecule. PLS-DA analysis was performed on the fentanyl derivatives using these three classes. The model performance shows discrimination between classes, but several compounds were predicted outside of their class that is predicted by the hierarchical classification (see Figure 17). Once the structural characteristics of each class were identified, the important fragments for each class were determined.

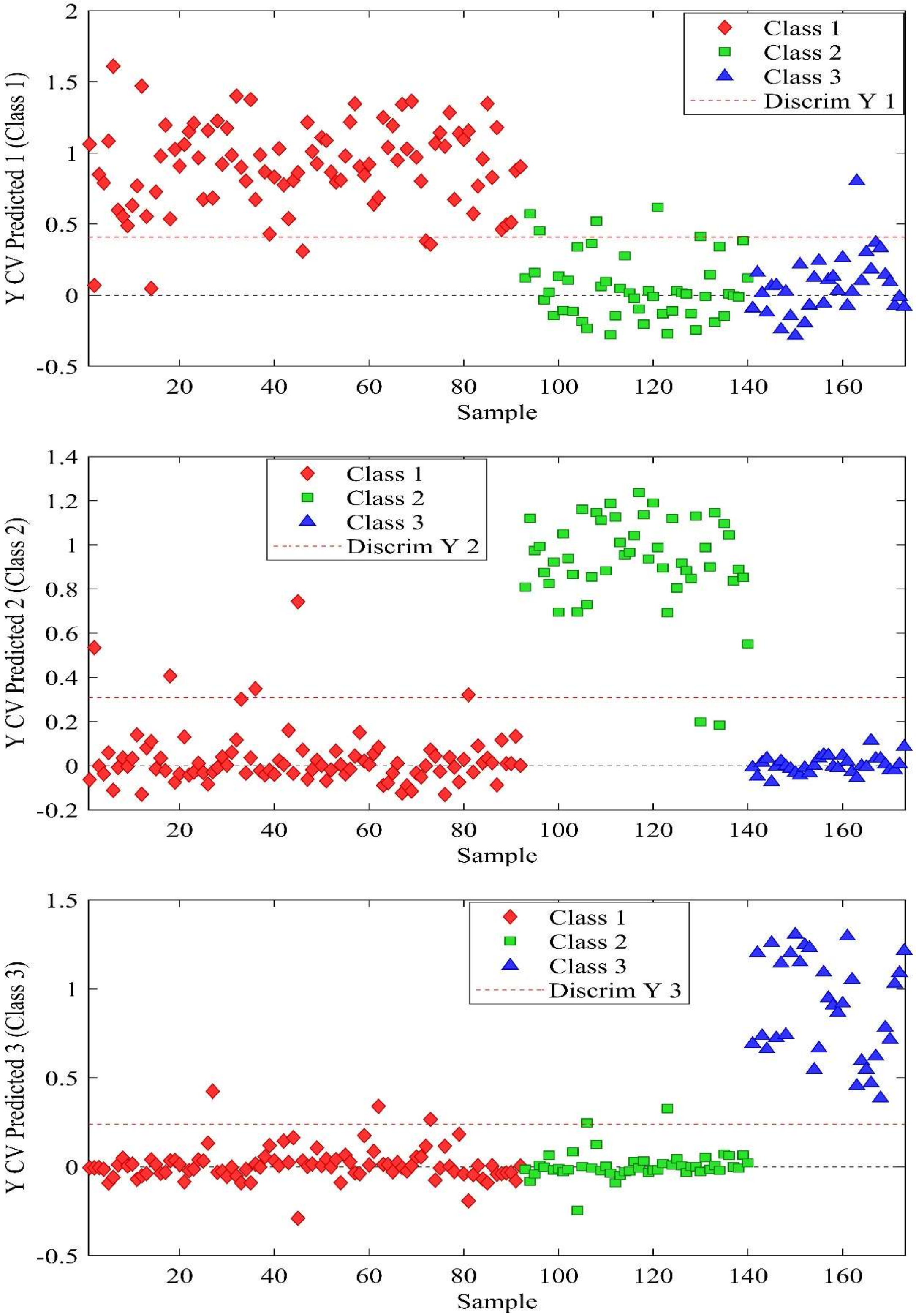

Figure 17. Initial fentanyl sub-model. significant differentiation, but with several misclassified compounds

Then, chemical structures of all fentanyl derivatives used in this study were analyzed to determine if hierarchical classification is accurate in predicting the class. During this process, it was discovered that several compounds in this class data were deuterium labelled compounds, which are identical to their non-deuterated counterparts but have a slightly higher mass. Deuterated compounds are commonly used as internal standards in forensic drug analysis. These samples were removed as the fragments created from these molecules will have heavier masses compared to non-deuterated counterparts. After the data was "cleaned" in this manner, the model performance improved. However, the cross-validation results show that there were ten misclassified samples. This required us to use GA to select the variables that are important in the discrimination. PLS-DA plots generated using GA is shown in Figure 18, and the TPR and TNR of this method can be found in Table 4.

The resulting sub-model accurately predicted the three classes of fentanyl derivatives. Despite some incorrect predictions, this study demonstrates that hierarchal clustering can effectively be used in the initial determination of the sub-classes when chemical structural similarities are not fully understood.

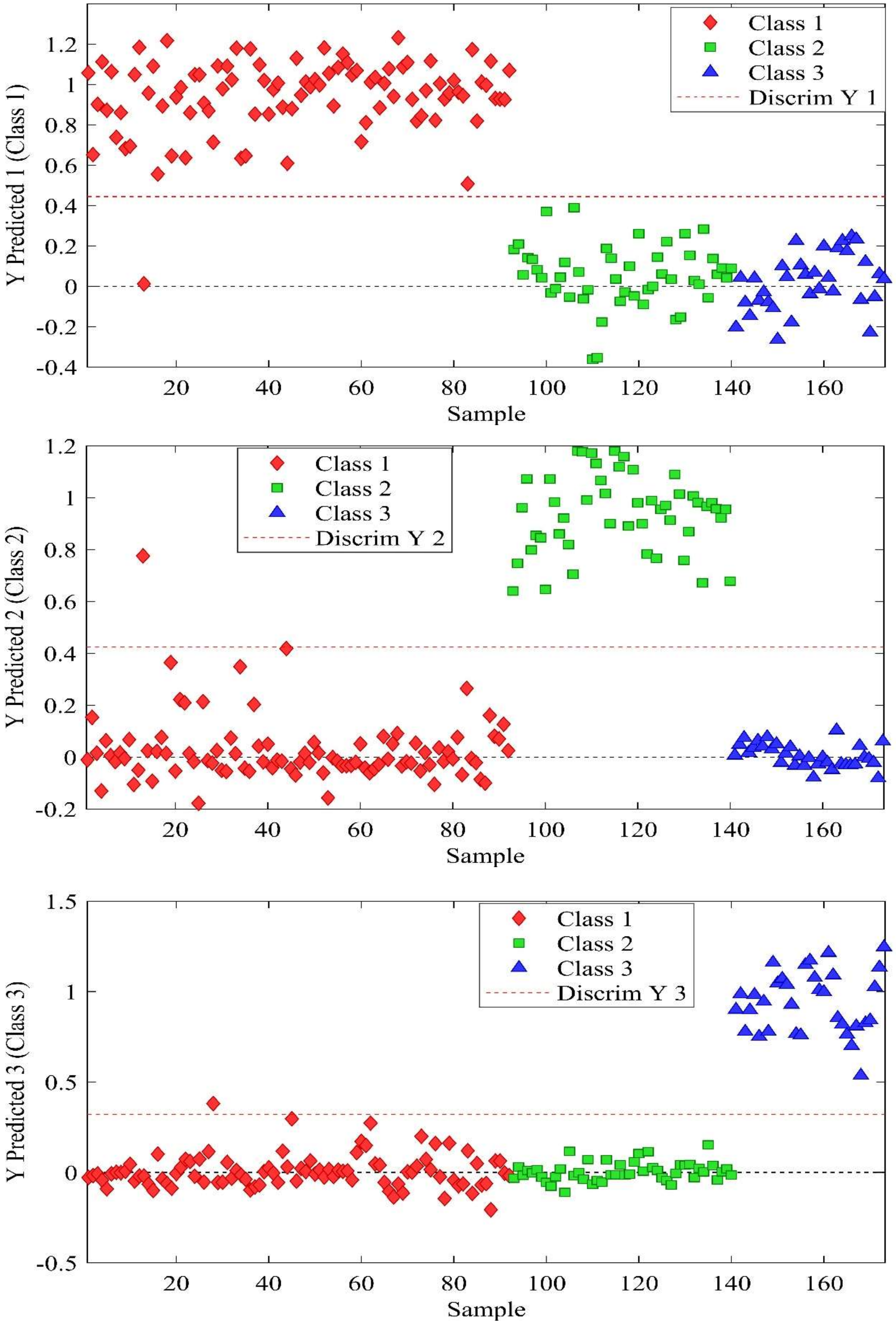

Figure 18: Fentanyl sub-model using cleaned data and GA. Reduced number of misclassified compounds

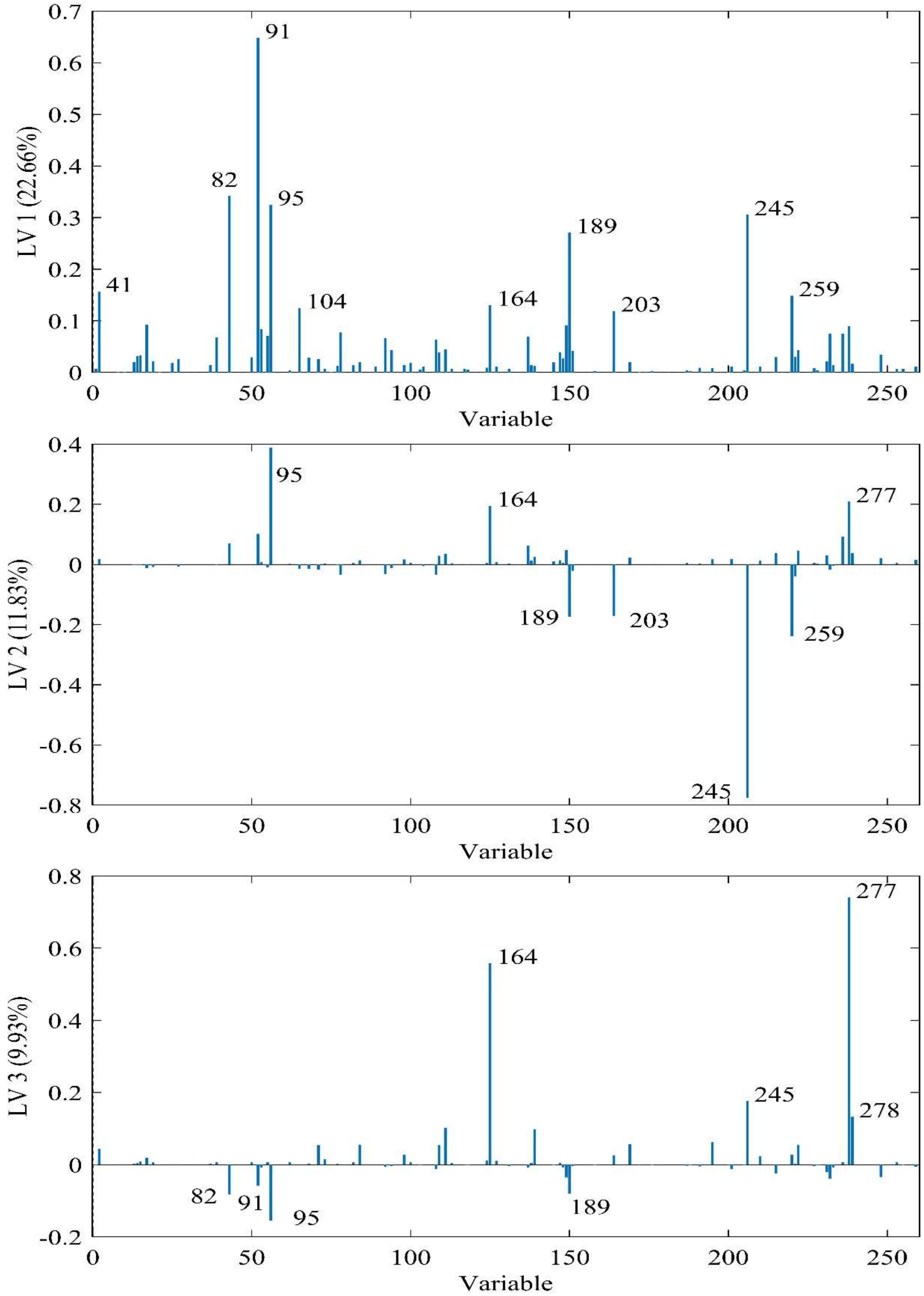

Figure 19: First three loading vectors for fentanyl sub model. Positive values in LV 1 correspond to class one compounds. Negative values in LV 2 correspond to class 2 compounds while positive values correspond to the other 2 classes. Positive values in LV3 correspond to class 3 compounds.

Table 4: Model prediction results and cross validation prediction results before and after variable selection for fentanyl sub-model.

Model Results					
	Accuracy	TPR	TNR	FPR	FNR
Class 1	0.983	0.989	0.975	0.025	0.011
Class 2	0.983	0.958	0.992	0.080	0.042
Class 3	1.000	1.000	1.000	0	0
Cross Validation Results					
	Accuracy	TPR	TNR	FPR	FNR
Class 1	0.965	0.957	0.975	0.025	0.043
Class 2	0.971	0.958	0.976	0.024	0.042
Class 3	0.994	1.000	0.993	0.007	0

Loading vectors calculated for this model reinforce (See Figure 19.) the structural identities of each class. LV1 has higher Emphasis on m/z of peaks related to the fentanyl core structure (m/z = 91, 189, and 245) and positive values of LV1 corresponds to class one. This explains the presence of have many important peaks from both other classes with less weight. The largest peak in LV 1, m/z = 91, is present in most compounds in the fentanyl derivative class. Class two has many of the same qualities of unmodified fentanyl, and this can be seen by the presence of 245 m/z, the base peak for fentanyl and 189 m/z, another common fragment of fentanyl. Class three generally contains heavier atoms in its modifications like fluorine and chlorine. This can be seen in the peaks in LV 3 which correspond to peaks in fentanyl but have higher m/z ratios.

CHAPTER FOUR: CONCLUSION

A new method was developed for the presumptive identification of novel fentanyl derivatives using mass spectral data of known compounds. Further, a sub-model was developed to discriminated fentanyl derivatives based on substitution patterns. Most importantly, we demonstrated that hierarchical clustering combined with classification methods such as PLS-DA can be used not only to develop a model to classify mass spectral data of compounds without knowing the structural modifications on them but also to find the important *m/z* ratios that can be used to discriminate the classes. Validation results show that these methods are robust with high accuracy, TPR and TNR. This method was simple and easy to develop and use in forensic labs without generating additional data. Routine drug analysis generates mass spectra of all the compounds present in a mixture injected into GCMS and the mass spectra of unknown compounds from the same analysis can be projected into these models and determine if fentanyl derivatives are present. Structural determination of these compounds can then be easily performed as the subclass of the drug can be predicted using sub-models. This method can also be useful when the amount of drug recovered is not adequate to be used in structural elucidation using HRMS, FTIR, or NMR.

Future experimentation will be focused on (I) validate the performance of the models by using newly reported NPSs. Mass spectra of these drugs will be obtained from SWGDrug or will be generated in-house using GCMS. (II) Develop sub-models for cannabinoids drug class using the same methodology.

REFERENCES

(1) Feeney, W.; Moorthy, A.; Sisco, E. Spectral Trends in GC-EI-MS Data Obtained from the SWGDRUG Library and Literature: A Resource for the Identification of Unknown Compounds. Forensic Chemistry **2022**, 31.

(2) Strano Rossi, S.; Odoardi, S.; Gregori, A.; Peluso, G.; Ripani, L.; Ortar, G.; Serpelloni, G.; Romolo, F. S. An Analytical Approach to the Forensic Identification of Different Classes of New Psychoactive Substances (Npss) in Seized Materials. *Rapid Communications in Mass Spectrometry* **2014**, *28* (17), 1904–1916.

(3) Winokur, A. D.; Kaufman, L. M.; Almirall, J. R. Differentiation and Identification of Fentanyl Analogues Using GC-IRD. *Forensic Chemistry* **2020**, *20*, 100255.

(4) Current NPS threats. https://www.unodc.org/unodc/en/scientists/current-nps-threats.html (accessed Nov 2, 2022).

(5) Mohr, A. L.; Logan, B. K.; Fogarty, M. F.; Krotulski, A. J.; Papsun, D. M.; Kacinko, S. L.; Huestis, M. A.; Ropero-Miller, J. D. Reports of Adverse Events Associated with Use of Novel Psychoactive Substances, 2017–2020: A Review. *Journal of Analytical Toxicology* **2022**, *46* (6).

(6) Gilbert, N.; Mewis, R. E.; Sutcliffe, O. B. Classification of Fentanyl Analogues through Principal Component Analysis (PCA) and Hierarchical Clustering of GC–MS Data. *Forensic Chemistry* **2020**, *21*, 100287.

(7) Levitas, M. P.; Andrews, E.; Lurie, I.; Marginean, I. Discrimination of Synthetic Cathinones by GC–MS and GC–MS/MS Using Cold Electron Ionization. *Forensic Science International* **2018**, *288*, 107–114.

(8) Koshute, P.; Hagan, N.; Jameson, N. J. Machine Learning Model for Detecting Fentanyl Analogs from Mass Spectra. *Forensic Chemistry* **2022**, *27*, 100379.

(9) Granger, R. M.; Yochum, H. M.; Granger, J. N.; Sienerth, K. D. *Instrumental analysis*, First ed.; Oxford University Press: New York, NY, **2017**.

(10) Silverstein, R. M.; Webster, F. X.; Kiemle, D. J. *Spectrometric identification of Organic Compounds*; John Wiley & Sons: Hoboken, NJ, NJ, **2005**.

(11) https://www.swgdrug.org/index.htm (accessed Nov 2, 2022).

(12) Huffman, S. Chem455 course. https://doi.org/10.15139/S3/6J9ZAU (accessed Nov 4, 2022).

(13) Gromski, P. S.; Xu, Y.; Correa, E.; Ellis, D. I.; Turner, M. L.; Goodacre, R. A Comparative Investigation of Modern Feature Selection and Classification Approaches for the Analysis of Mass Spectrometry Data. *Analytica Chimica Acta*. **2014**, 829. DOI: 10.1016/j.aca.2014.03.039

(14) Kranenburg, R. F.; Peroni, D.; Affourtit, S.; Westerhuis, J. A.; Smilde, A. K.; van Asten, A. C. Revealing Hidden Information in GC-MS Spectra from Isomeric Drugs: Chemometrics Based Identification from 15 eV and 70 eV EI Mass Spectra. *Forensic Chemistry.* **2020**, 18. DOI: 10.1016/j.forc.2020.100225

(15) Cheek M. E. An exploration of chemometric regression techniques to analyze infrared spectra of aqueous sugar mixtures (dissertation). ProQuest LLC. **2019**

(16) Pereira, L. S.A.; Lisboa, F. L.C.; Neto, J. C.; Valladao, F. N.; Sena, M. M. Screening Method for Rapid Classification of Psychoactive Substances in Illicit Tablets Using Mid Infrared Spectroscopy and PLS-DA. *Forensic Science*

(17) Valdez, C. A. Gas Chromatography-Mass Spectrometry Analysis of Synthetic Opioids Belonging to the Fentanyl Class: A Review. *Critical Reviews in Analytical Chemistry* **2021**, 1–31.

(18) Werther, W.; Lohninger, H.; Stancl F.; Varmuza K. Classification of Mass Spectra A Comparison of Yes/No Classification Methods for the Recognition of Simple Structural Properties. *Chemometrics and Intelligent Laboratory Systems*. **1994**, 22 63-76.

(19) Bell, S. *Forensic chemistry*, Third ed.; CRC Press: Boca Raton, FL, **2022**.

(20) Lappas, N. T.; Lappas, C. M. *Forensic toxicology: Principles and concepts*, Second ed.; Academic Press, an imprint of Elsevier: San Diego, CA, **2022**.

(21) Hassanien, S. H.; Bassman, J. R.; Perrien Naccarato, C. M.; Twarozynski, J. J.; Traynor, J. R.; Iula, D. M.; Anand, J. P. In Vitro Pharmacology of Fentanyl Analogs at the Human Mu Opioid Receptor and Their Spectroscopic Analysis. *Drug Testing and Analysis* **2020**, *12* (8), 1212–1221.

(22) United States Drug Enforcement Administration. https://www.dea.gov/factsheets?keywords=&page=0 (accessed Nov 2, 2022).

(23) Valdez, C. A. Gas Chromatography-Mass Spectrometry Analysis of Synthetic Opioids Belonging to the Fentanyl Class: A Review. *Critical Reviews in Analytical Chemistry* **2021**, 1–31.

APPENDIX A: LOADING VECTORS

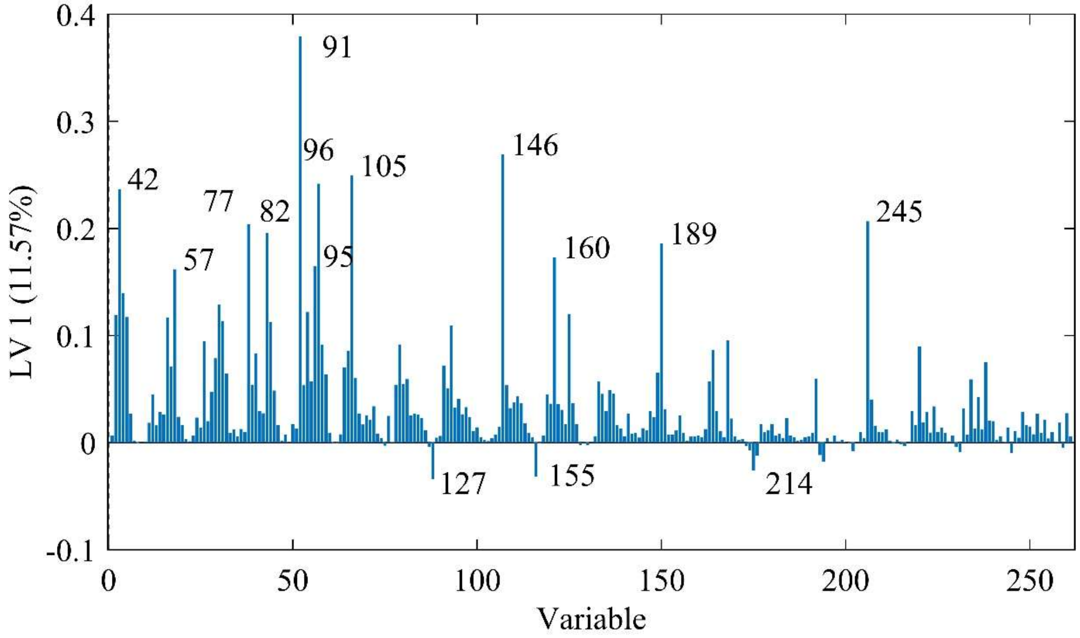

Figure A1: Loading vector 1 of Figure 6 – also shown in the text

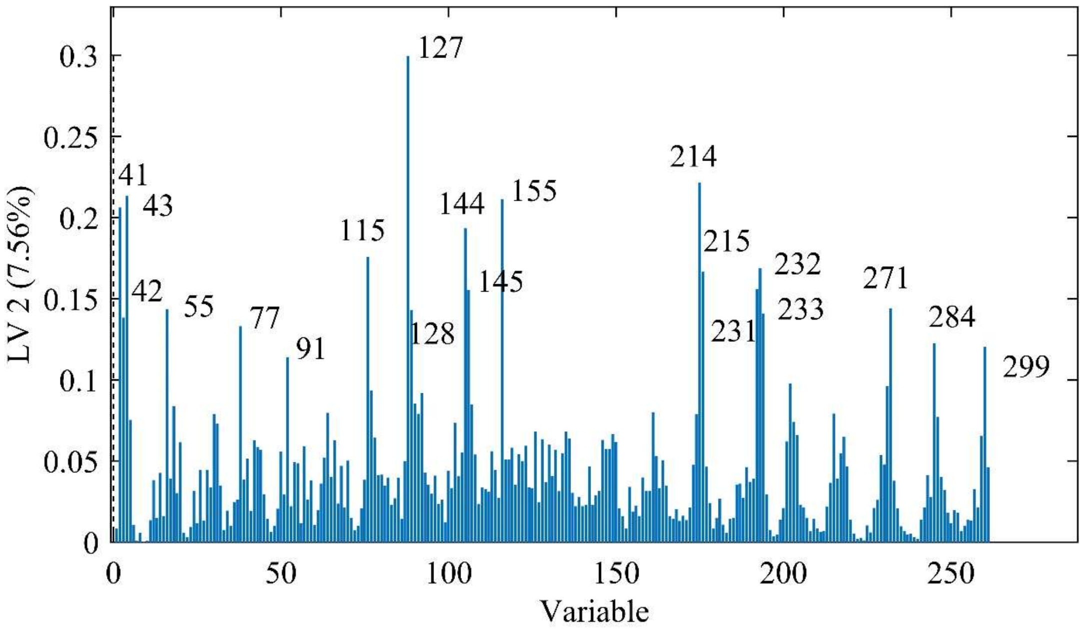

Figure A2: Loading vector 2 of Figure 6 – also shown in the text

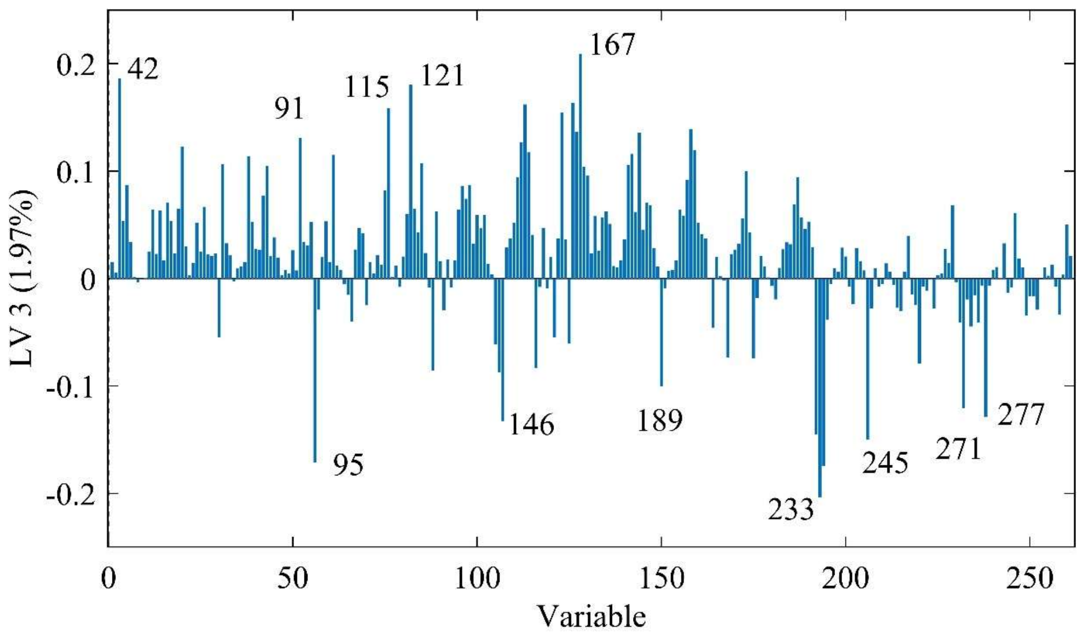

Figure A3: Loading vector 3 of Figure 6

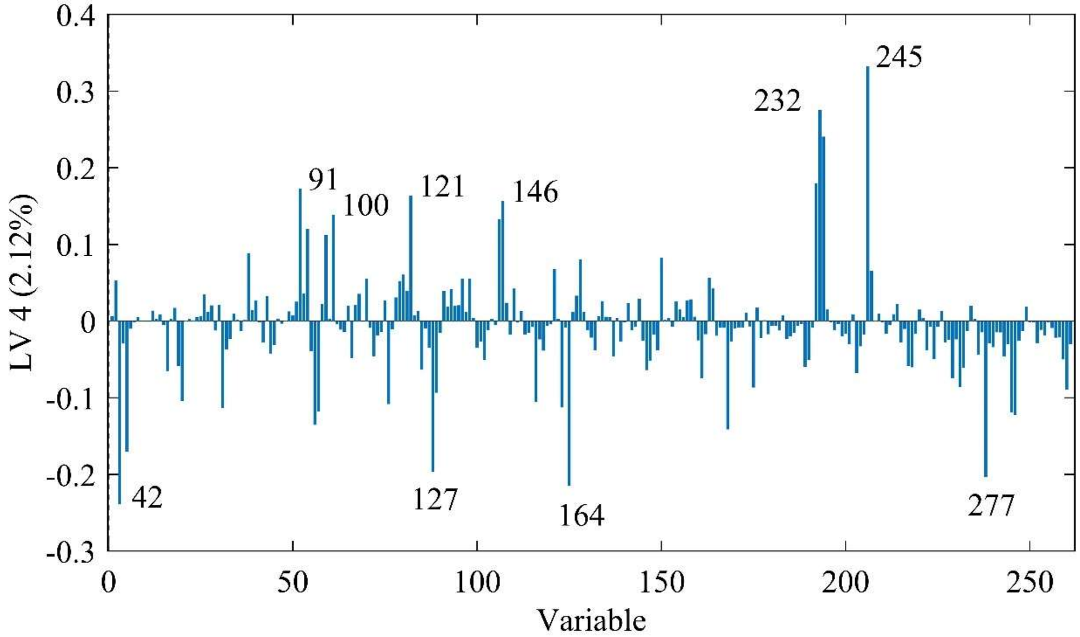

Figure A4: Loading vector 4 of Figure 6

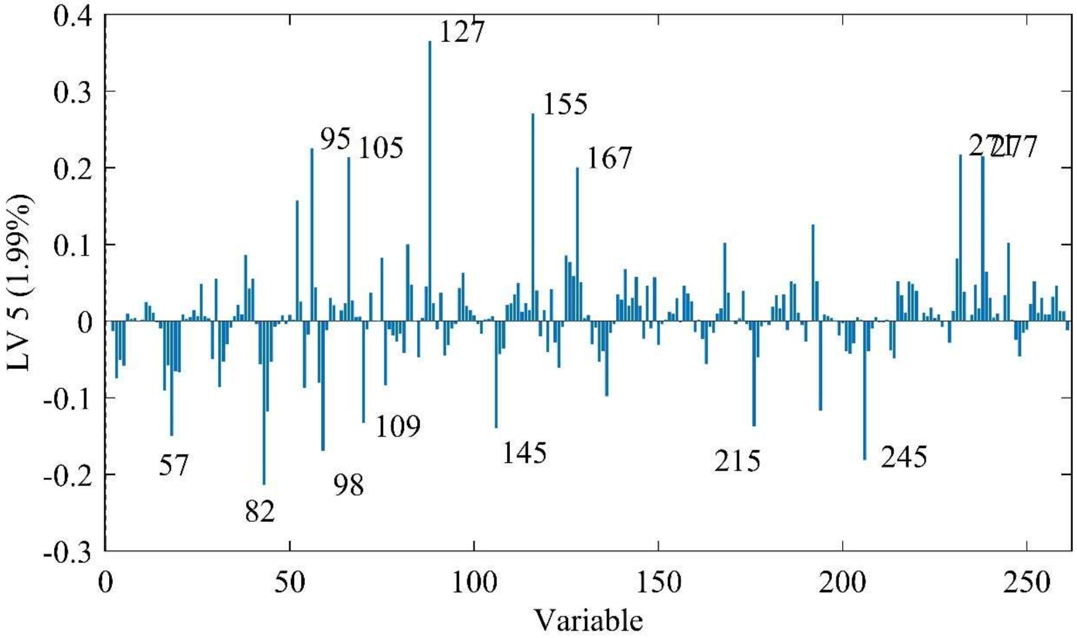

Figure A5: Loading vector 5 of Figure 6

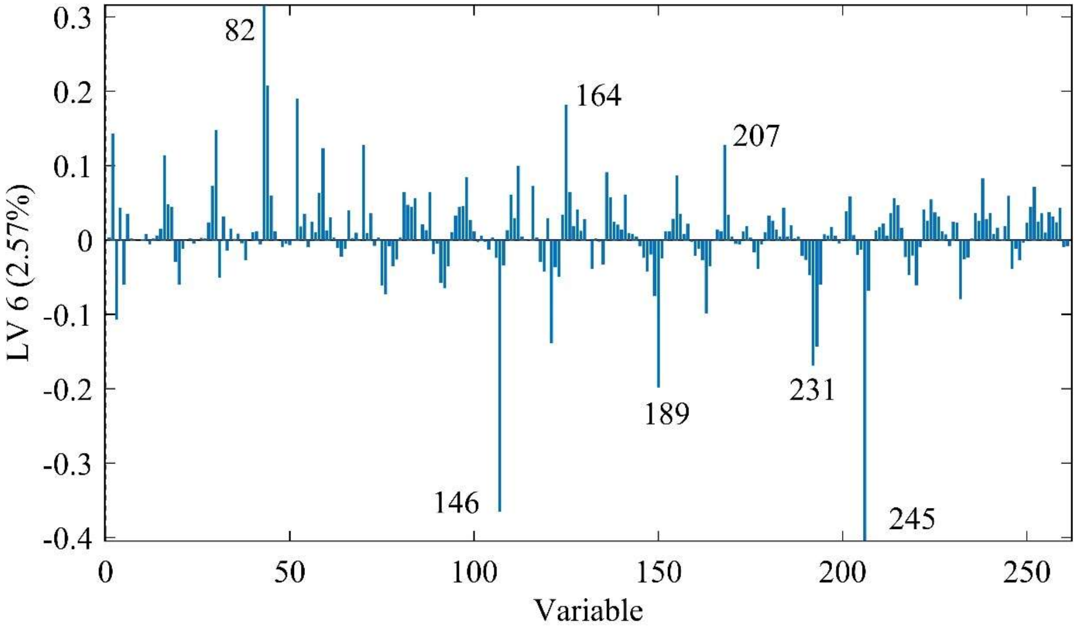

Figure A6: Loading vector 6 of Figure 6

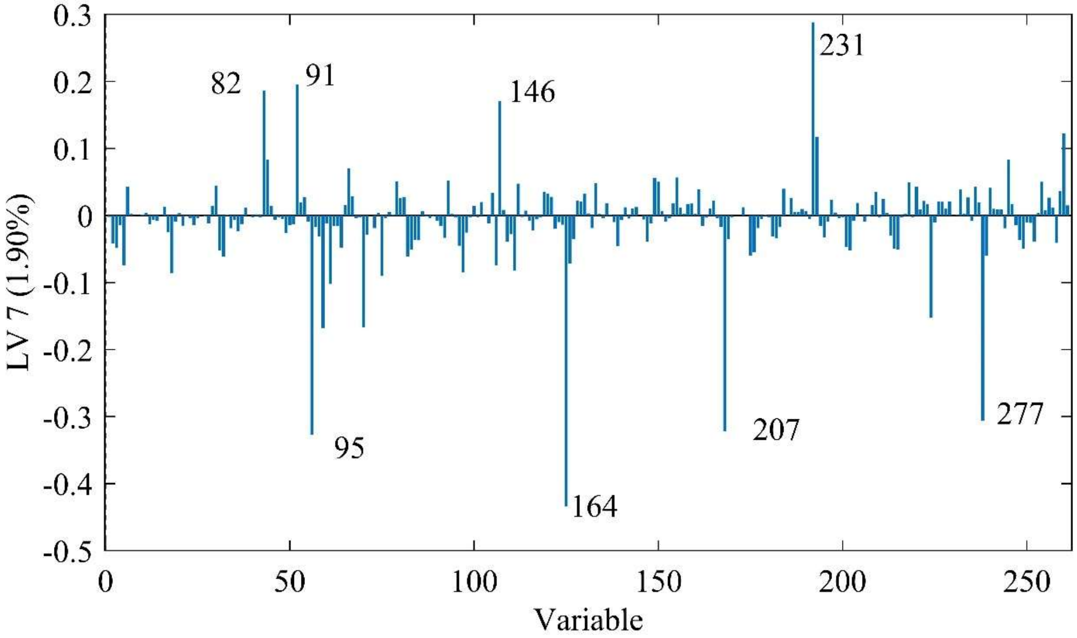

Figure A7: Loading vector 7 of Figure 6

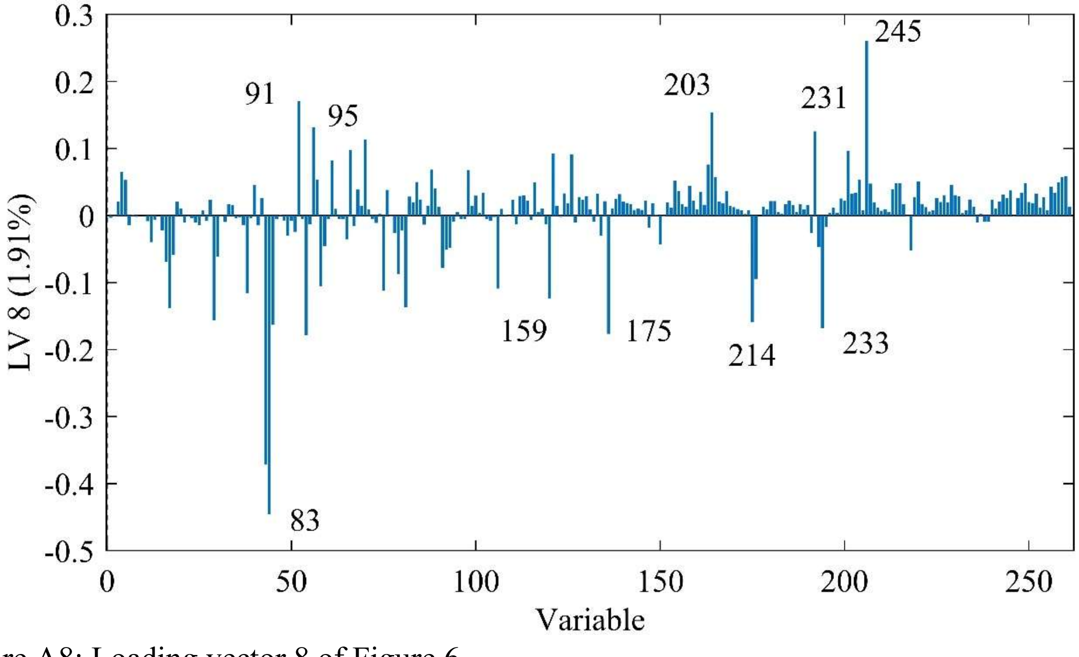

Figure A8: Loading vector 8 of Figure 6

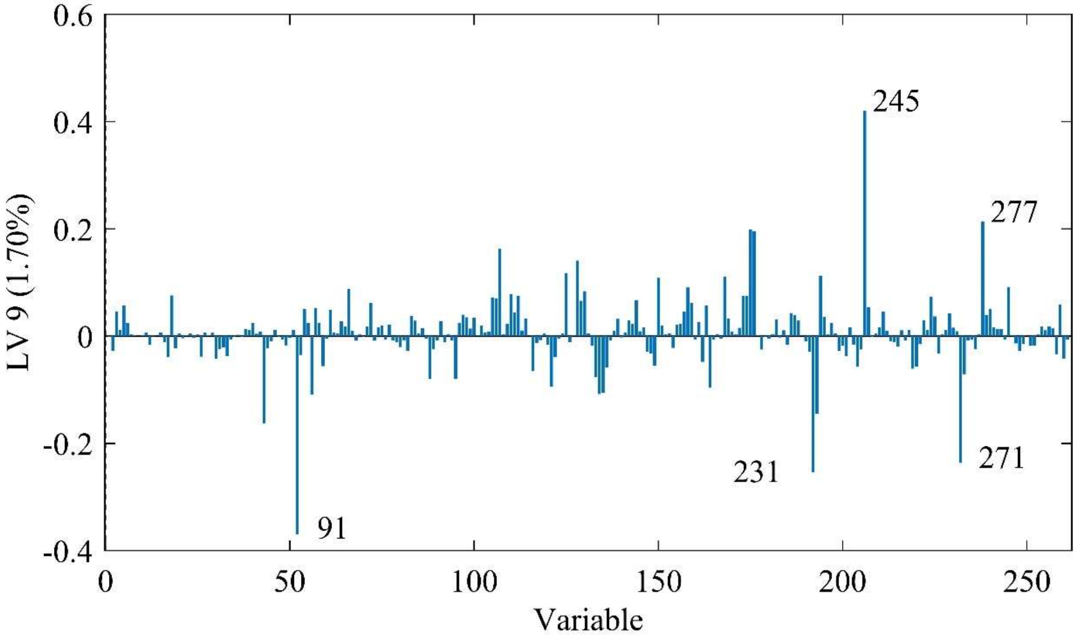

Figure A9: Loading vector 9 of Figure 6

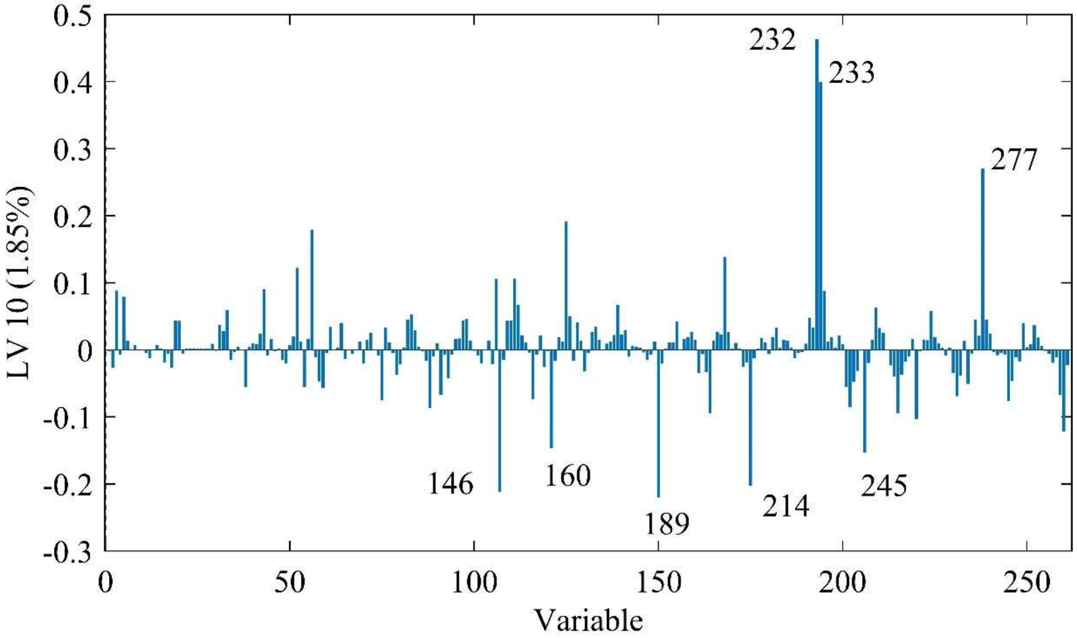

Figure A10: Loading vector 10 of Figure 6

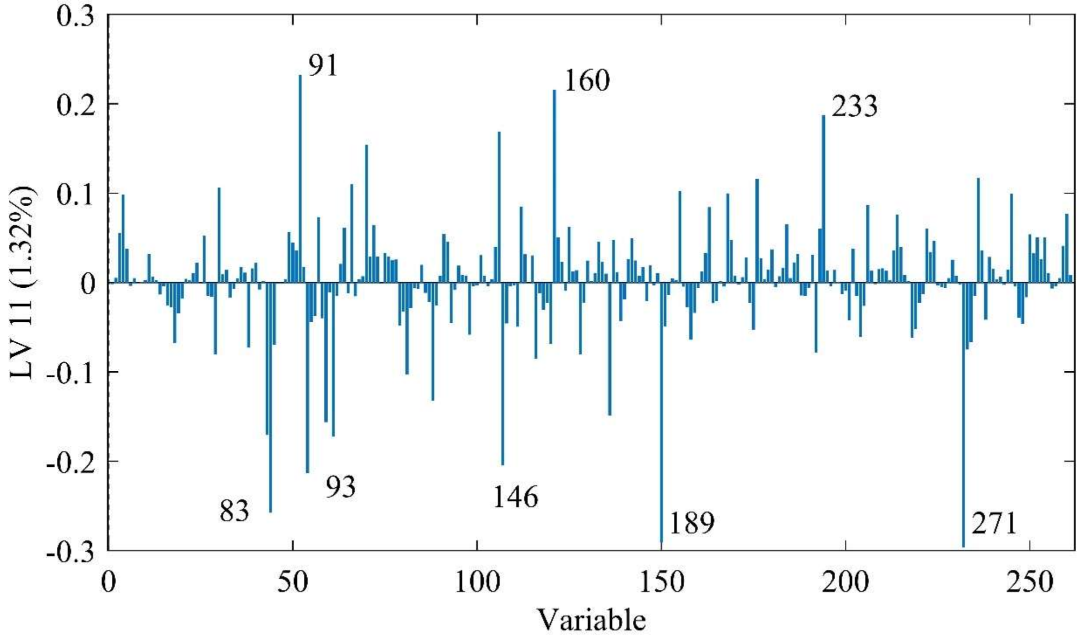

Figure A11: Loading vector 11 of Figure 6

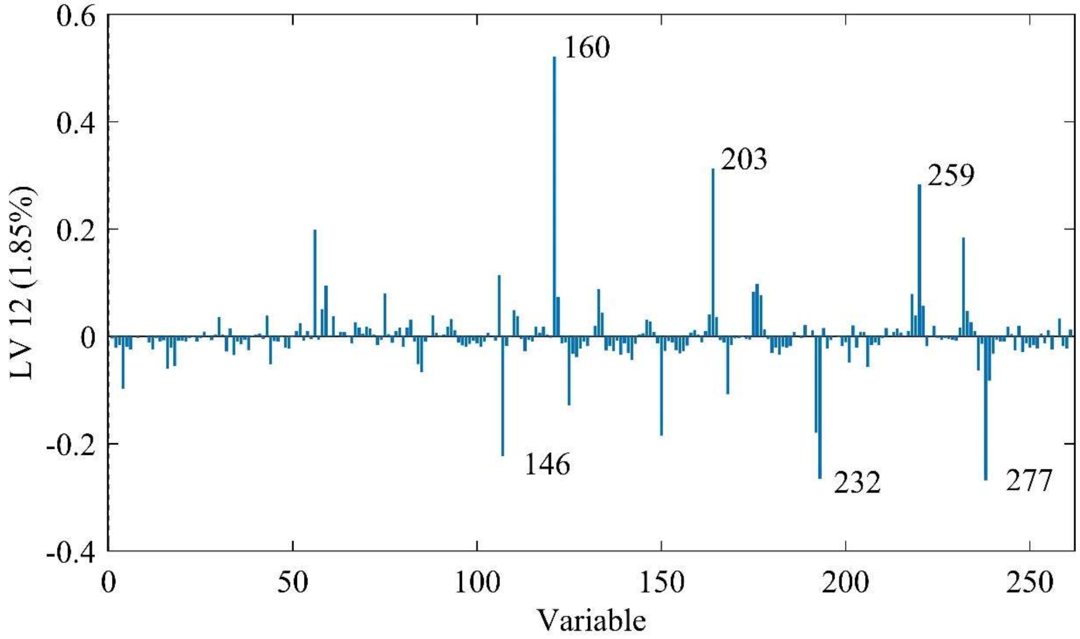

Figure A12: Loading vector 12 of Figure 6

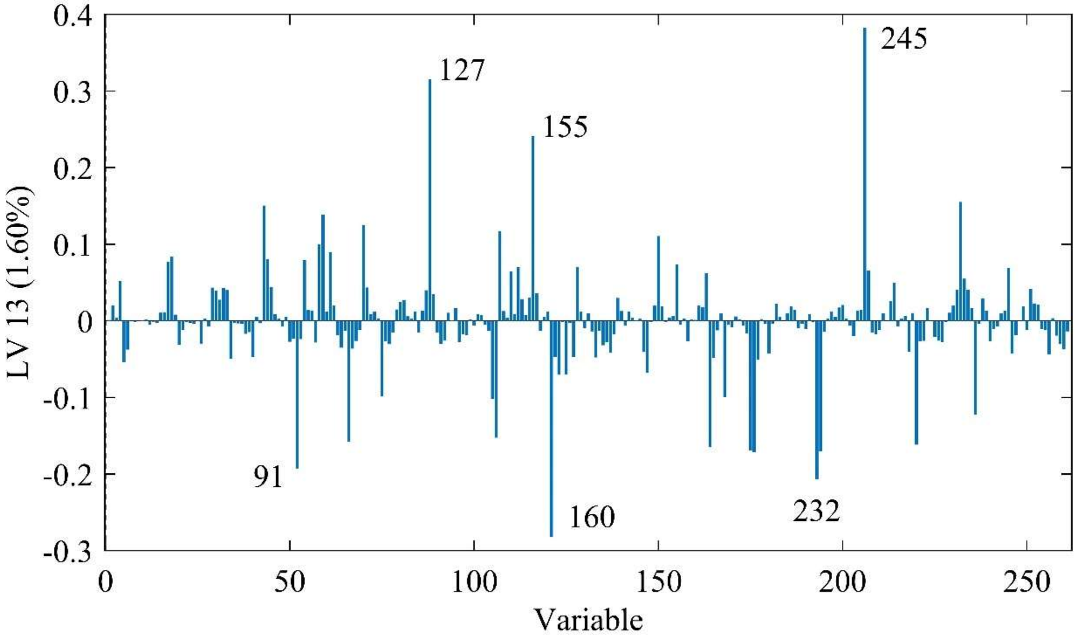

Figure A13: Loading vector 13 of Figure 6

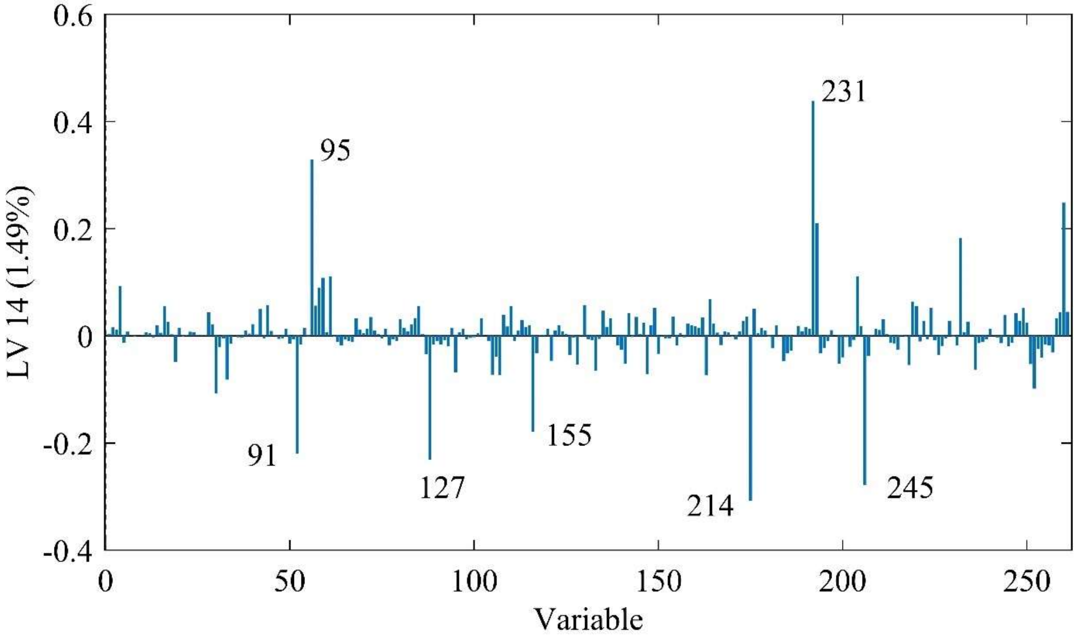

Figure A14: Loading vector 14 of Figure 6

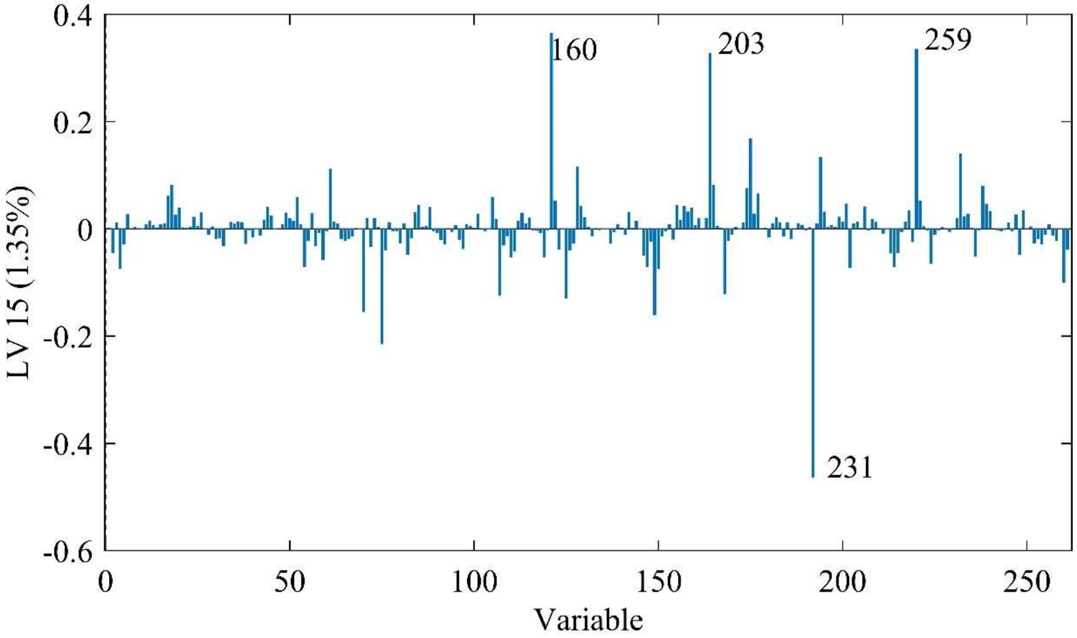

Figure A15: Loading vector 15 of Figure 6

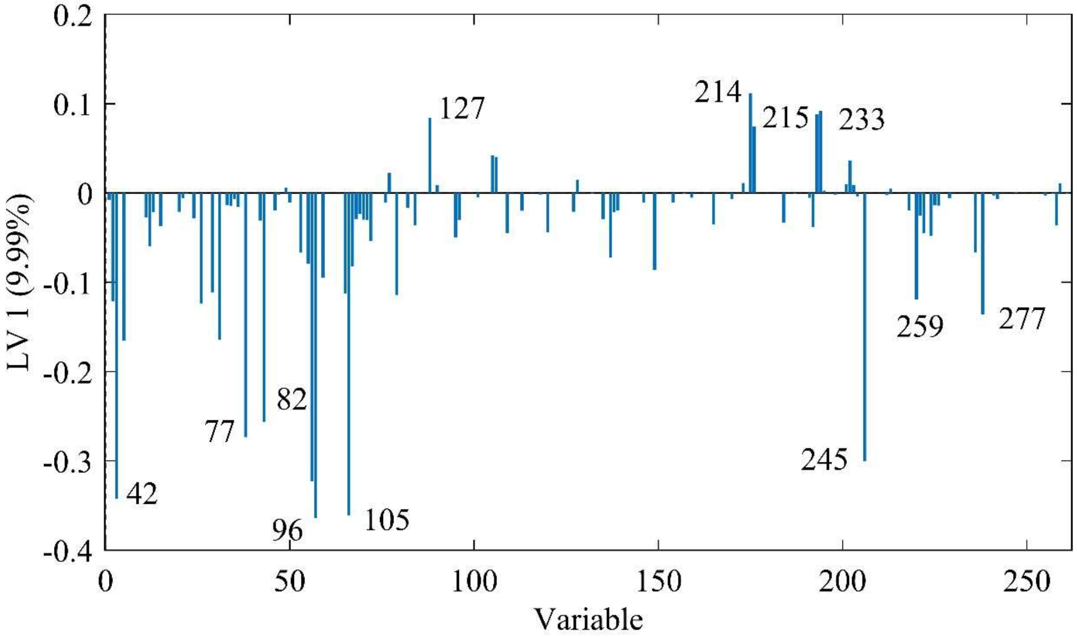

Figure A16: Loading vector 1 of Figure 8 – also shown in the text

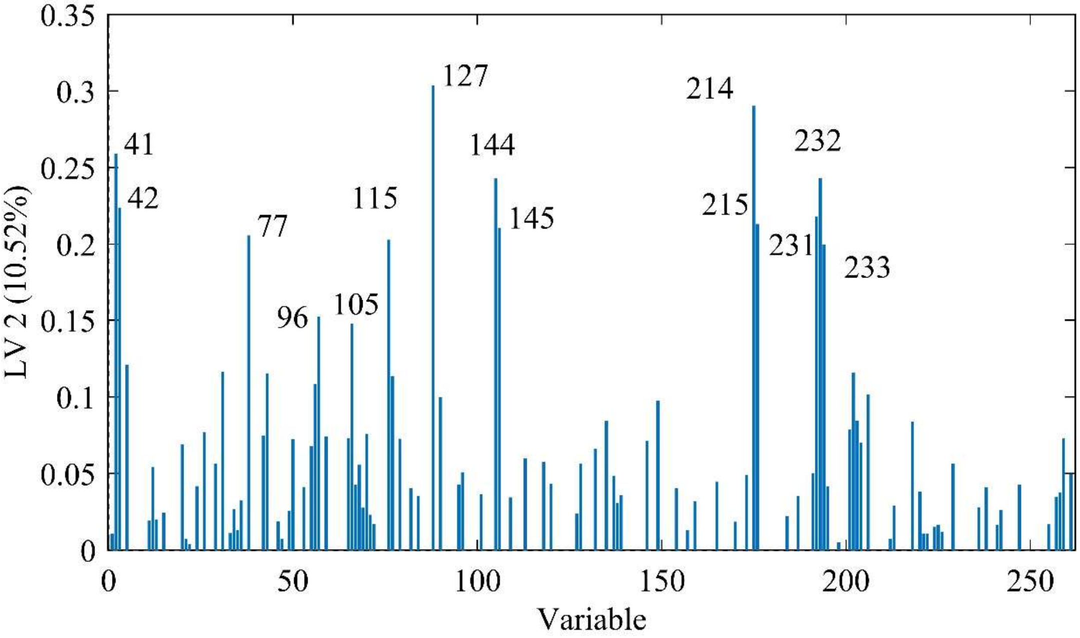

Figure A17: Loading vector 2 of Figure 8 – also shown in the text

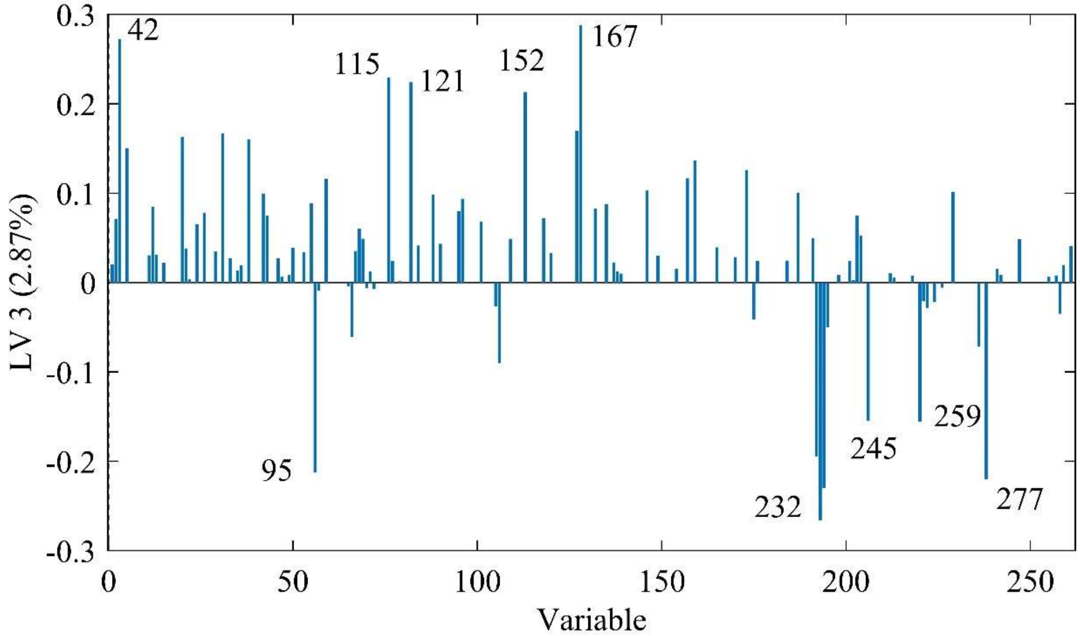

Figure A18: Loading vector 3 of Figure 8

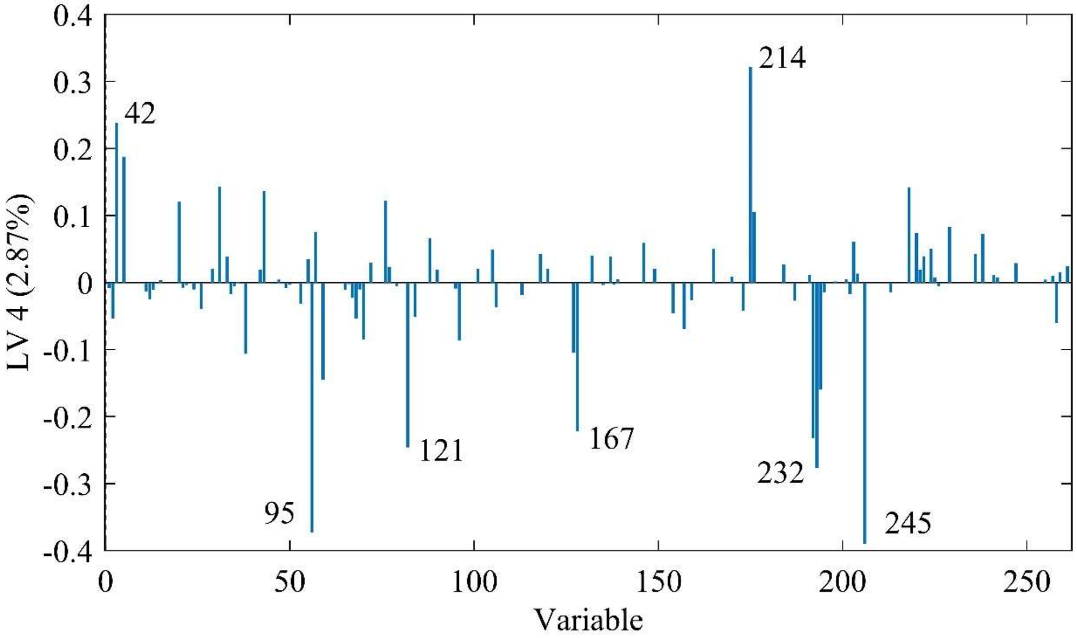

Figure A19: Loading vector 4 of Figure 8

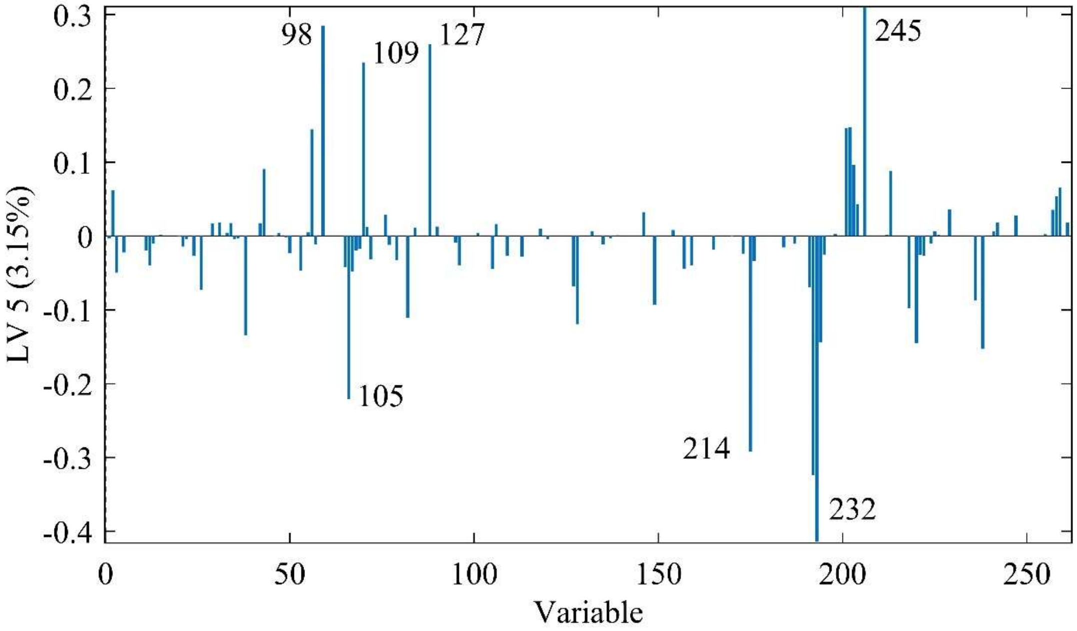

Figure A20: Loading vector 5 of Figure 8

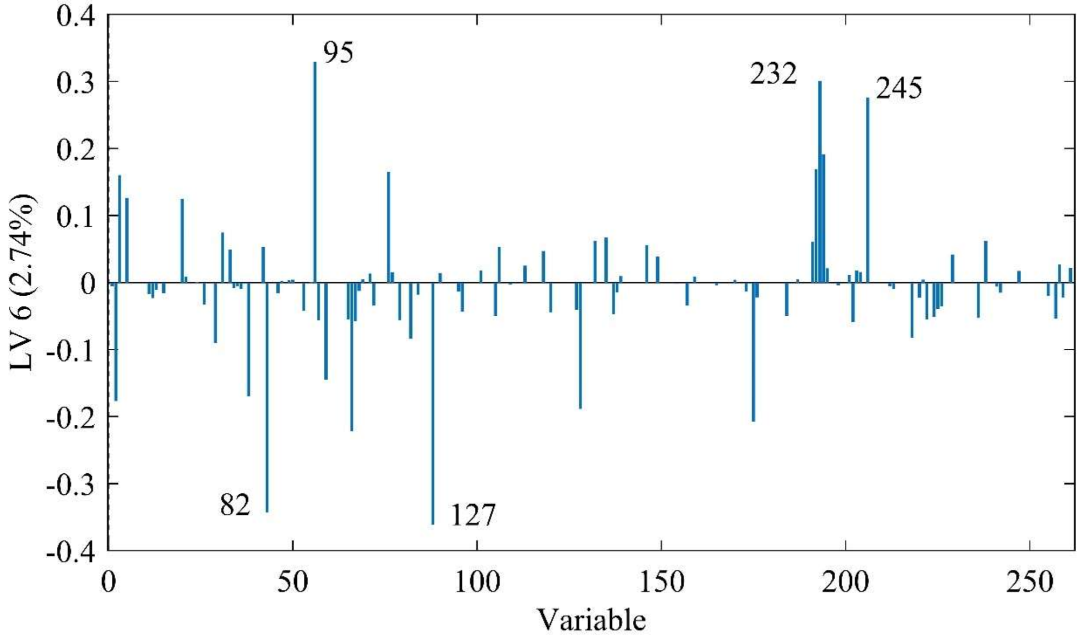

Figure A21: Loading vector 6 of Figure 8

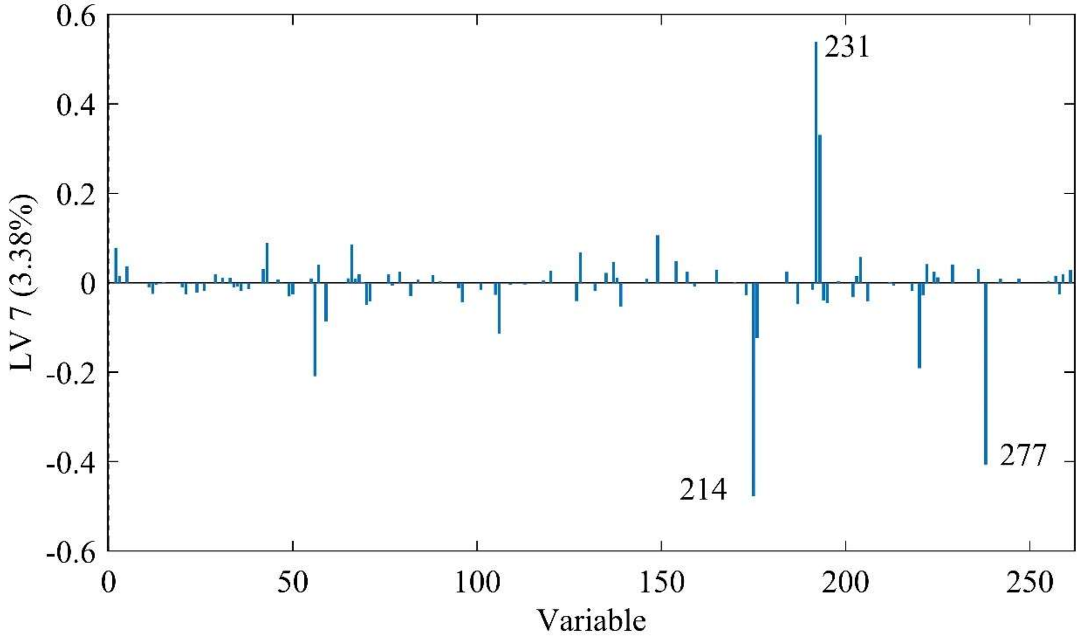

Figure A22: Loading vector 7 of Figure 8

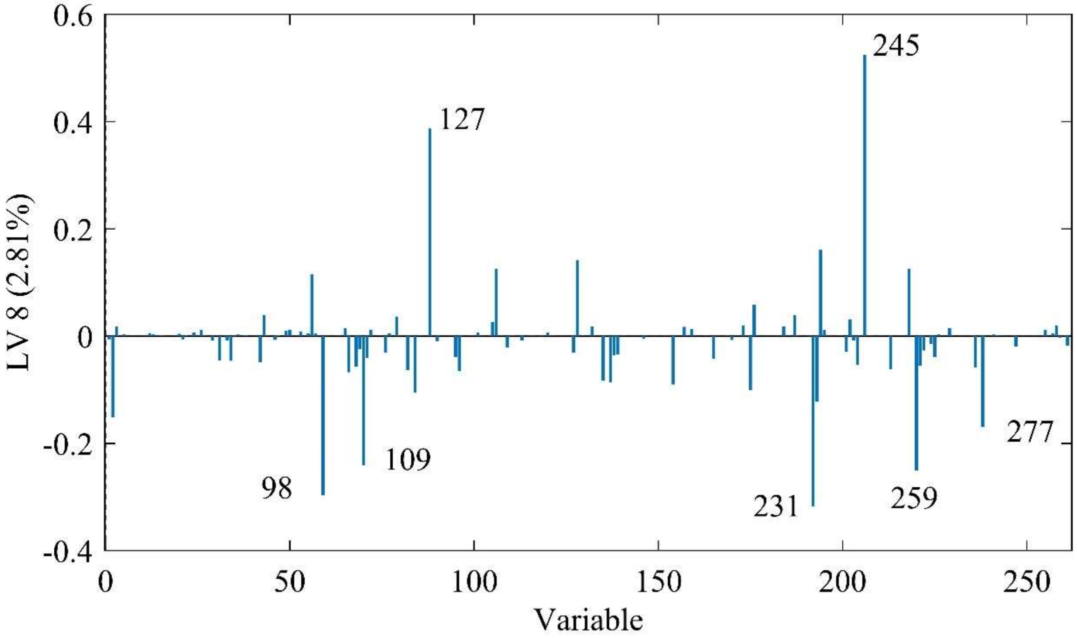

Figure A23: Loading vector 8 of Figure 8

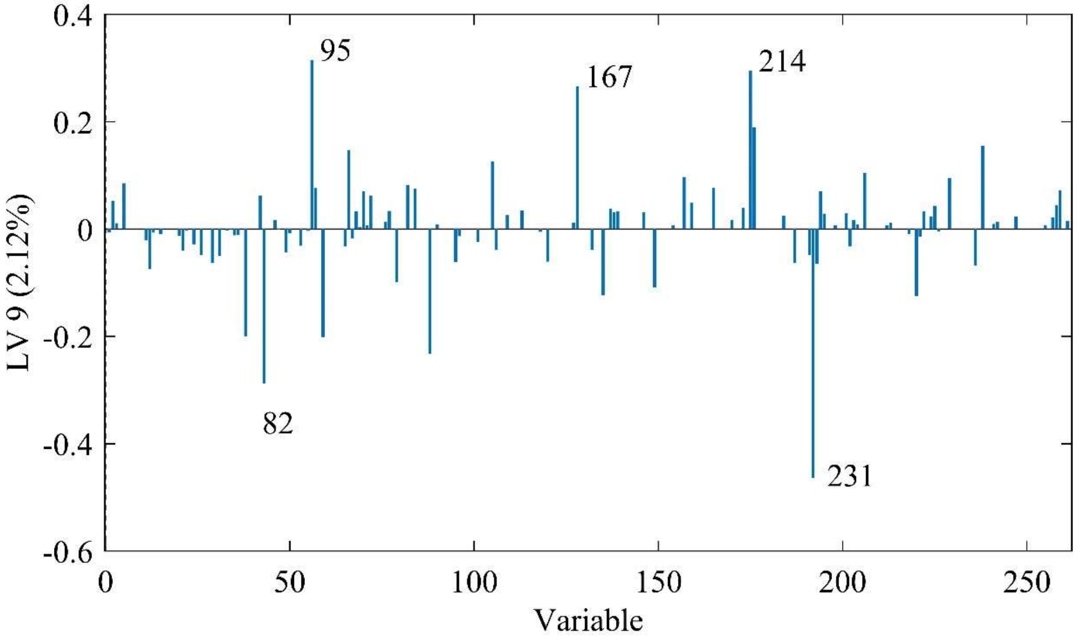

Figure A24: Loading vector 9 of Figure 8

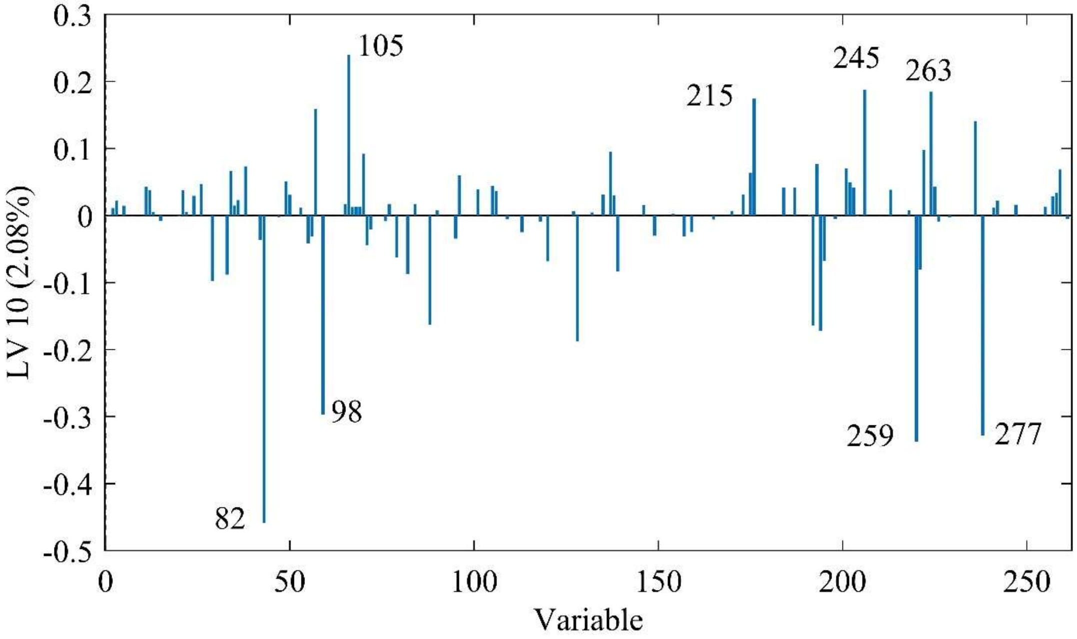

Figure A25: Loading vector 10 of Figure 8

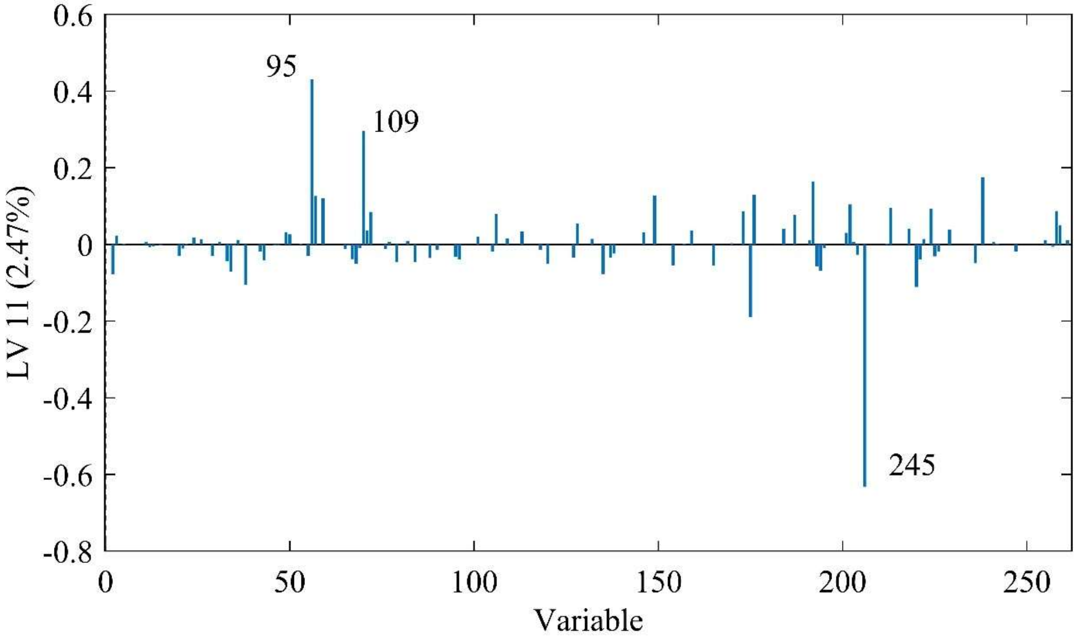

Figure A26: Loading vector 11 of Figure 8

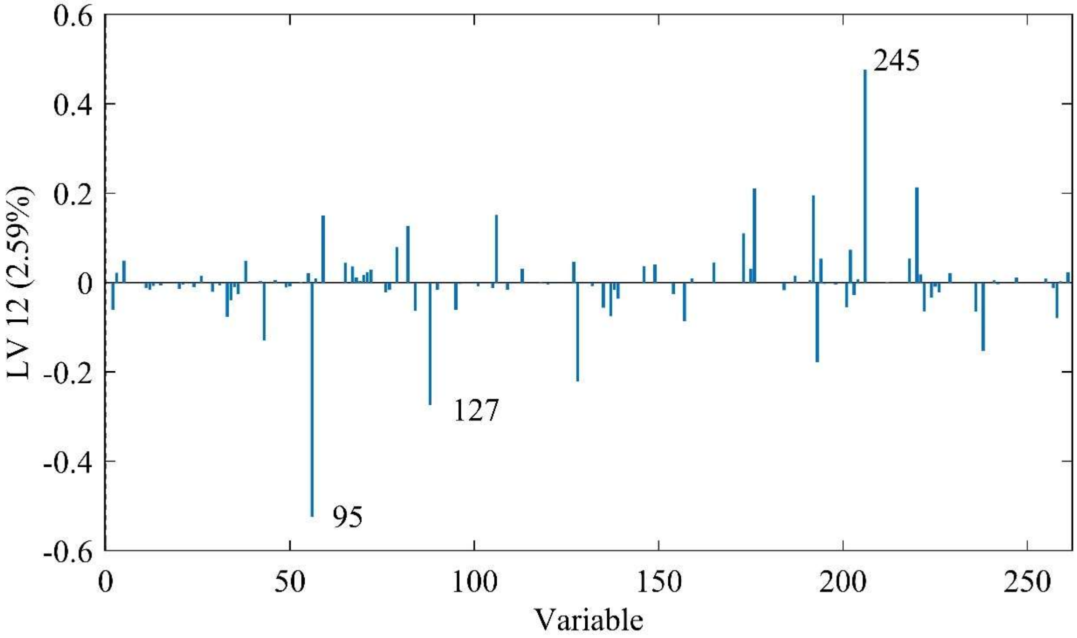

Figure A27: Loading vector 12 of Figure 8

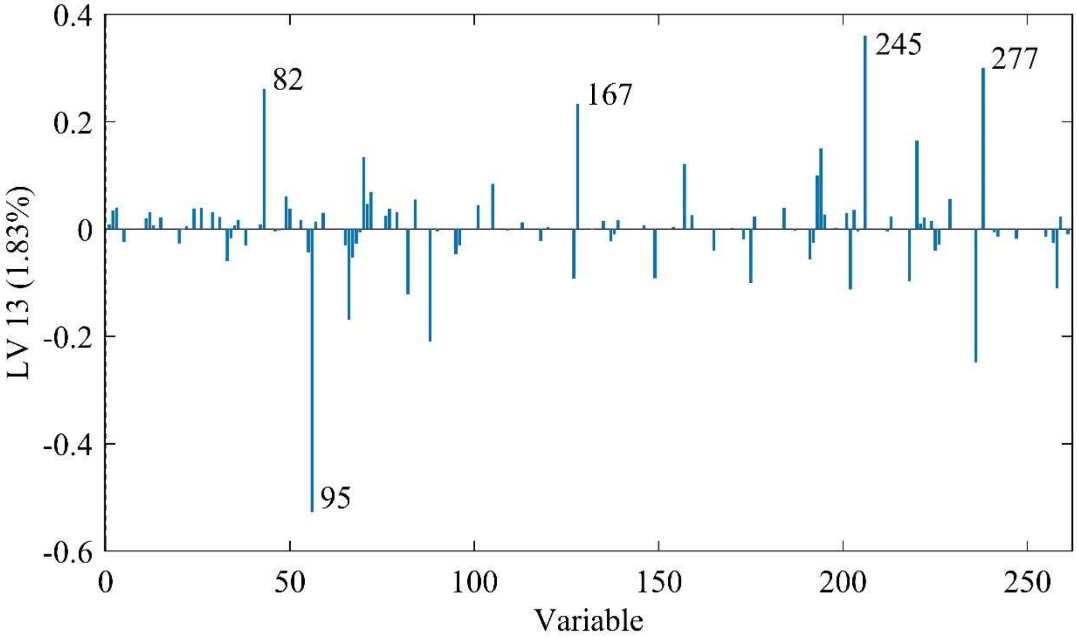

Figure A28: Loading vector 13 of Figure 8

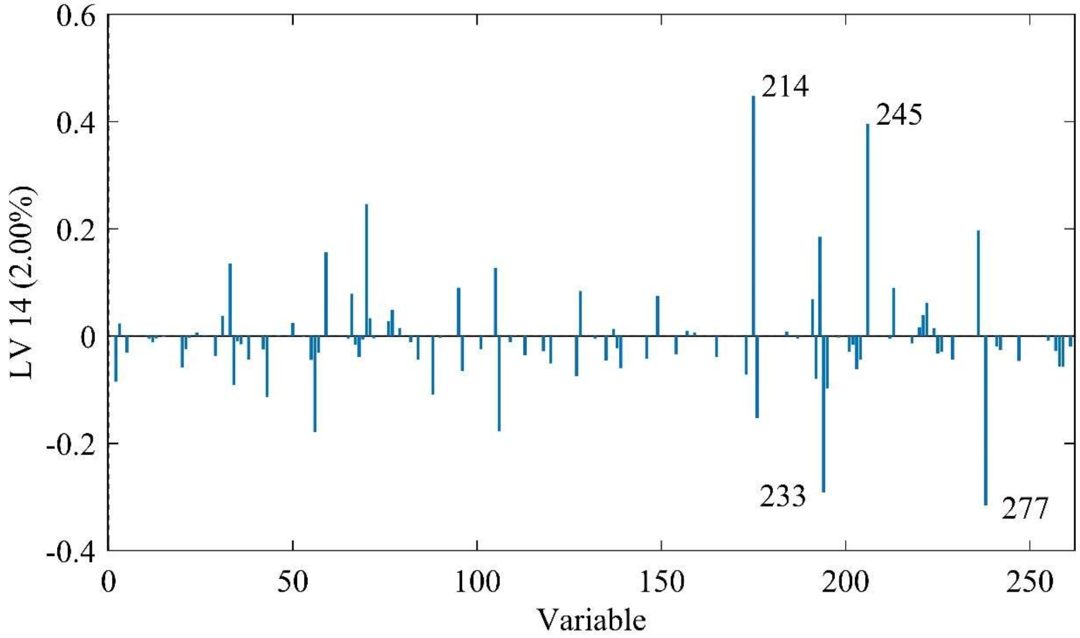

Figure A29: Loading vector 14 of Figure 8

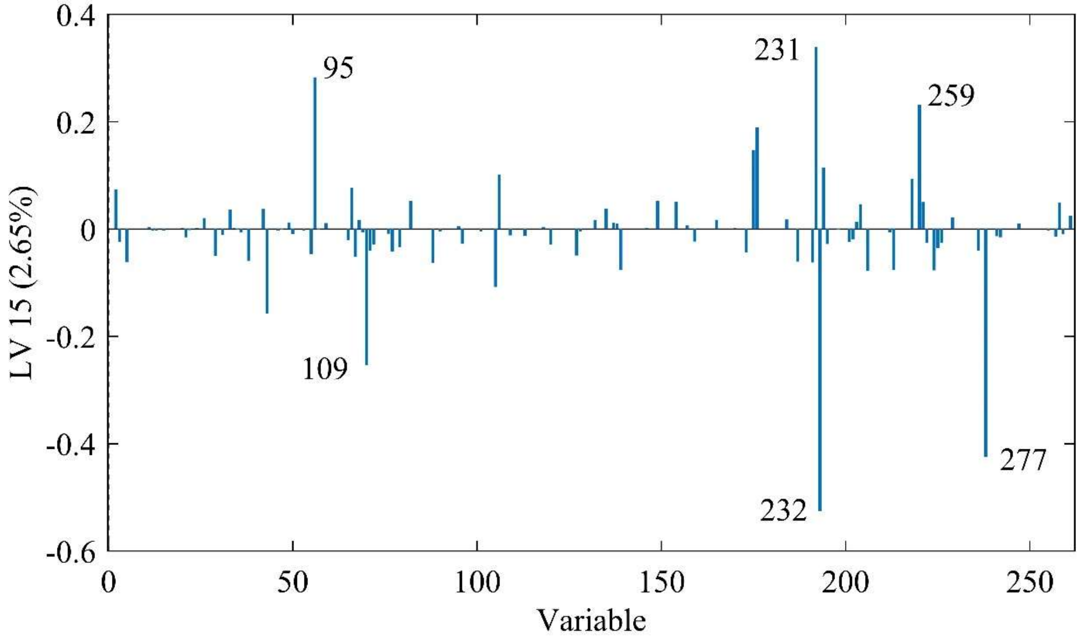

Figure A30: Loading vector 15 of Figure 8

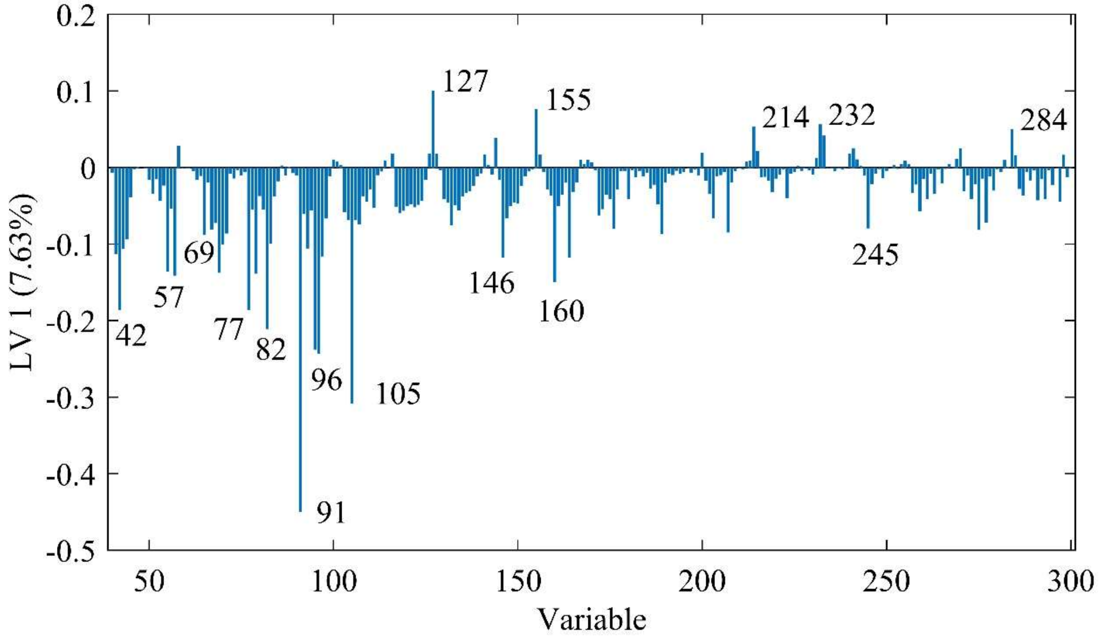

Figure A31: Loading vector 1 of Figure 11a – also shown in the text

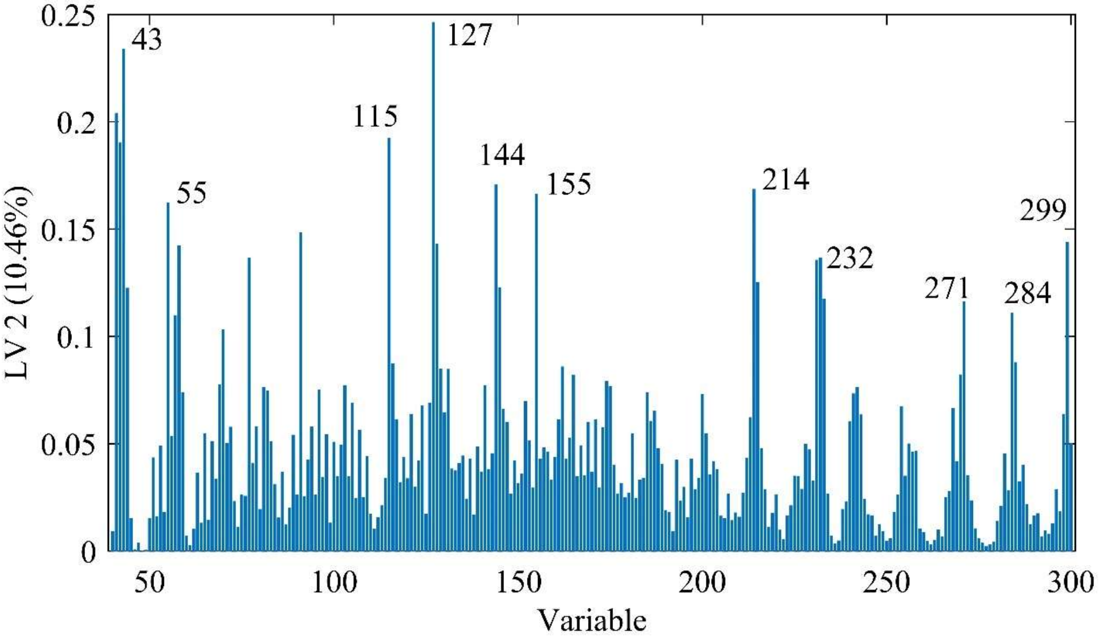

Figure A32: Loading vector 2 of Figure 11a

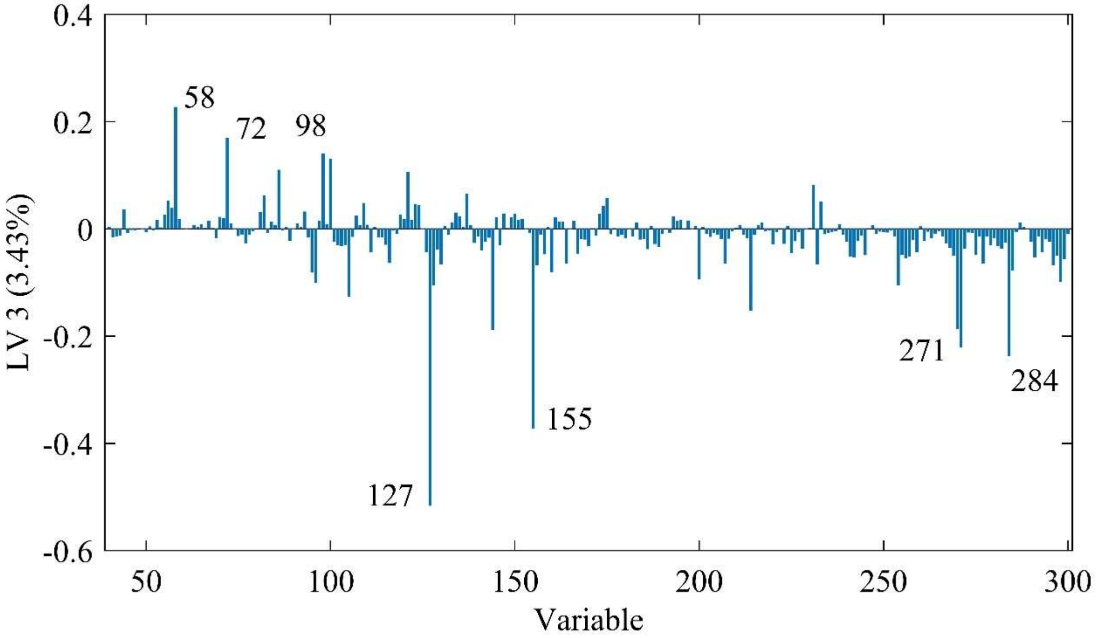

Figure A33: Loading vector 3 of Figure 11a

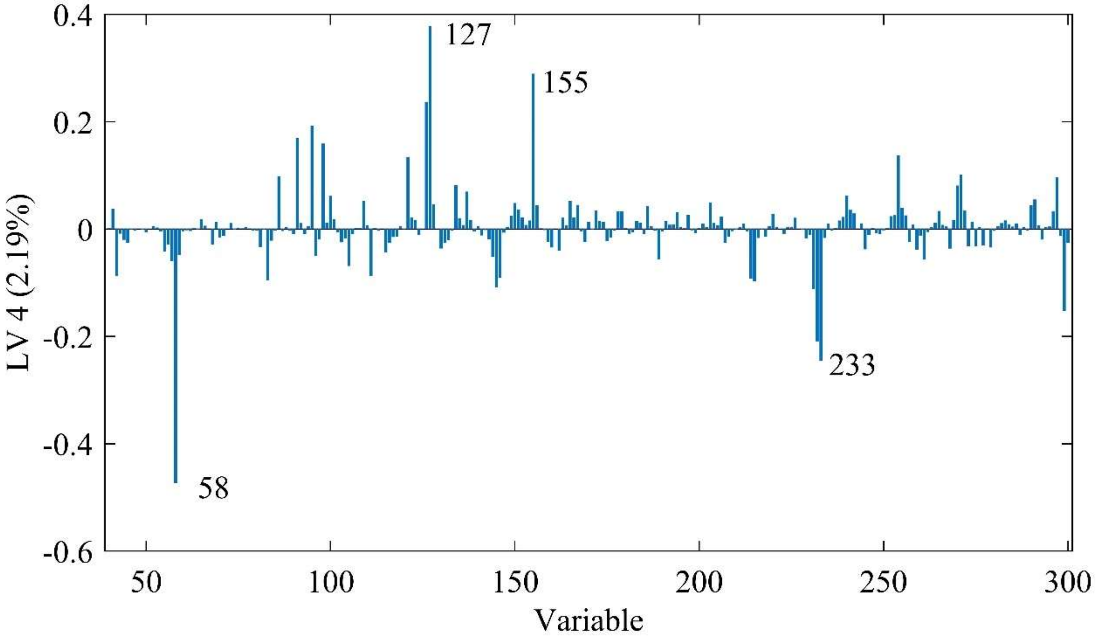

Figure A34: Loading vector 4 of Figure 11a

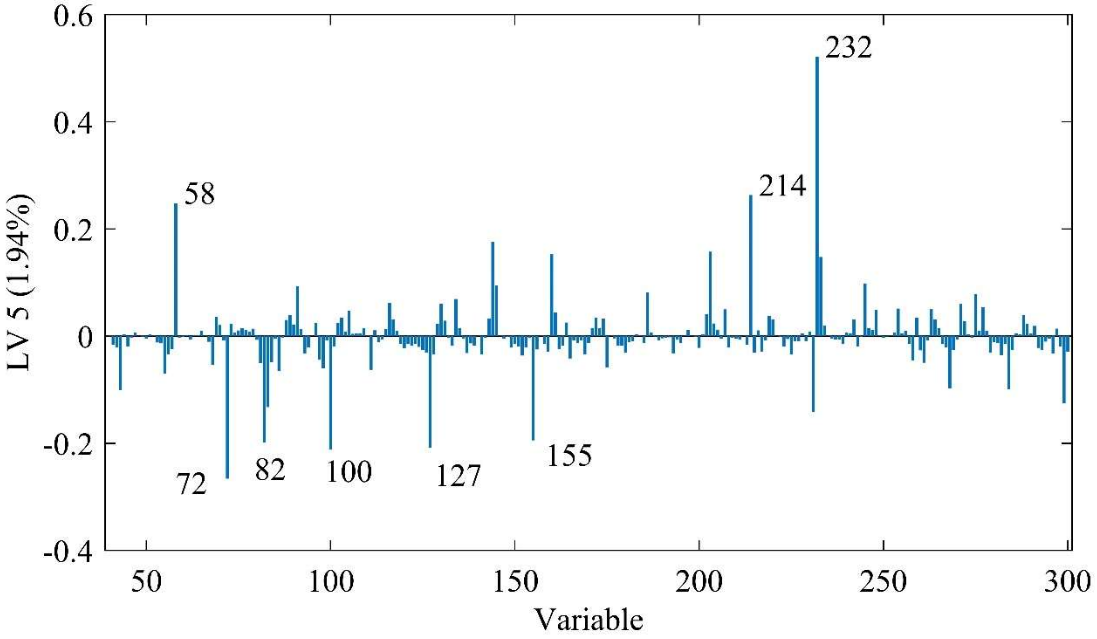

Figure A35: Loading vector 5 of Figure 11a

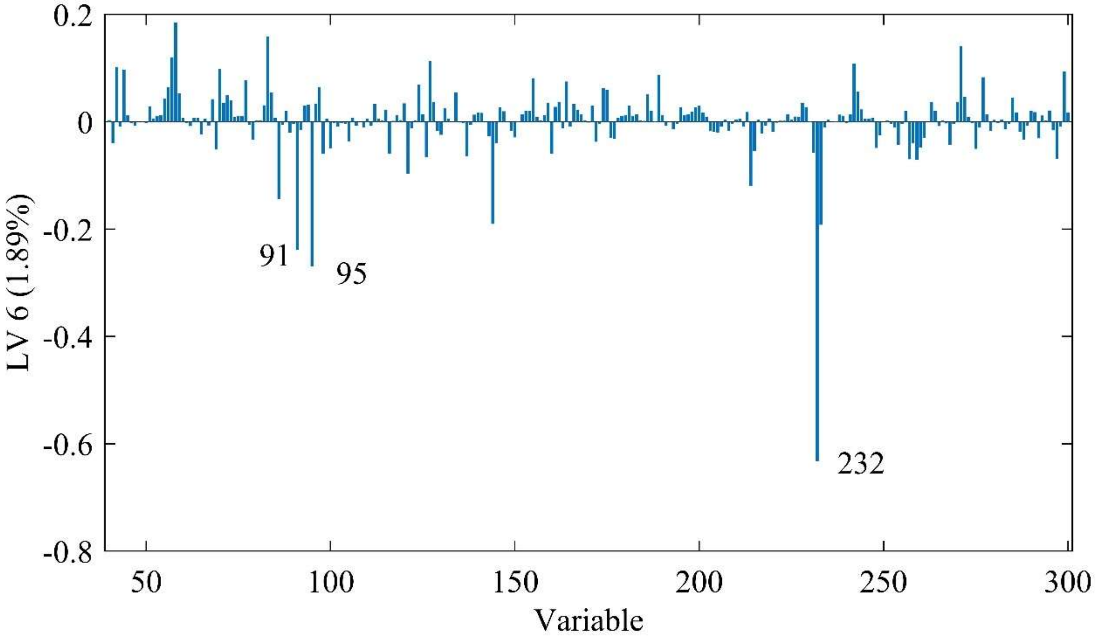

Figure A36: Loading vector 6 of Figure 11a

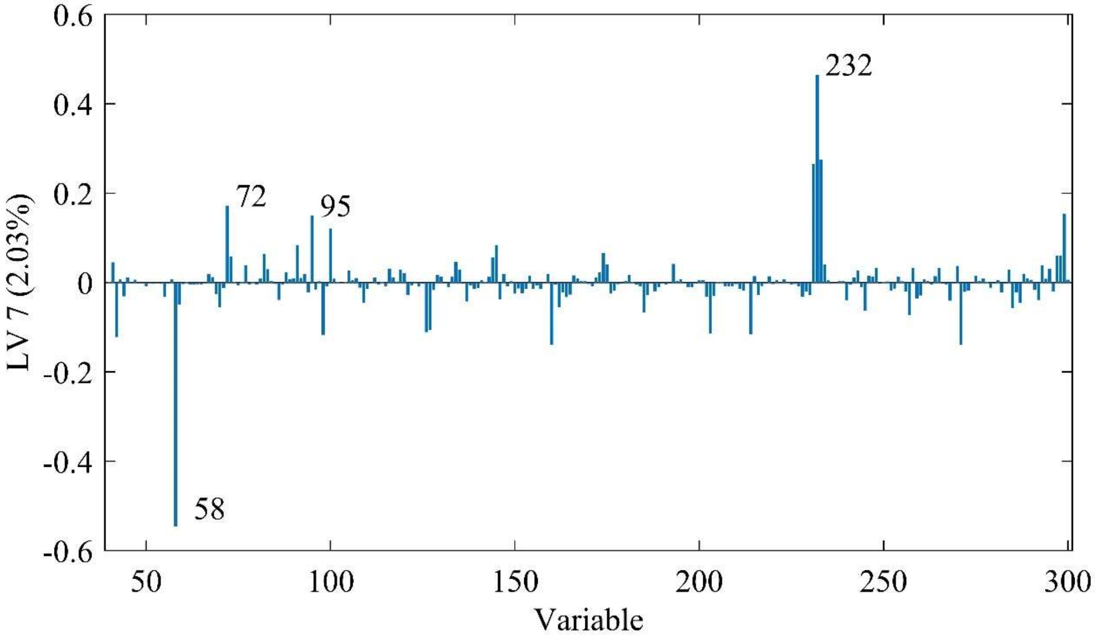

Figure A37: Loading vector 7 of Figure 11a

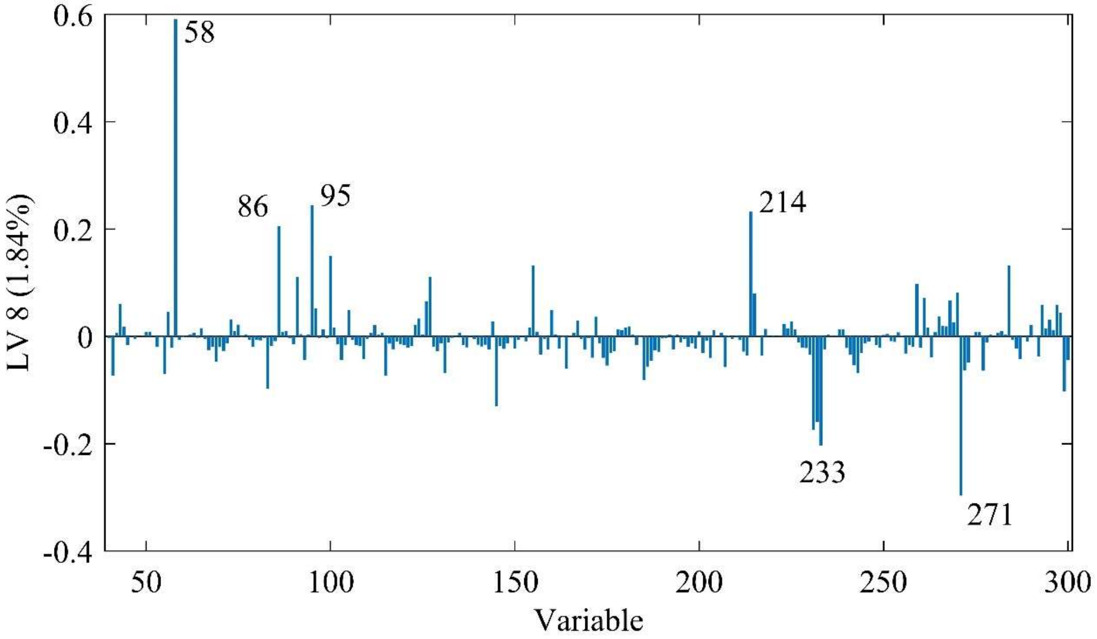

Figure A38: Loading vector 8 of Figure 11a

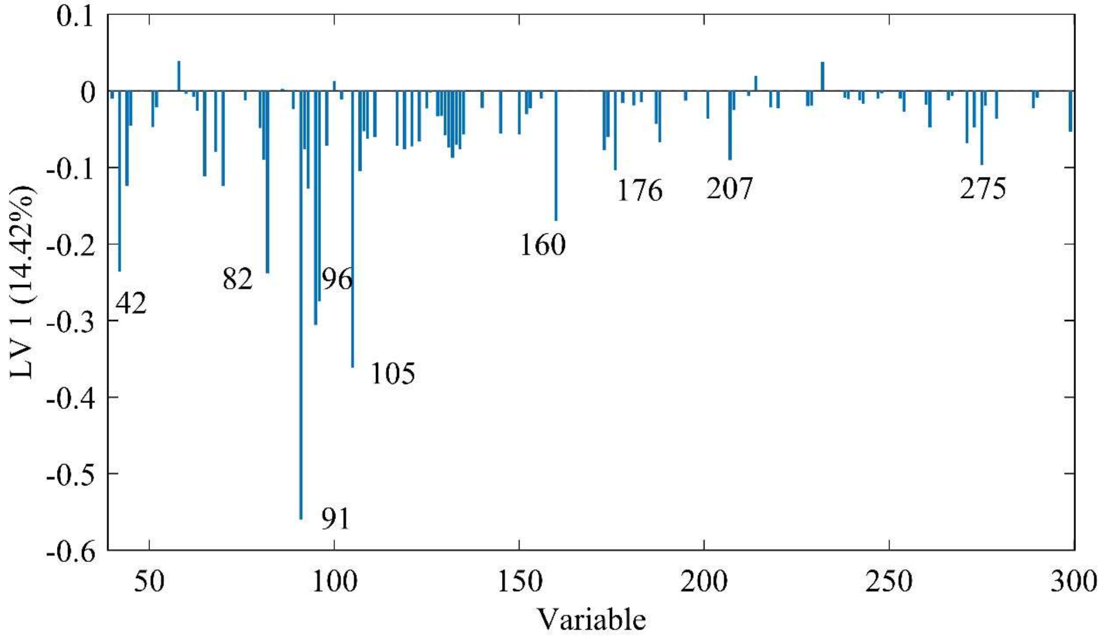

Figure A39: Loading vector 1 of Figure 12a – also shown in the text

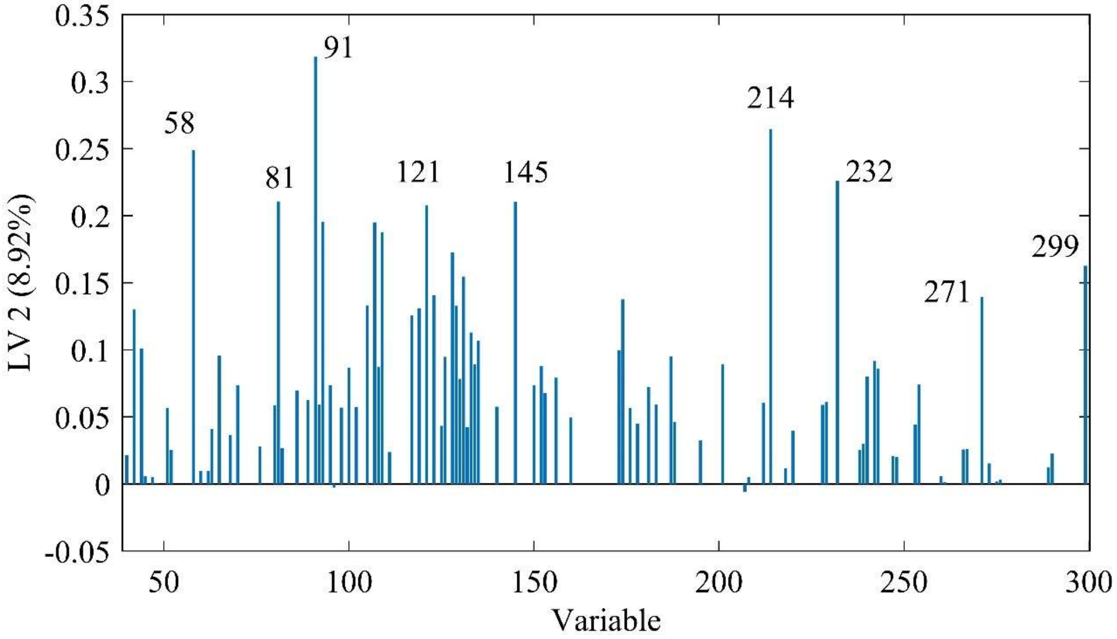

Figure A40: Loading vector 2 of Figure 12a

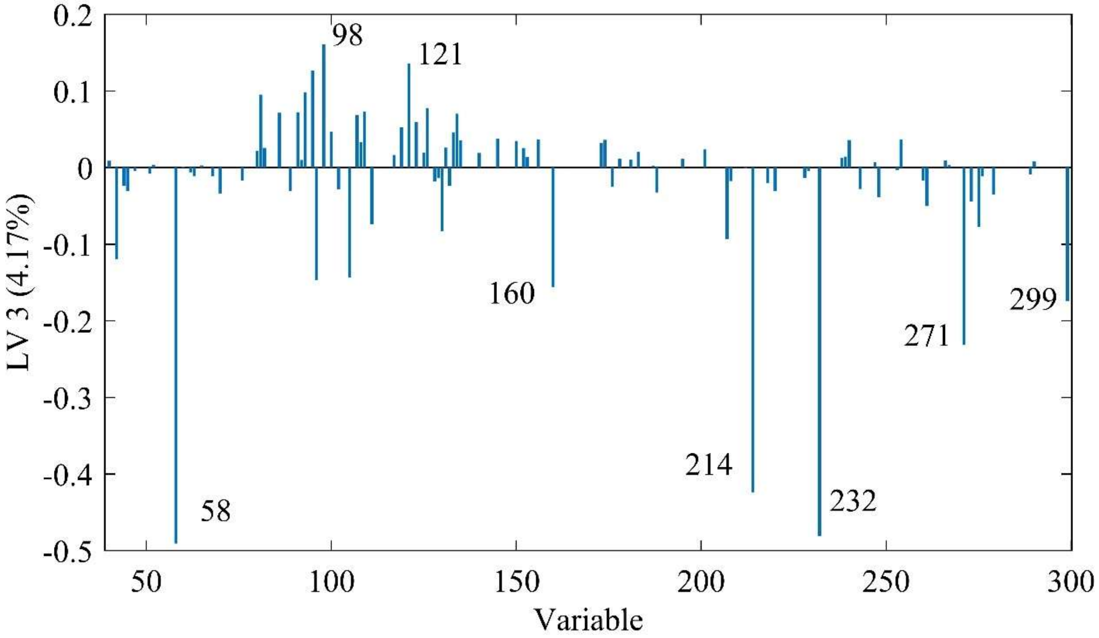

Figure A41: Loading vector 3 of Figure 12a

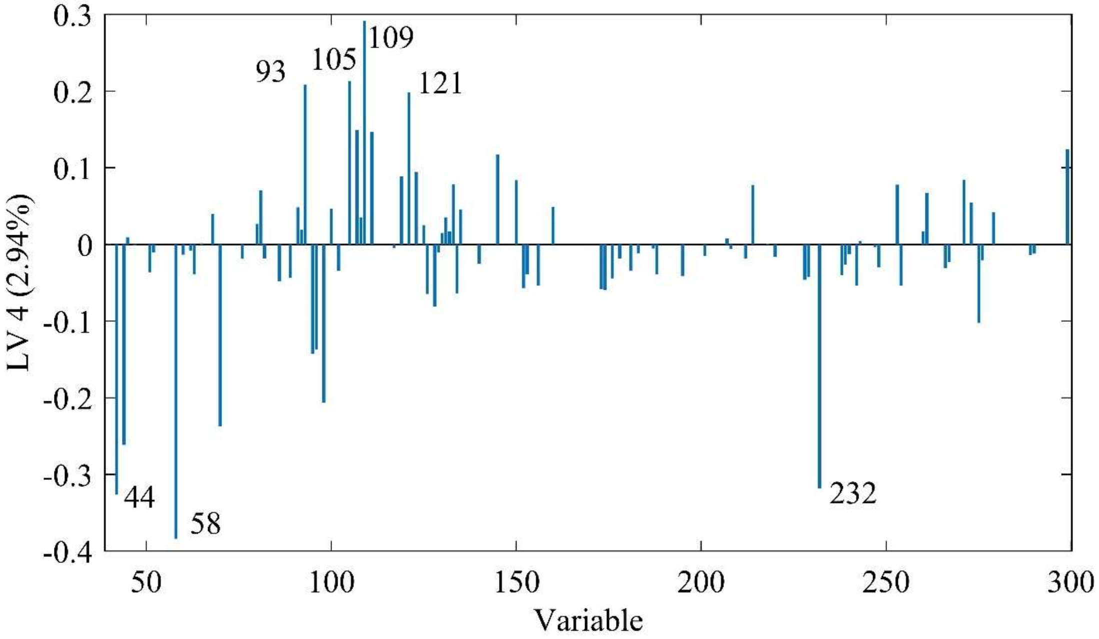

Figure A42: Loading vector 4 of Figure 12a

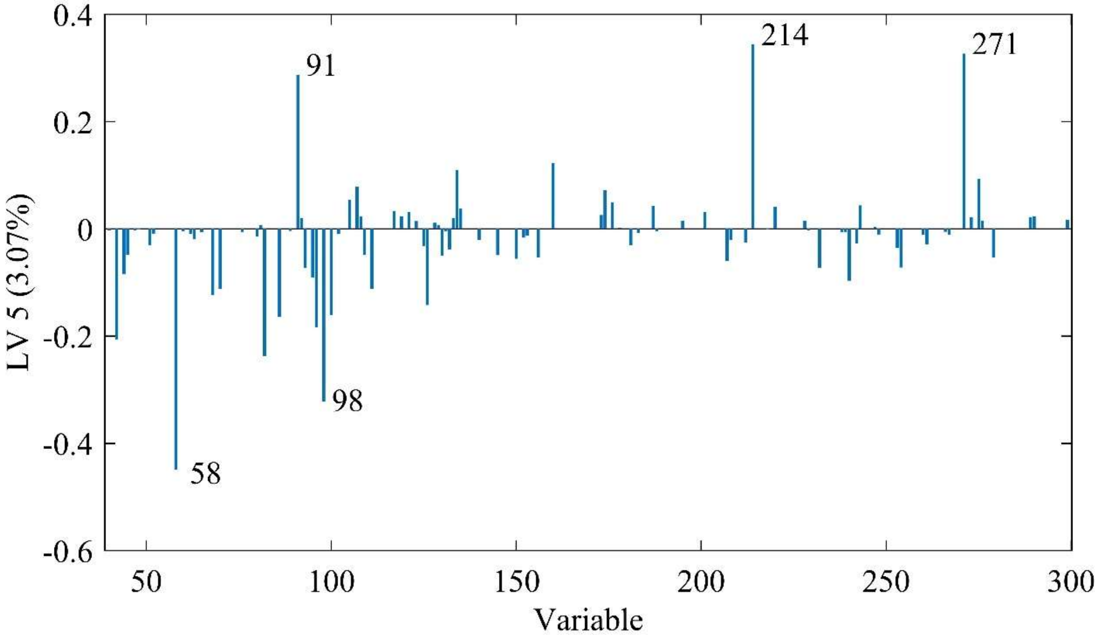

Figure A43: Loading vector 5 of Figure 12a

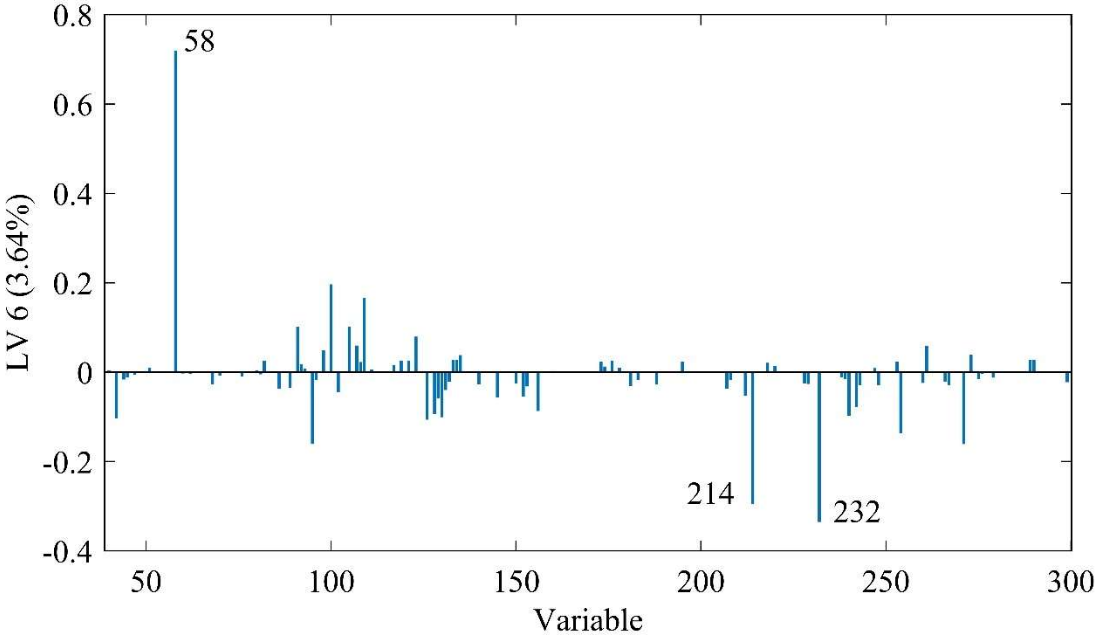

Figure A44: Loading vector 6 of Figure 12a

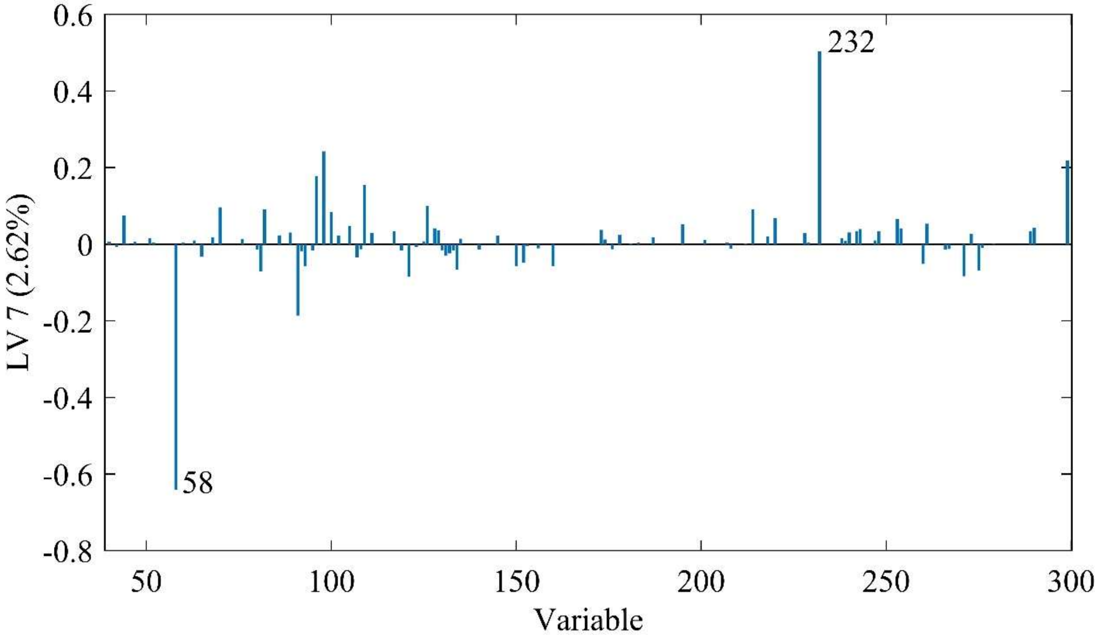

Figure A45: Loading vector 7 of Figure 12a

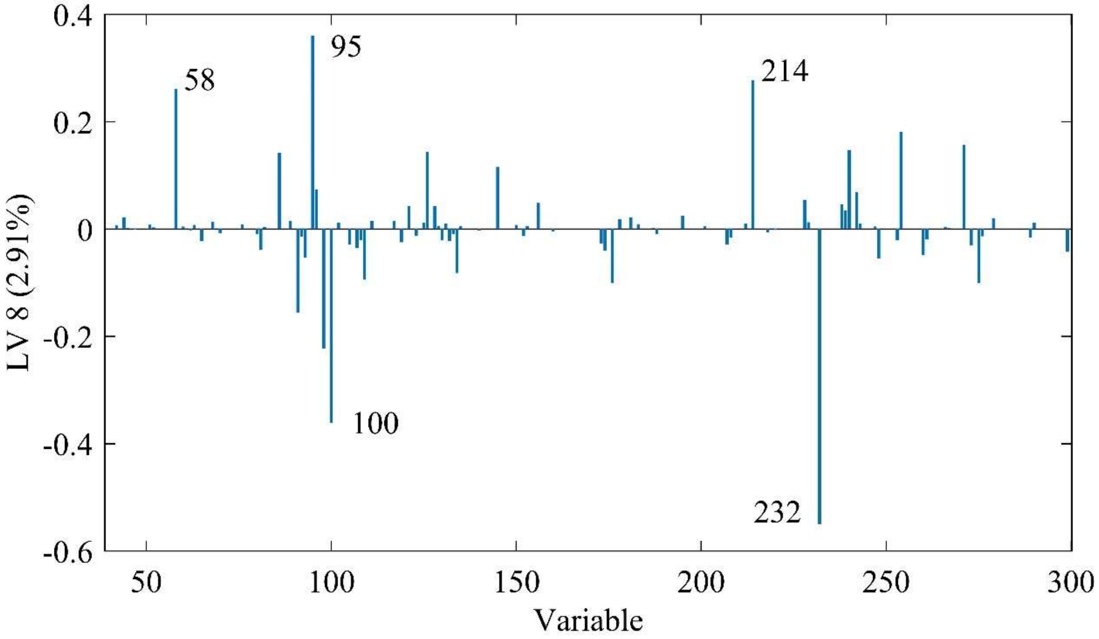

Figure A46: Loading vector 8 of Figure 12a

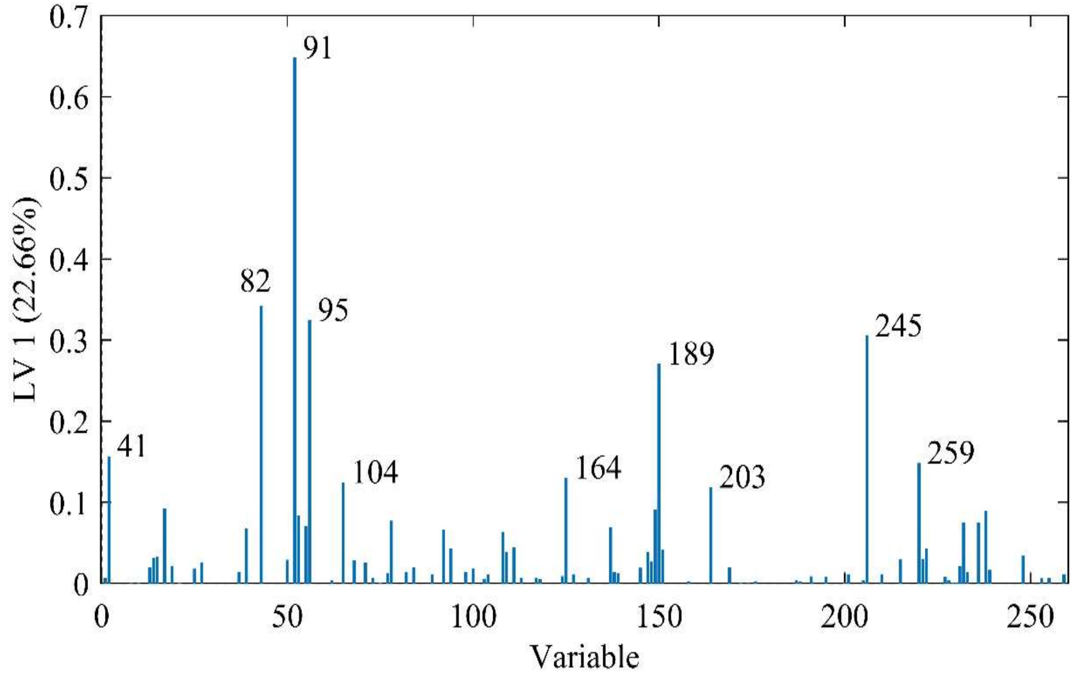

Figure A47: Loading vector 1 of Figure 18 – also shown in the text

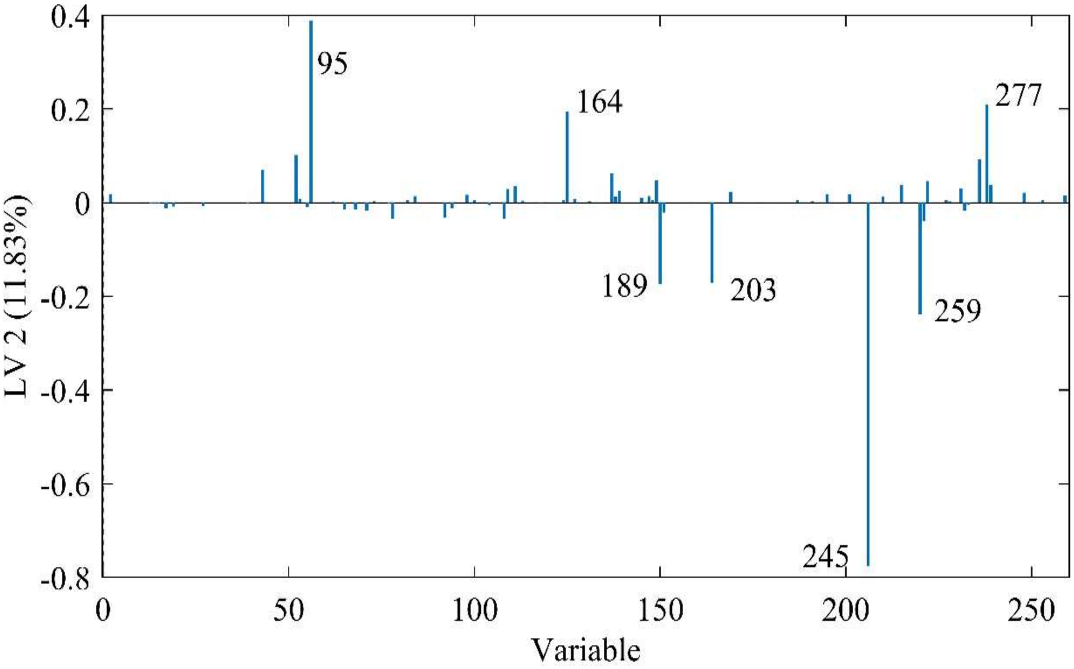

Figure A48: Loading vector 2 of Figure 18 – also shown in the text

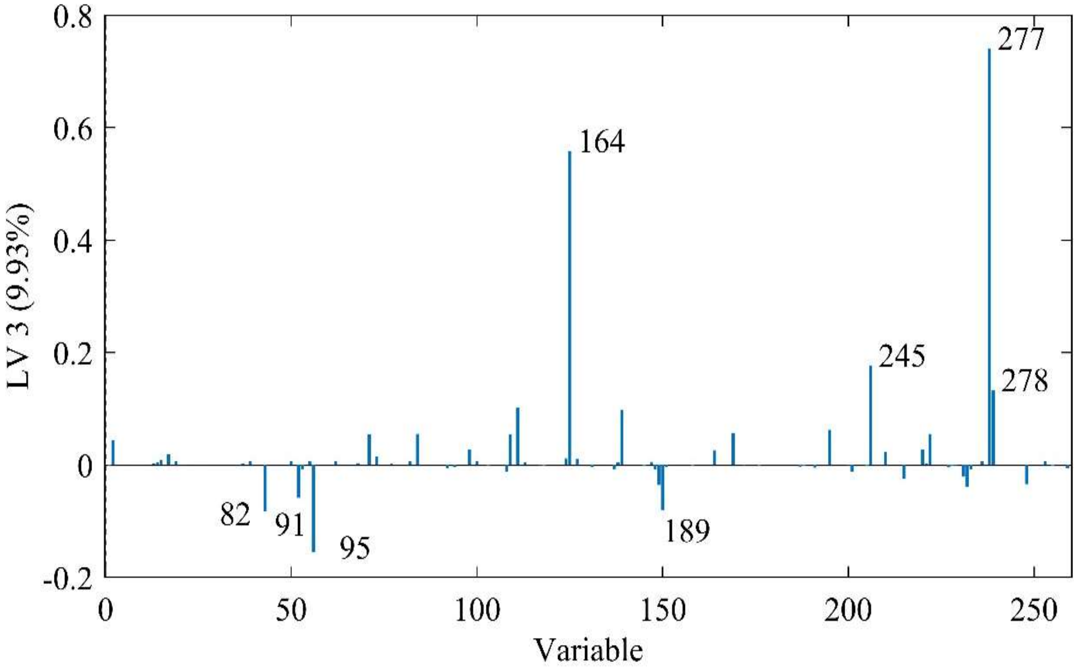

Figure A49: Loading vector 3 of Figure 18 – also shown in the text

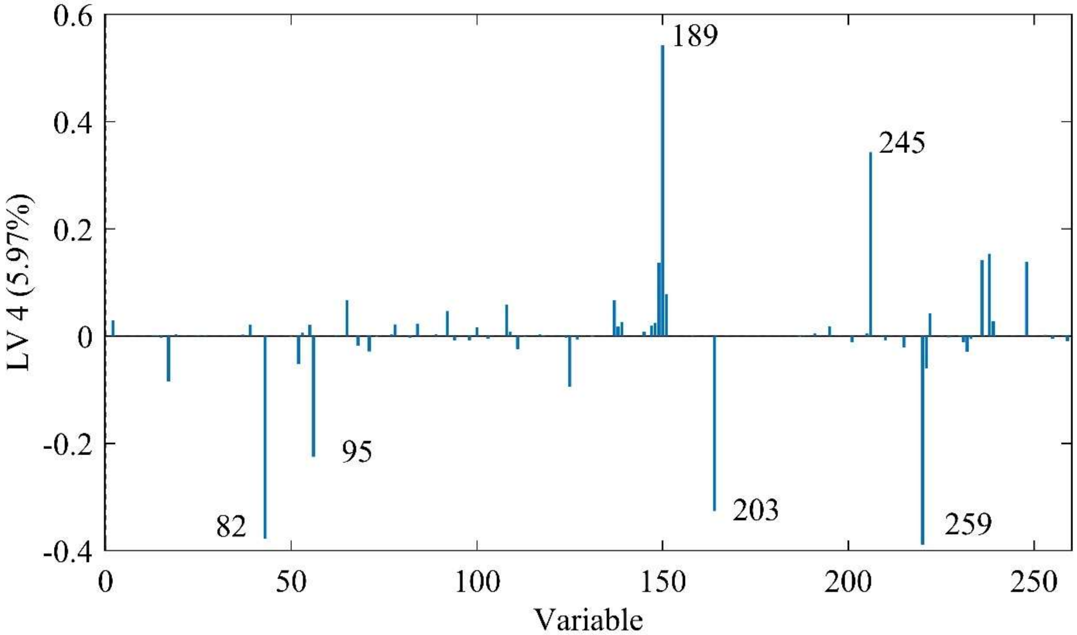

Figure A50: Loading vector 4 of Figure 18

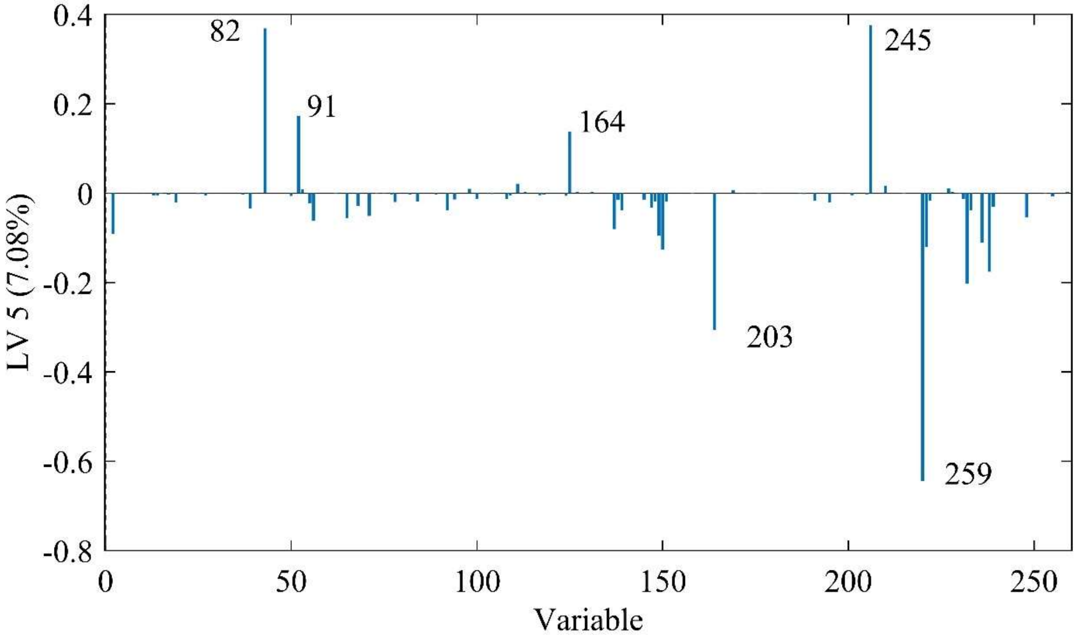

Figure A51: Loading vector 5 of Figure 18

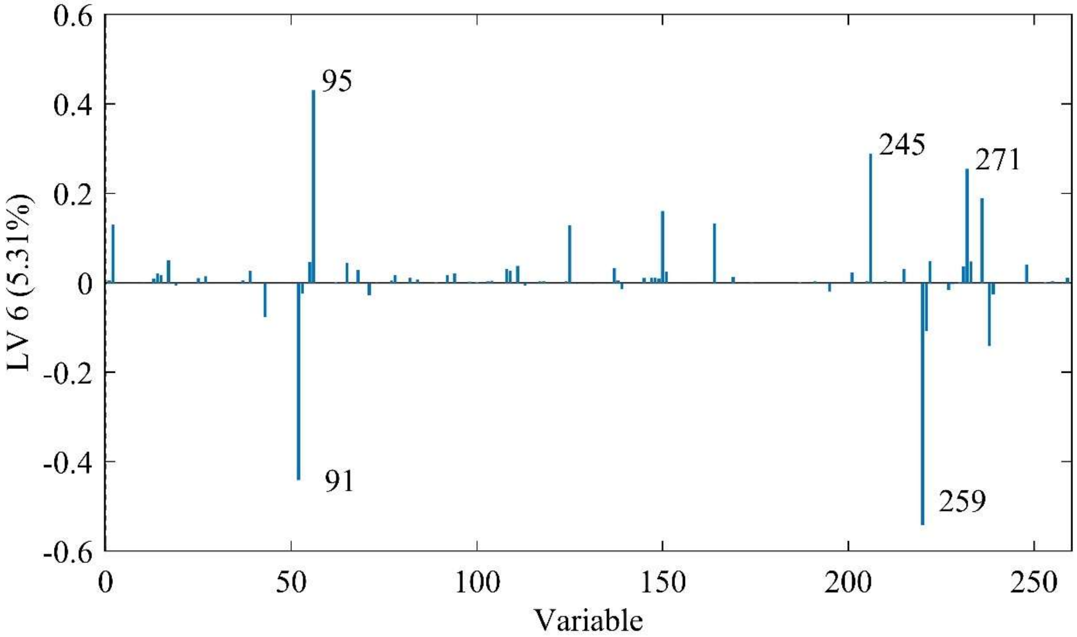

Figure A52: Loading vector 6 of Figure 18

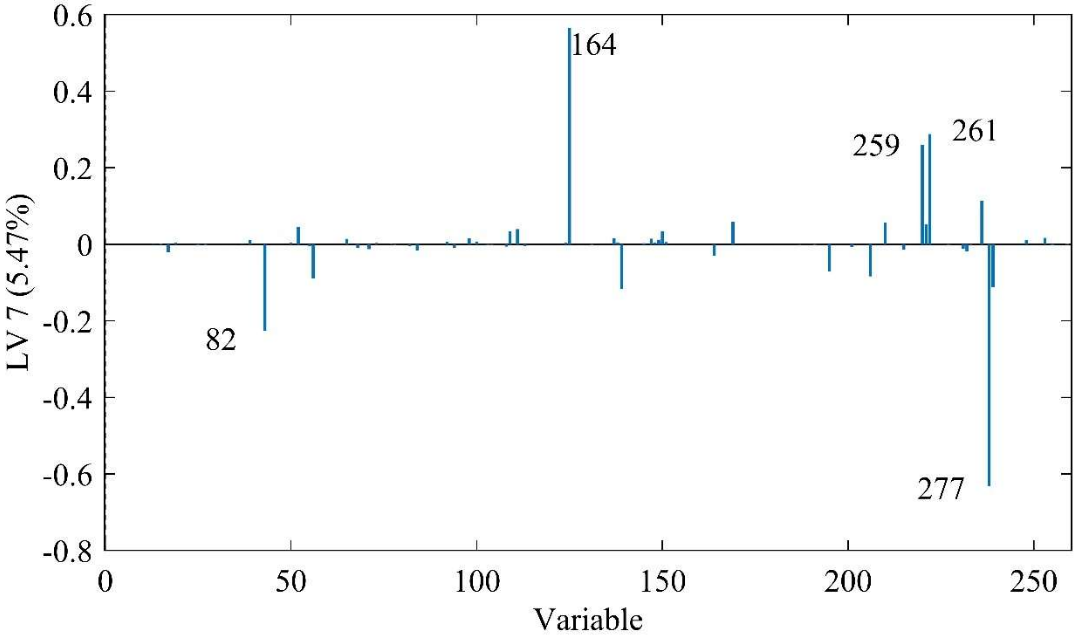

Figure A53: Loading vector 7 of Figure 18

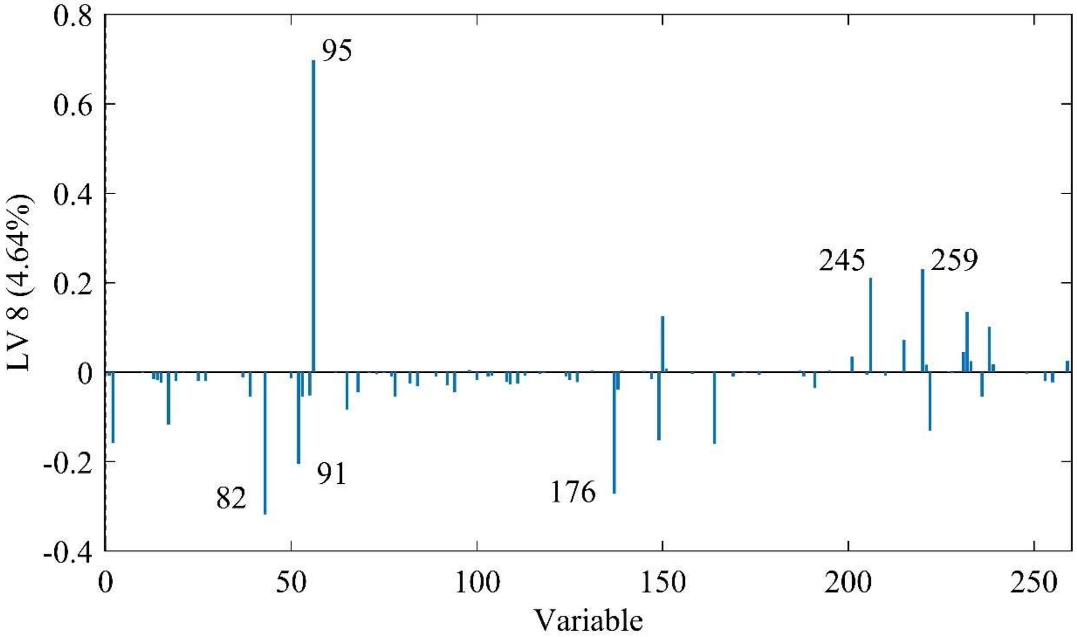

Figure A54: Loading vector 8 of Figure 18

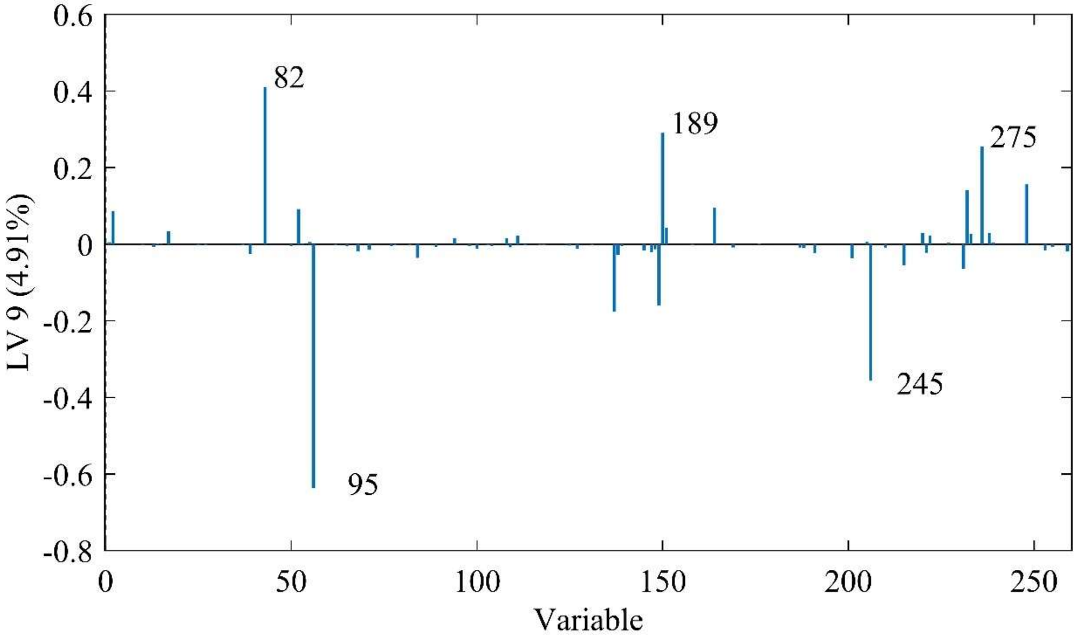

Figure A55: Loading vector 9 of Figure 18

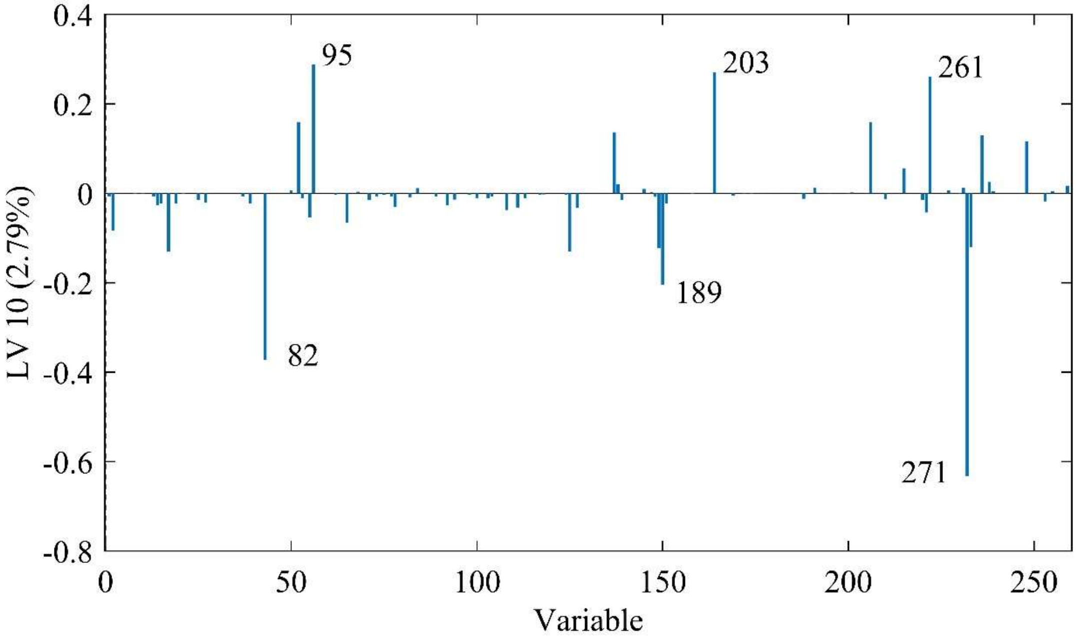

Figure A56: Loading vector 10 of Figure 18

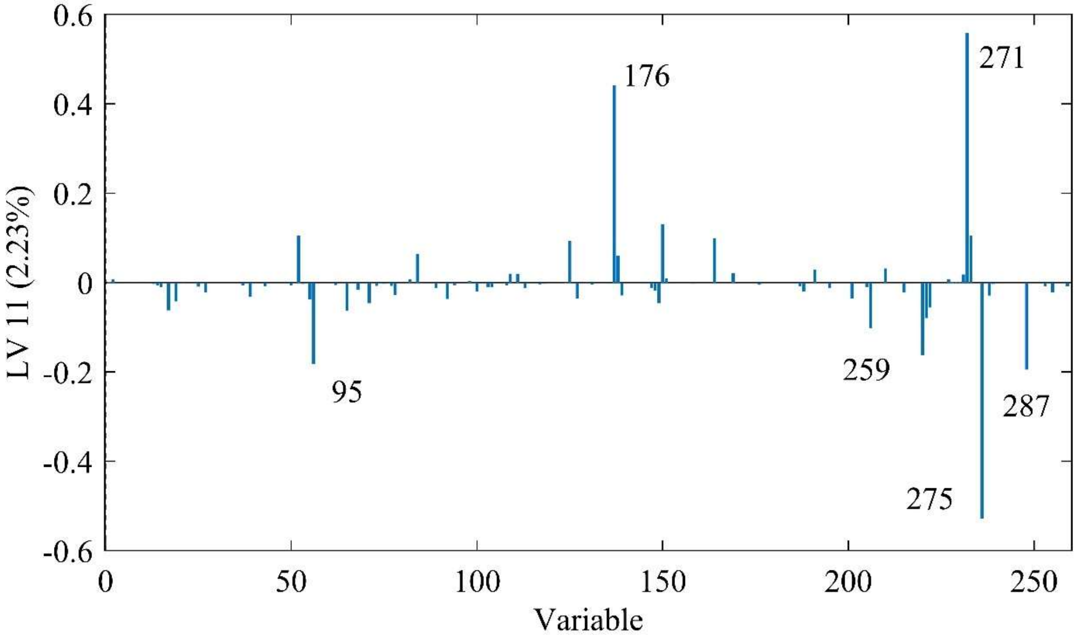

Figure A57: Loading vector 11 of Figure 18

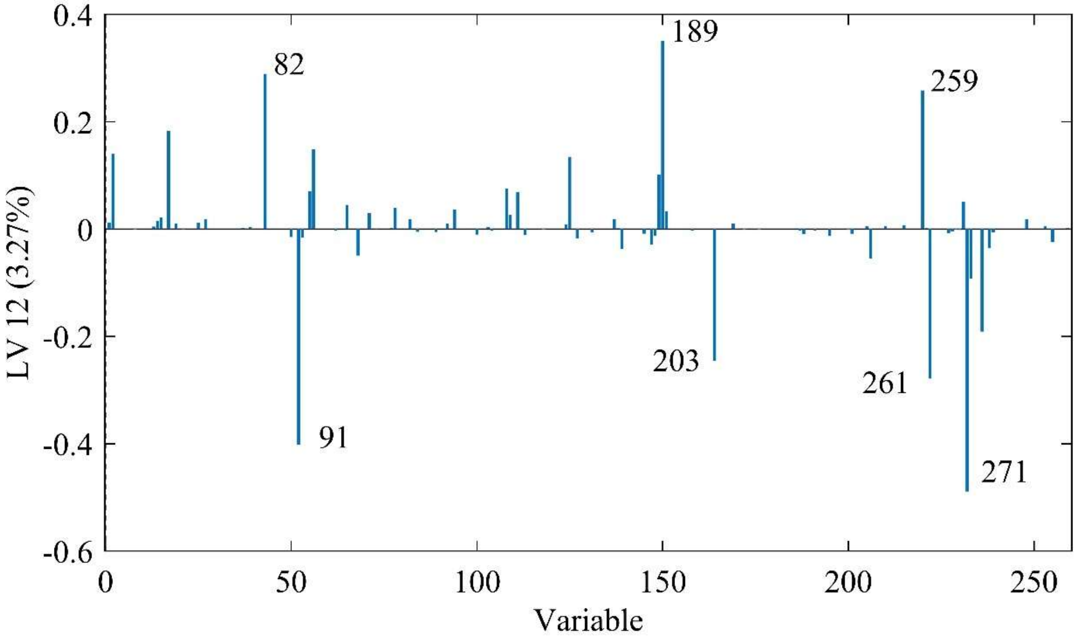

Figure A58: Loading vector 12 of Figure 18

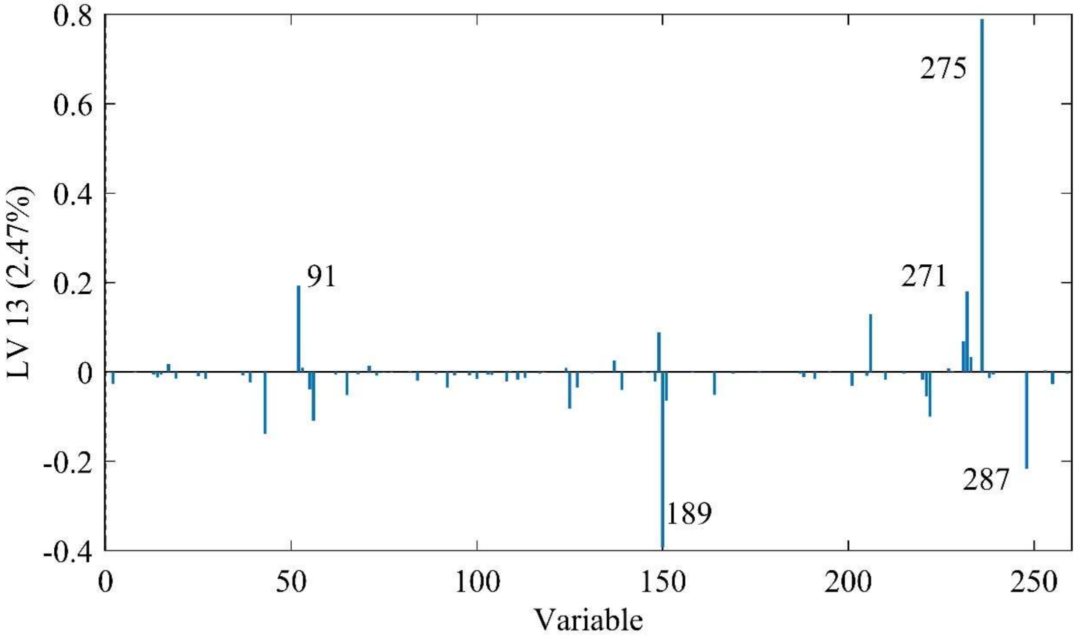

Figure A59: Loading vector 13 of Figure 18

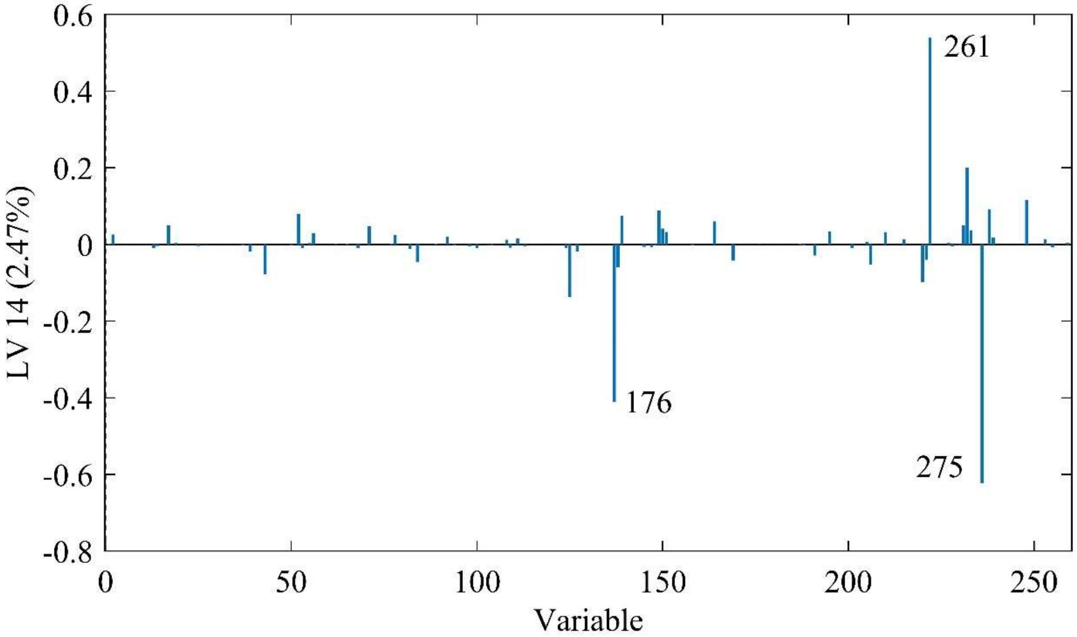

Figure A60: Loading vector 14 of Figure 18

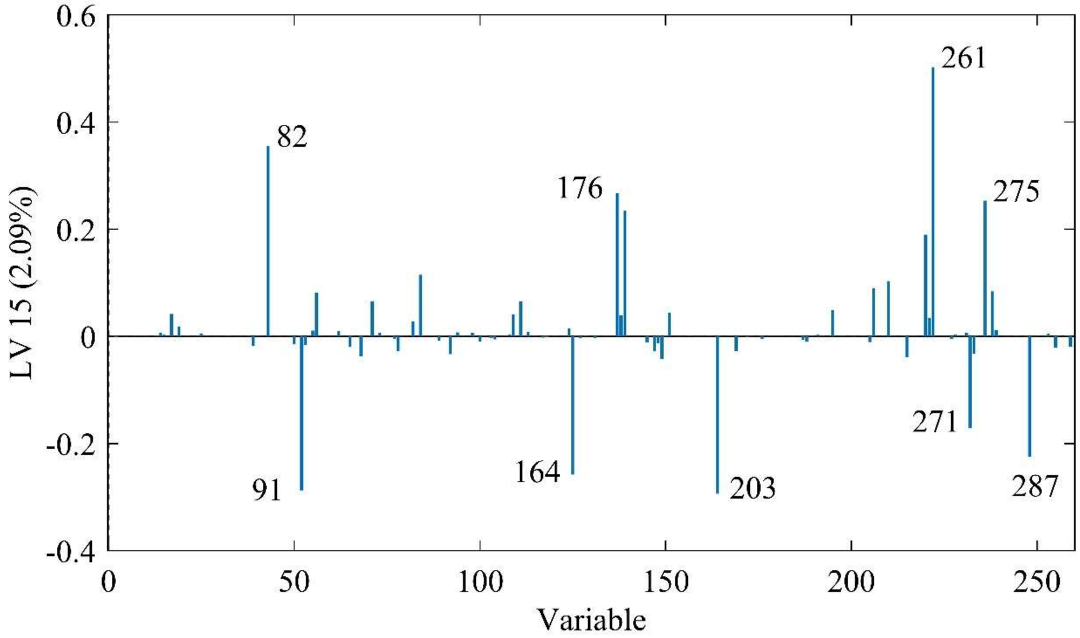

Figure A61: Loading vector 15 of Figure 18

APPENDIX B: *Y* LOADING VECTORS

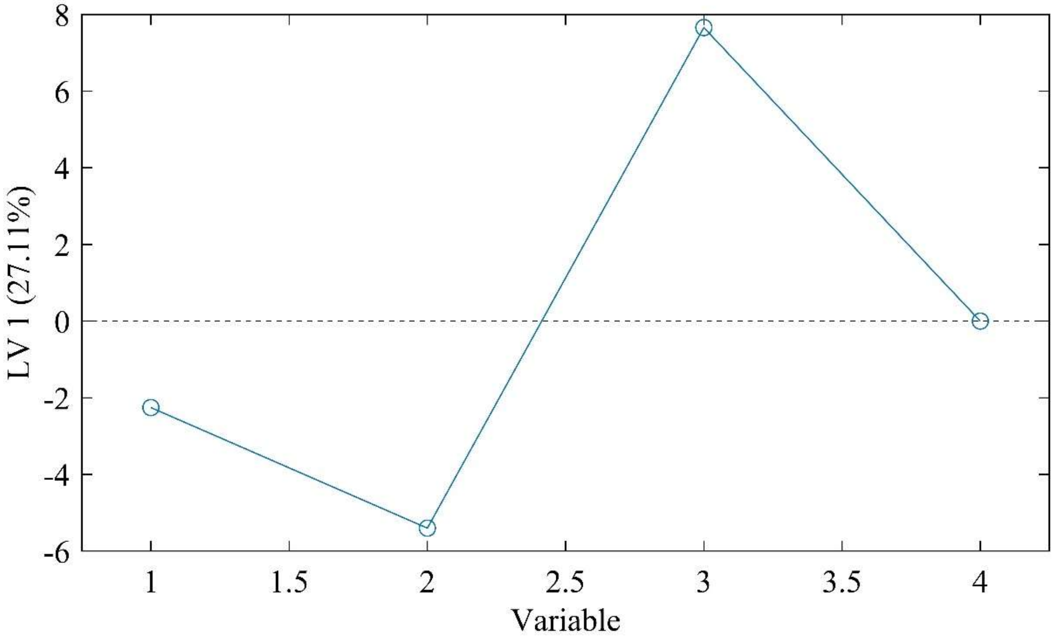

Figure B1: Y Loading vector 1 of Figure 6

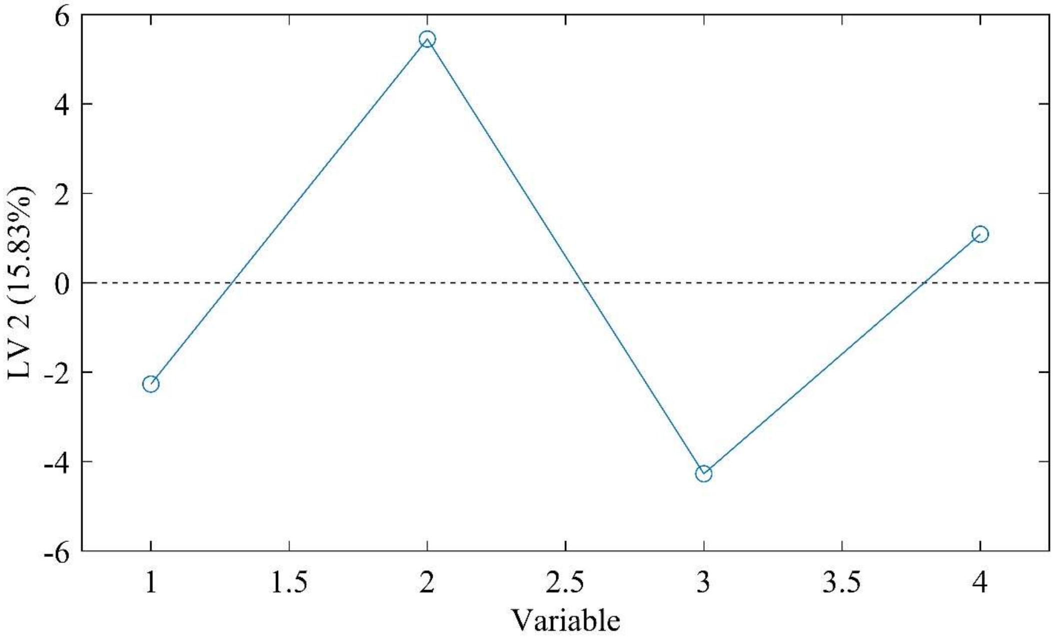

Figure B2: Y Loading vector 2 of Figure 6

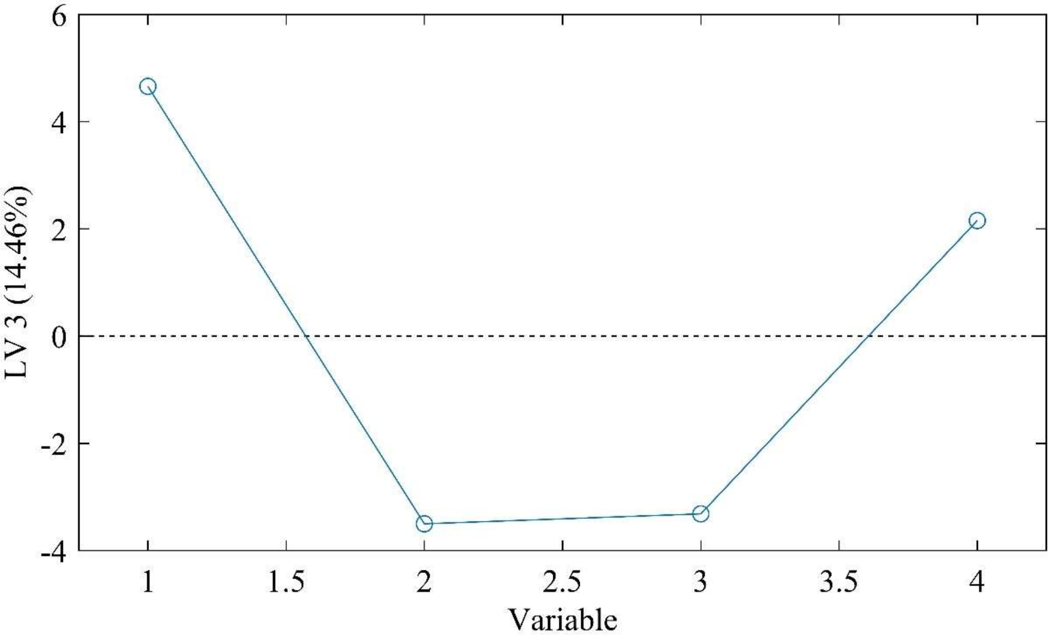

Figure B3: Y Loading vector 3 of Figure 6

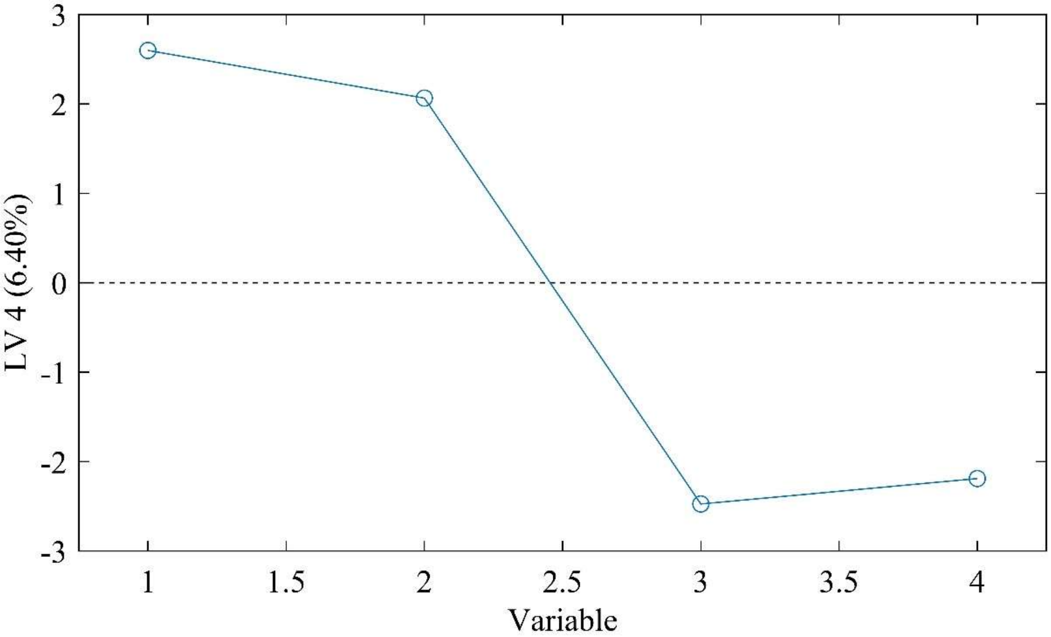

Figure B4: Y Loading vector 4 of Figure 6

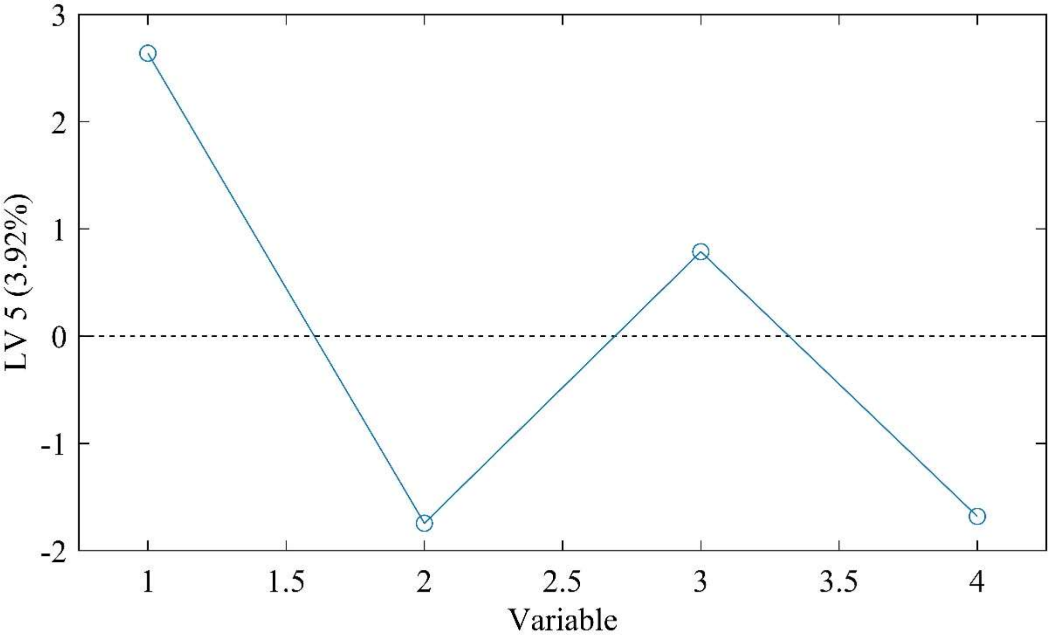

Figure B5: Y Loading vector 5 of Figure 6

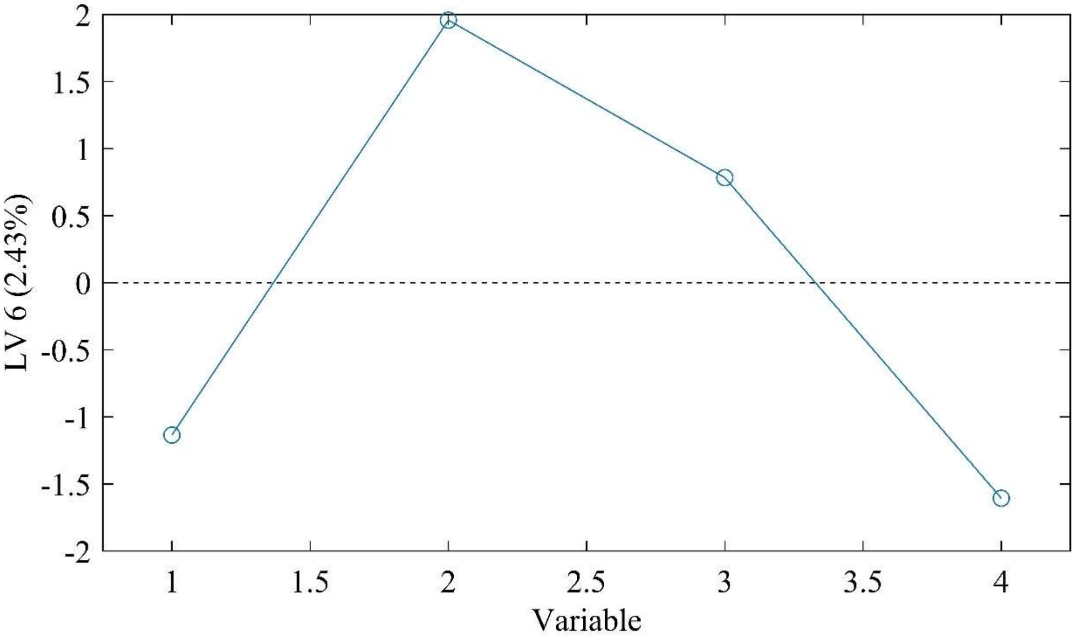

Figure B6: Y Loading vector 6 of Figure 6

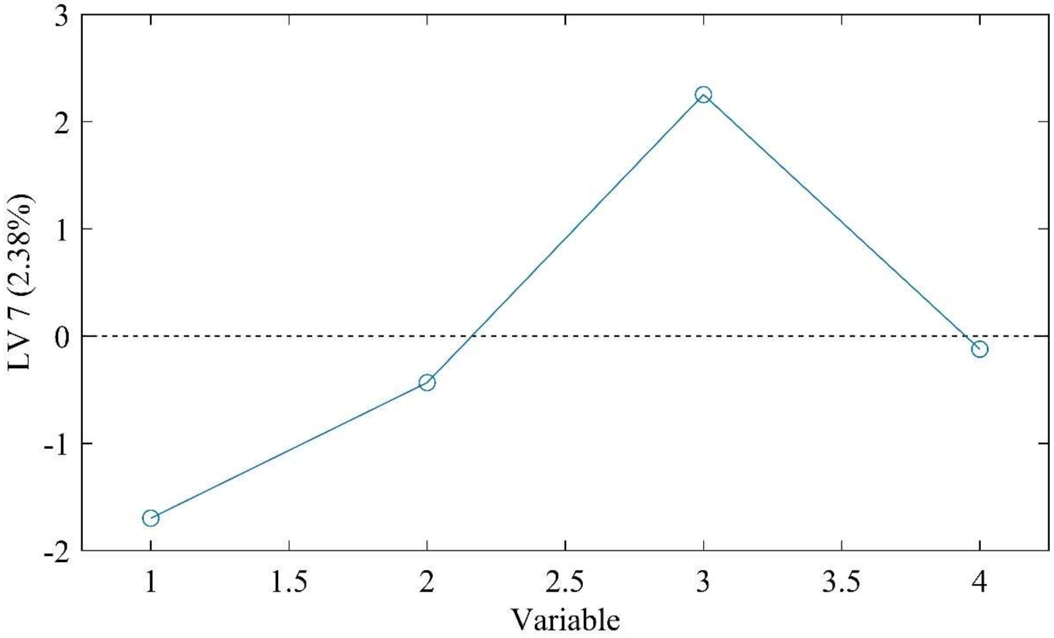

Figure B7: Y Loading vector 7 of Figure 6

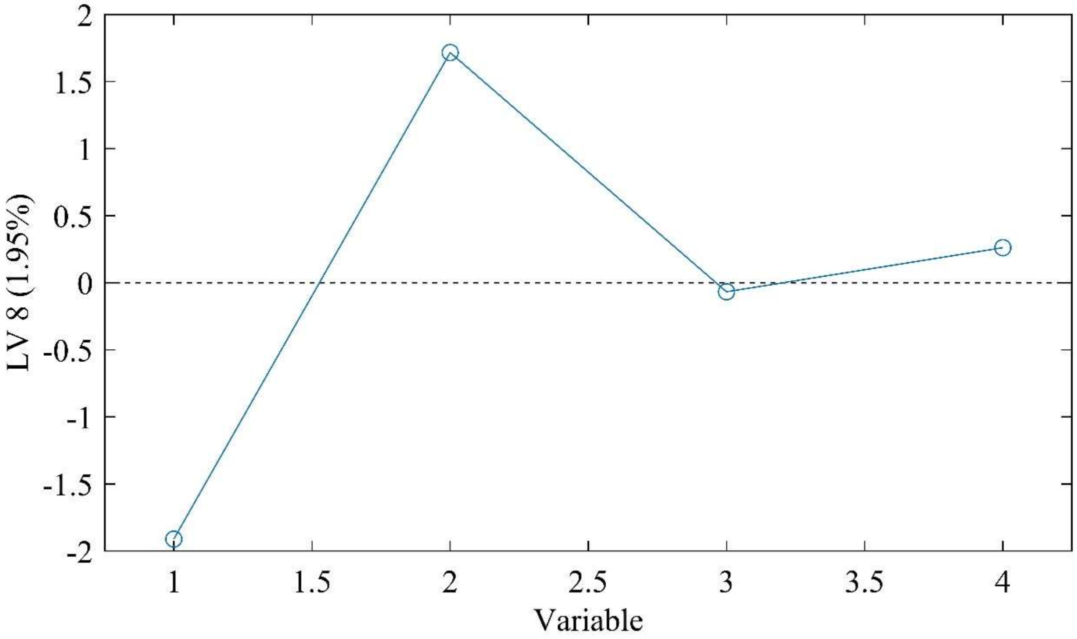

Figure B8: Y Loading vector 8 of Figure 6

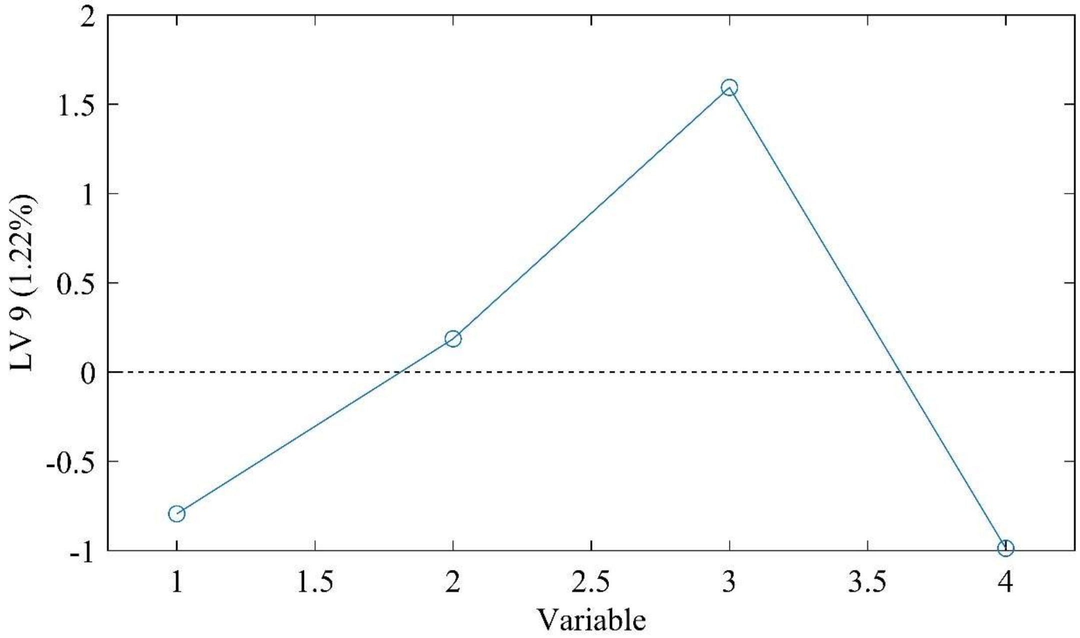

Figure B9: Y Loading vector 9 of Figure 6

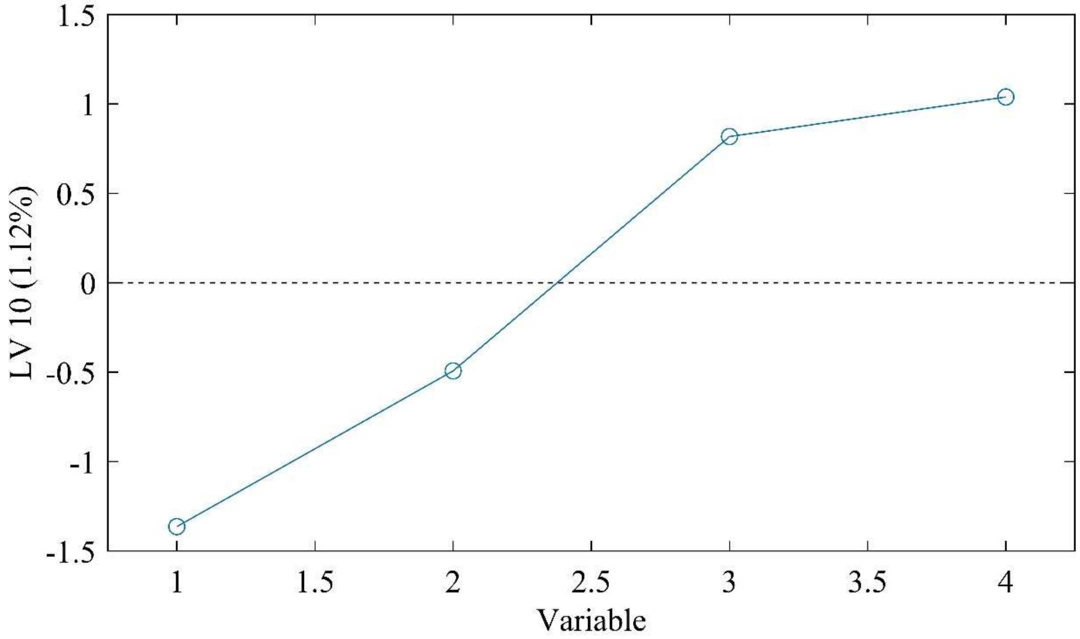

Figure B10: Y Loading vector 10 of Figure 6

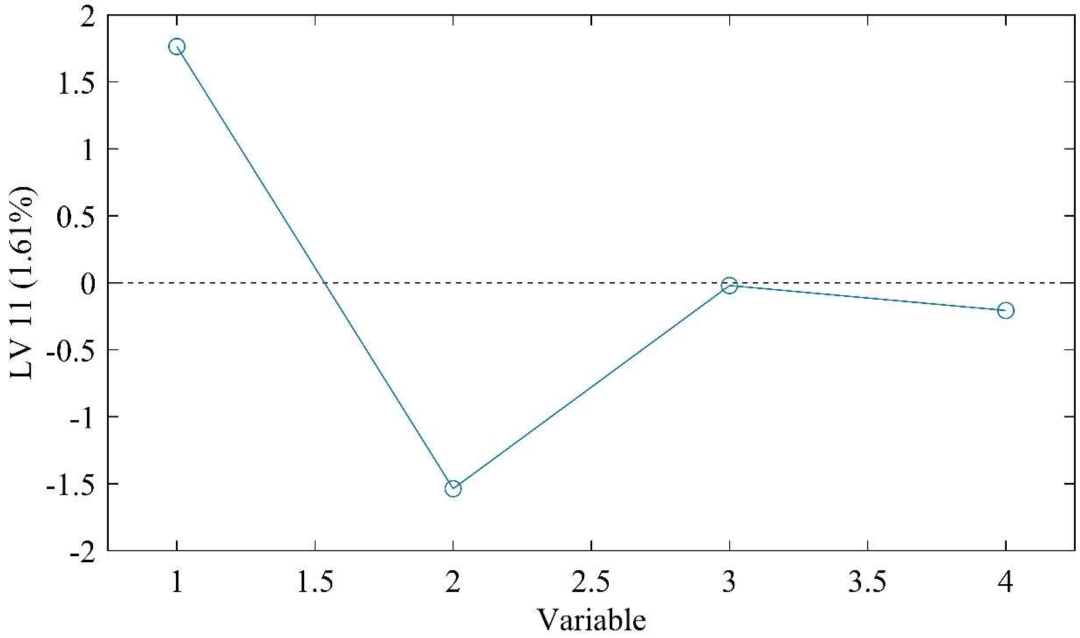

Figure B11: Y Loading vector 11 of Figure 6

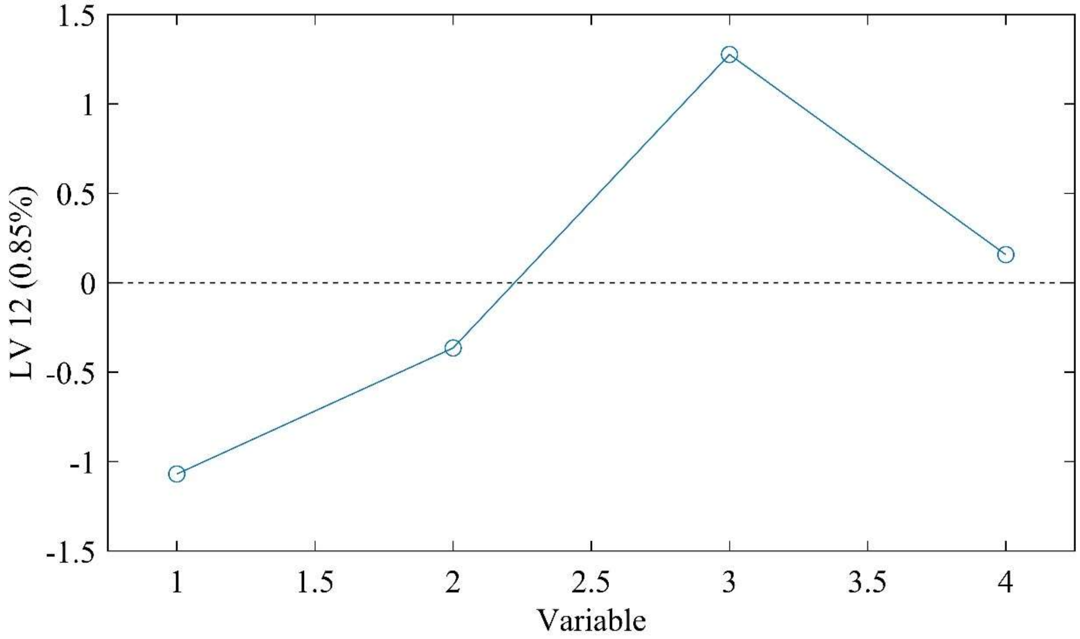

Figure B12: Y Loading vector 12 of Figure 6

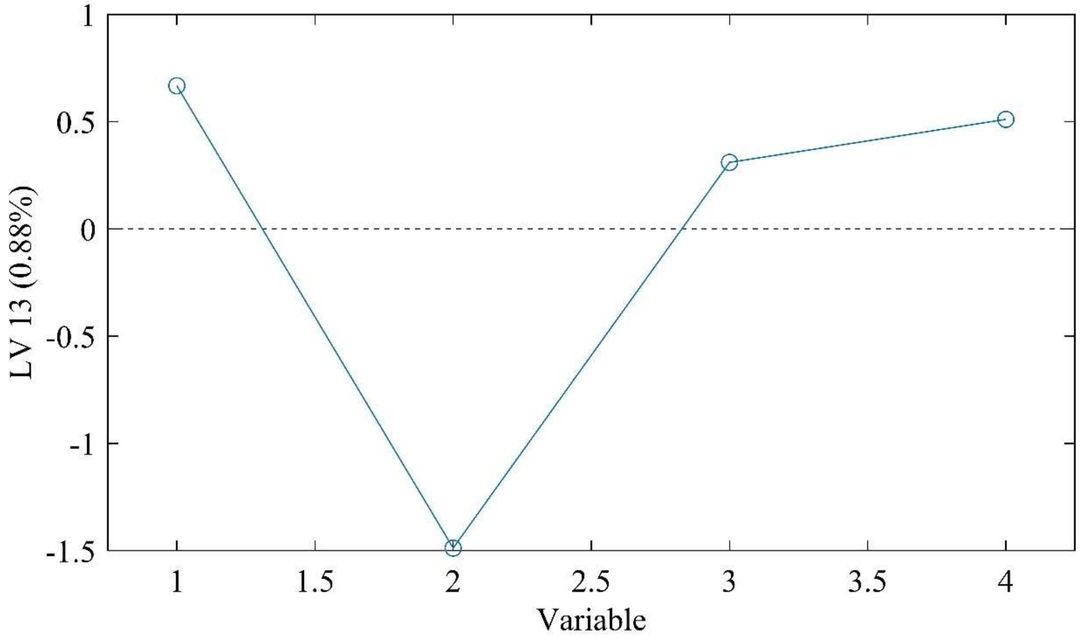

Figure B13: Y Loading vector 13 of Figure 6

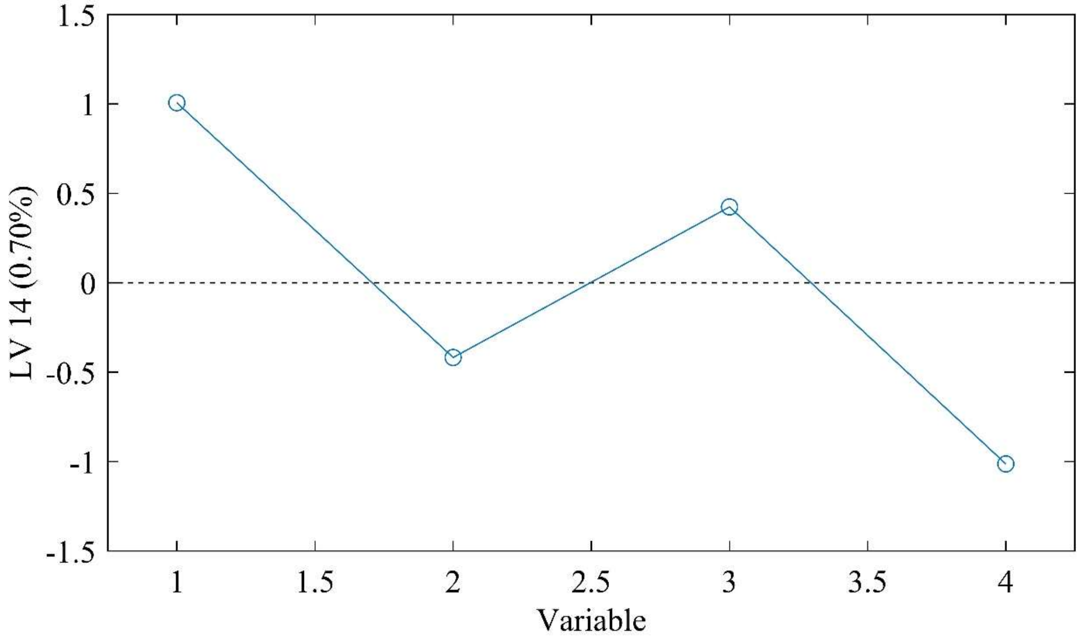

Figure B14: Y Loading vector 14 of Figure 6

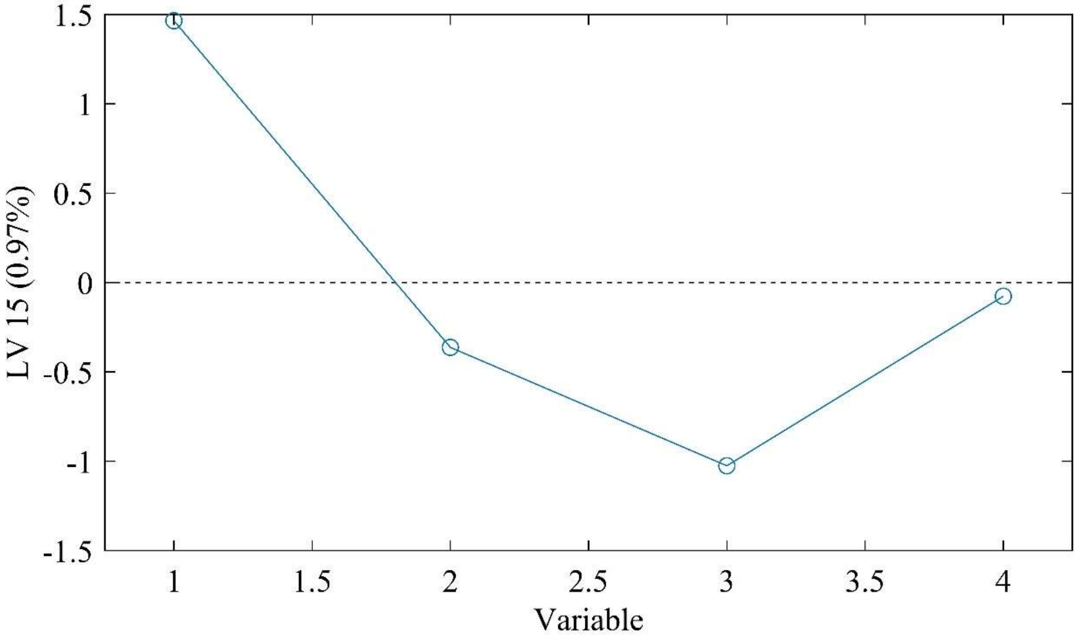

Figure B15: Y Loading vector 15 of Figure 6

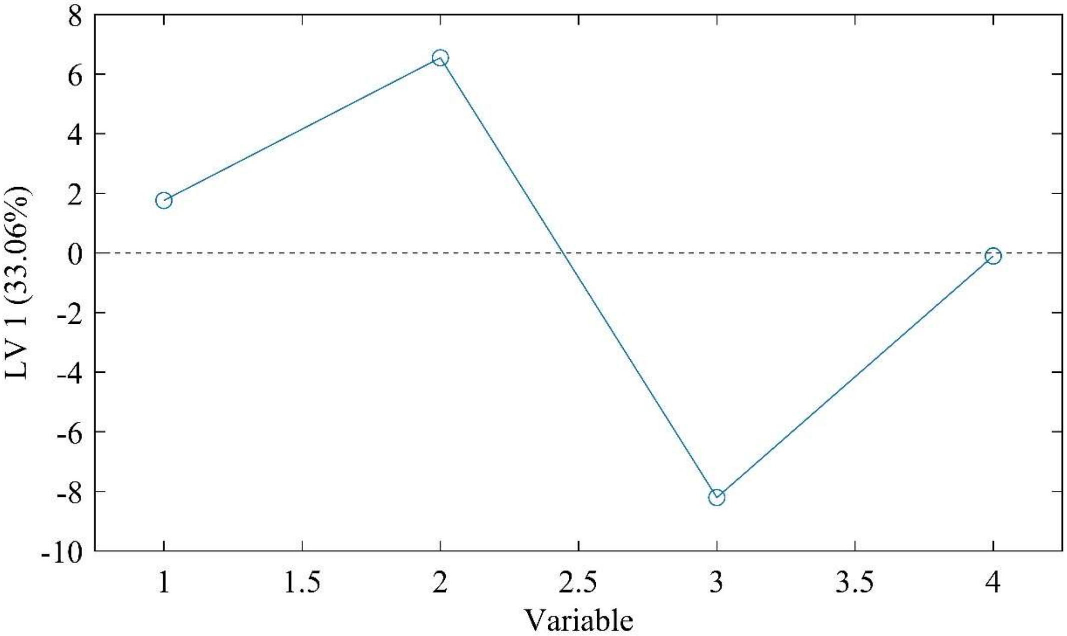

Figure B16: Y Loading vector 1 of Figure 8

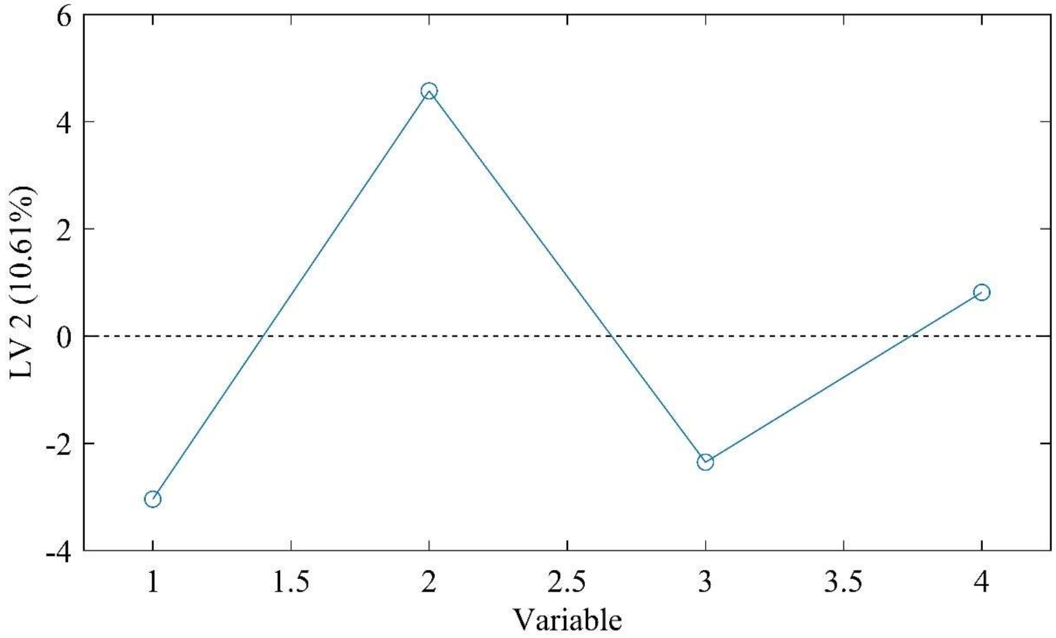

Figure B17: Y Loading vector 2 of Figure 8

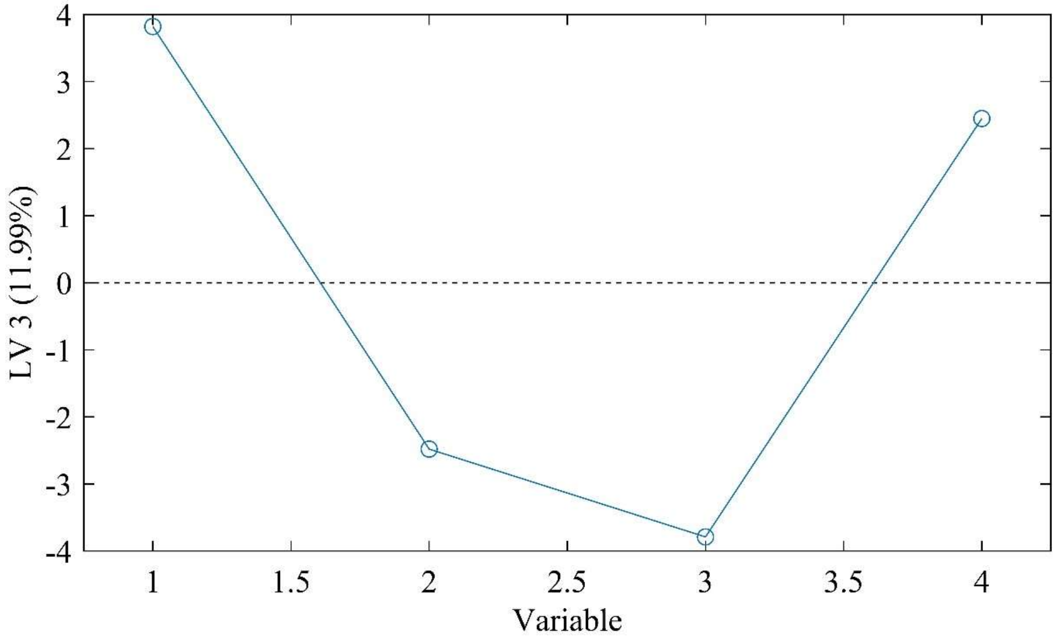

Figure B18: Y Loading vector 3 of Figure 8

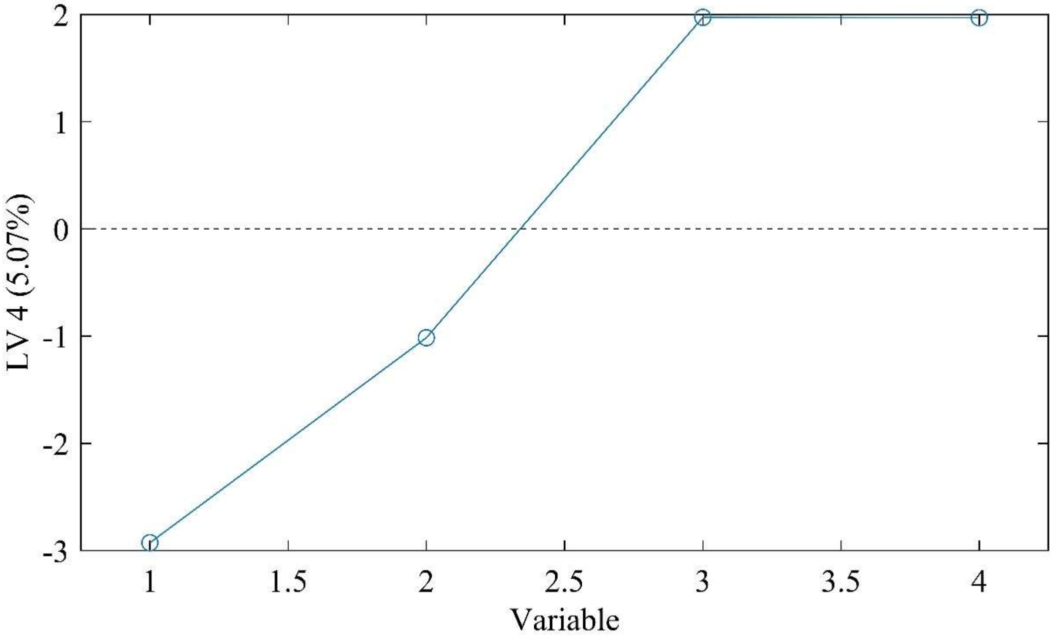

Figure B19: Y Loading vector 4 of Figure 8

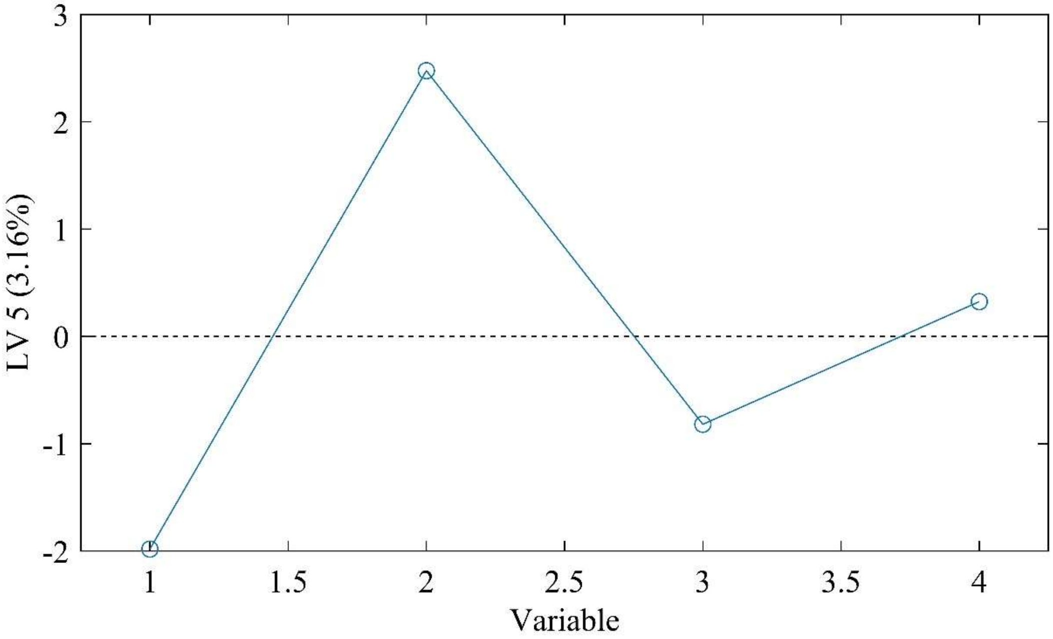

Figure B20: Y Loading vector 5 of Figure 8

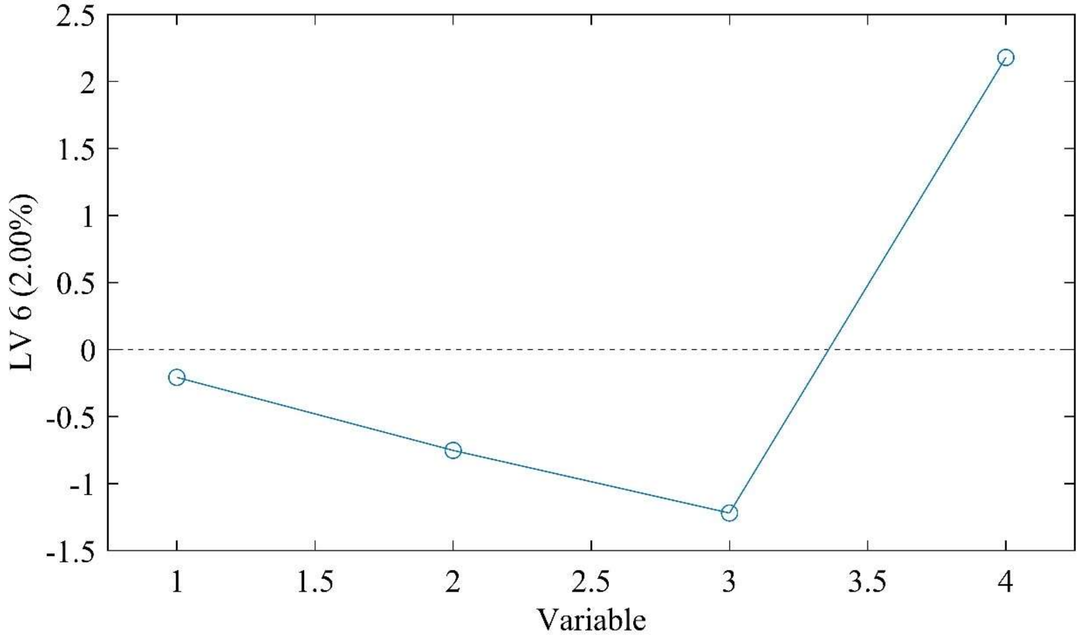

Figure B21: Y Loading vector 6 of Figure 8

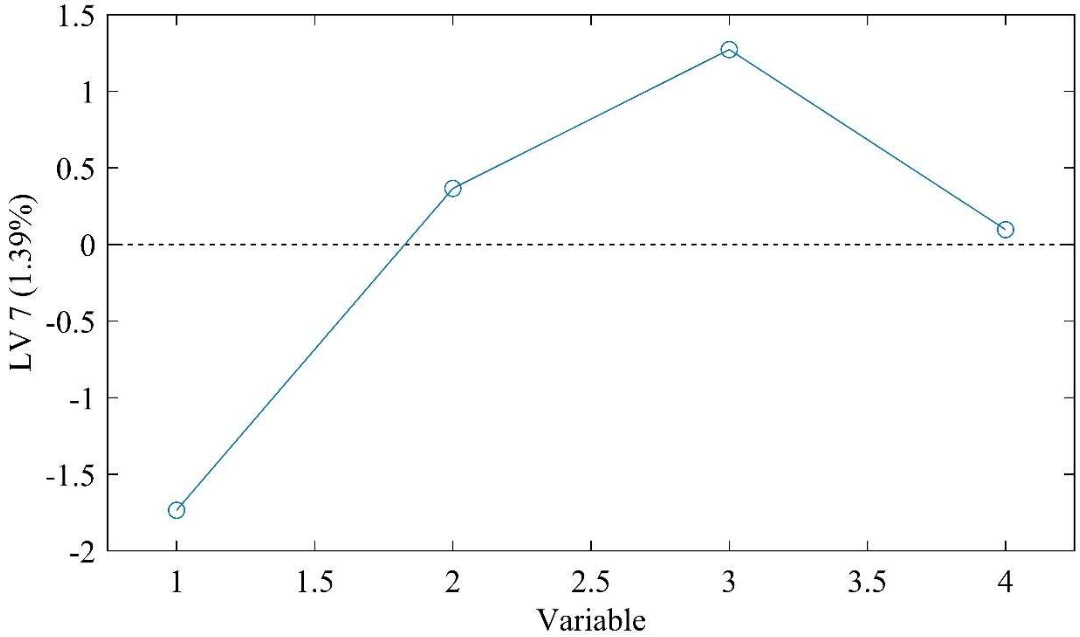

Figure B22: Y Loading vector 7 of Figure 8

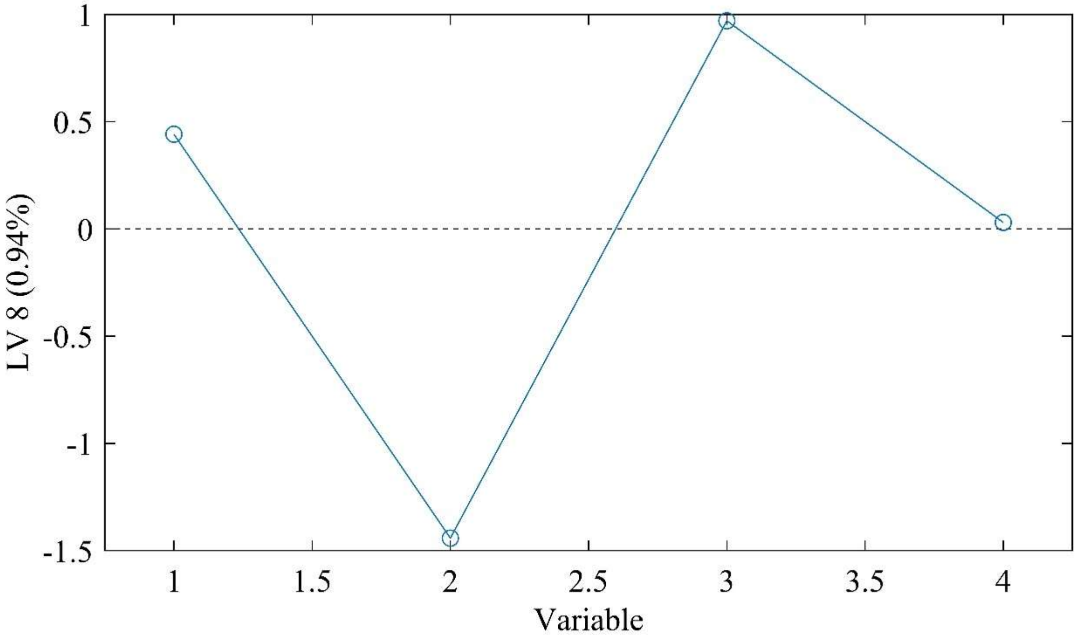

Figure B23: Y Loading vector 8 of Figure 8

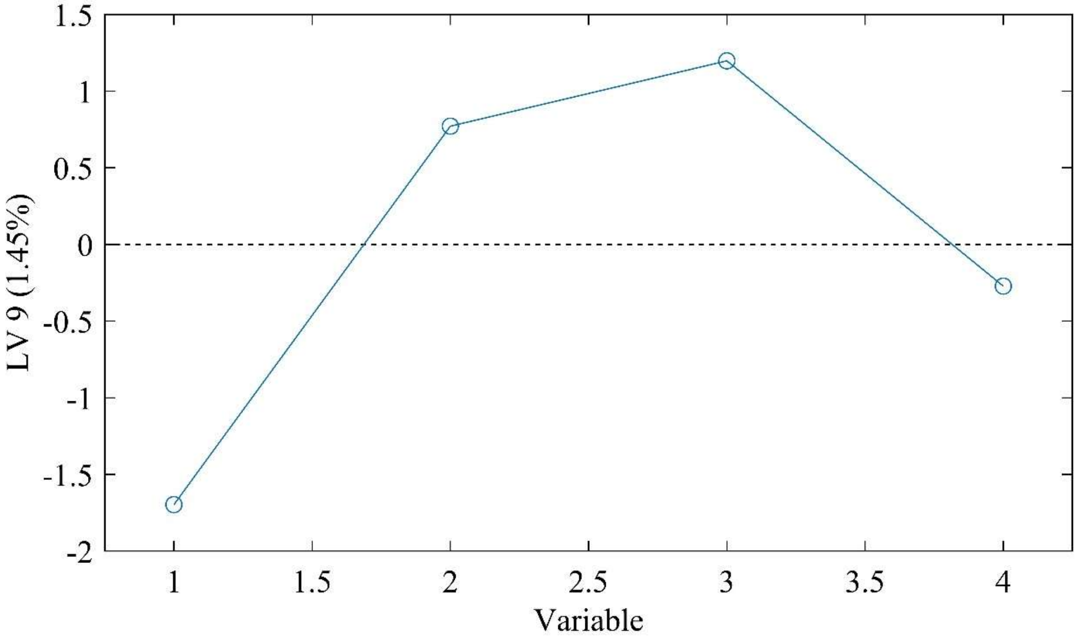

Figure B24: Y Loading vector 9 of Figure 8

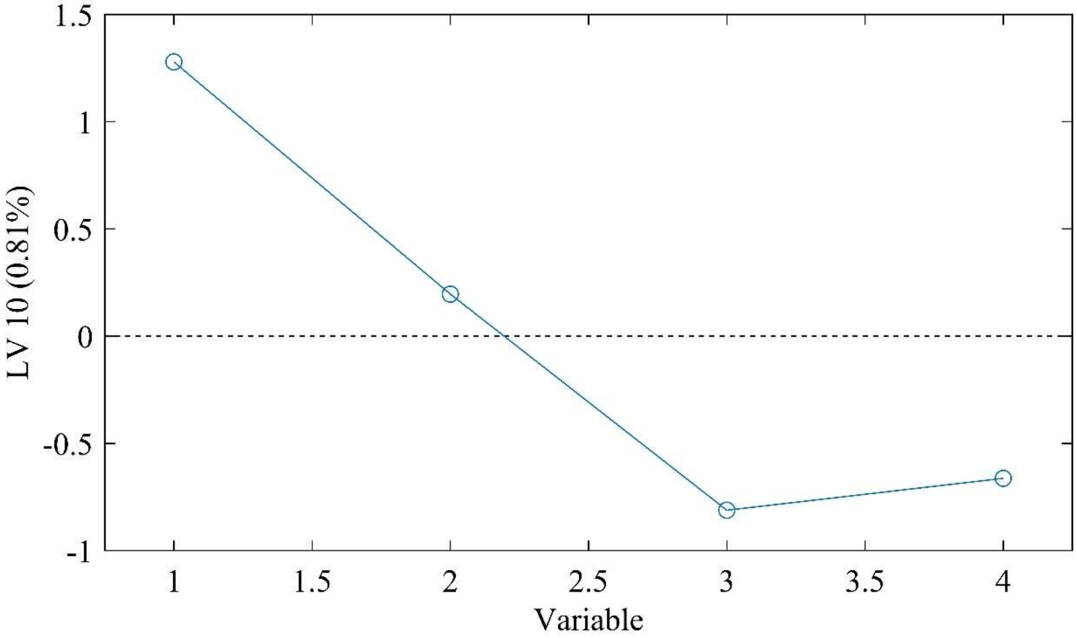

Figure B25: Y Loading vector 10 of Figure 8

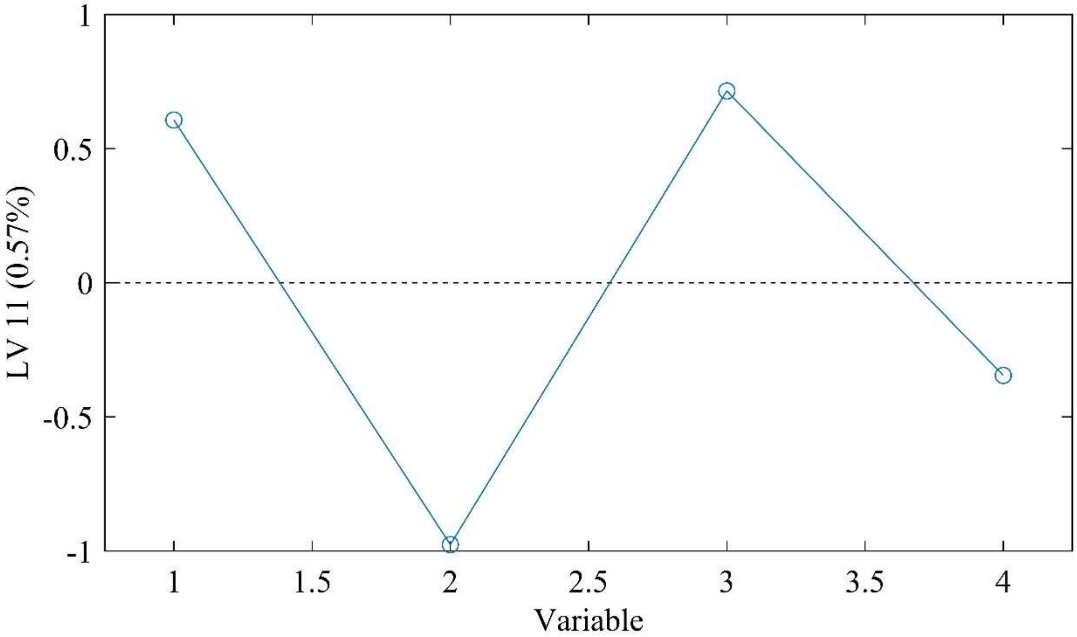

Figure B26: Y Loading vector 11 of Figure 8

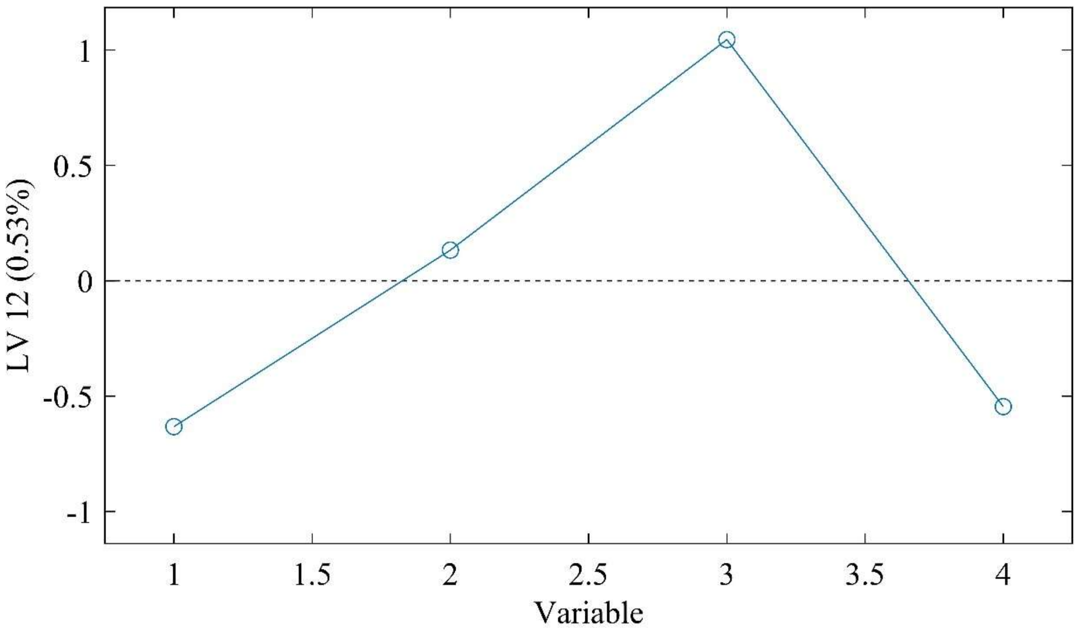

Figure B27: Y Loading vector 12 of Figure 8

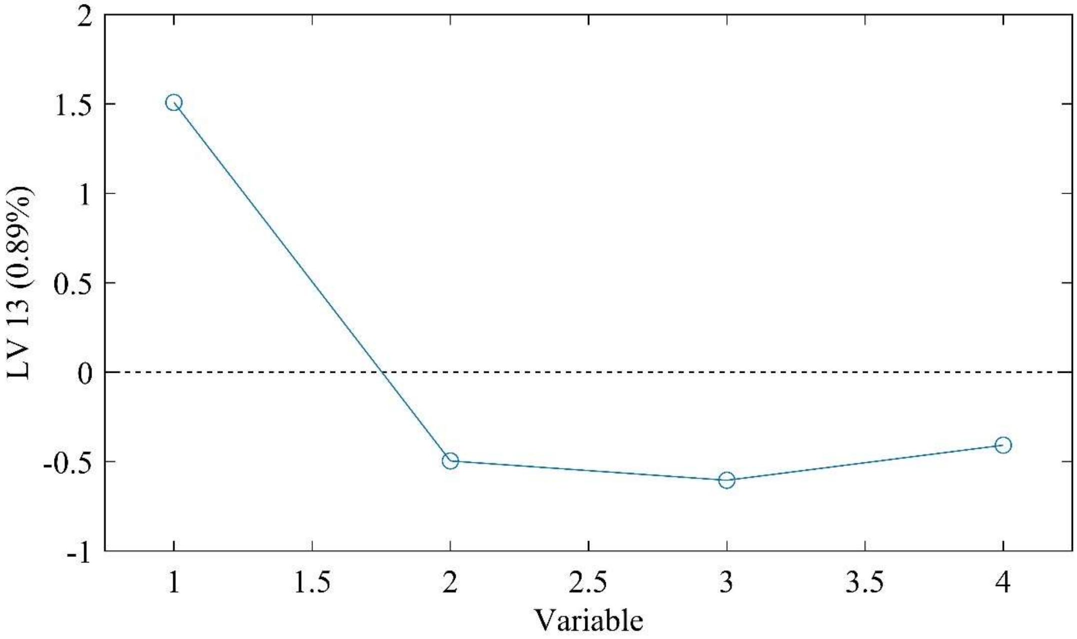

Figure B28: Y Loading vector 13 of Figure 8

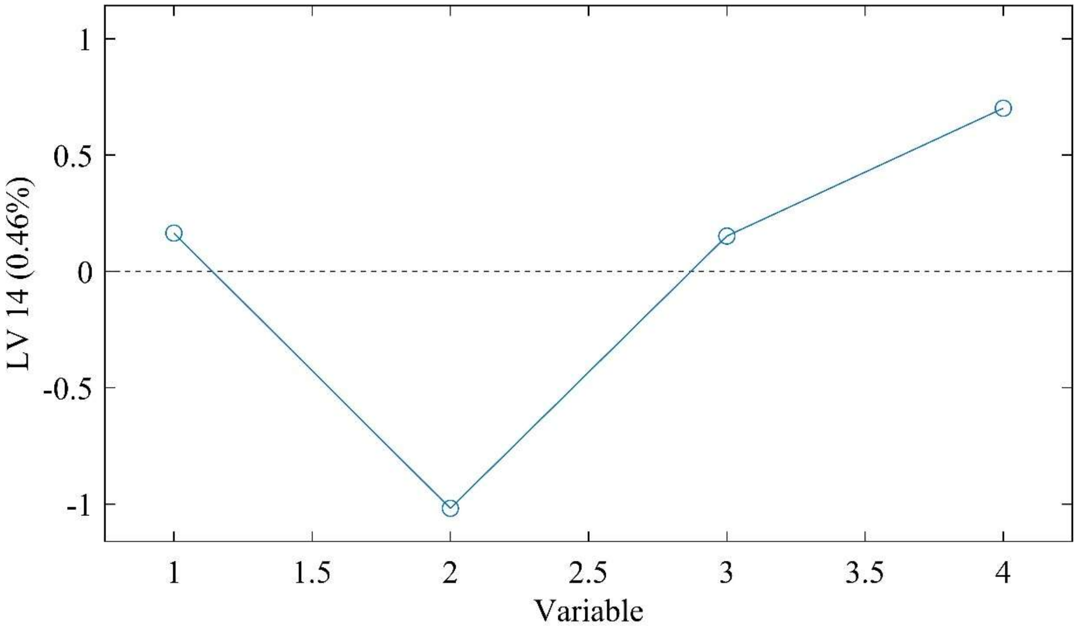

Figure B29: Y Loading vector 14 of Figure 8

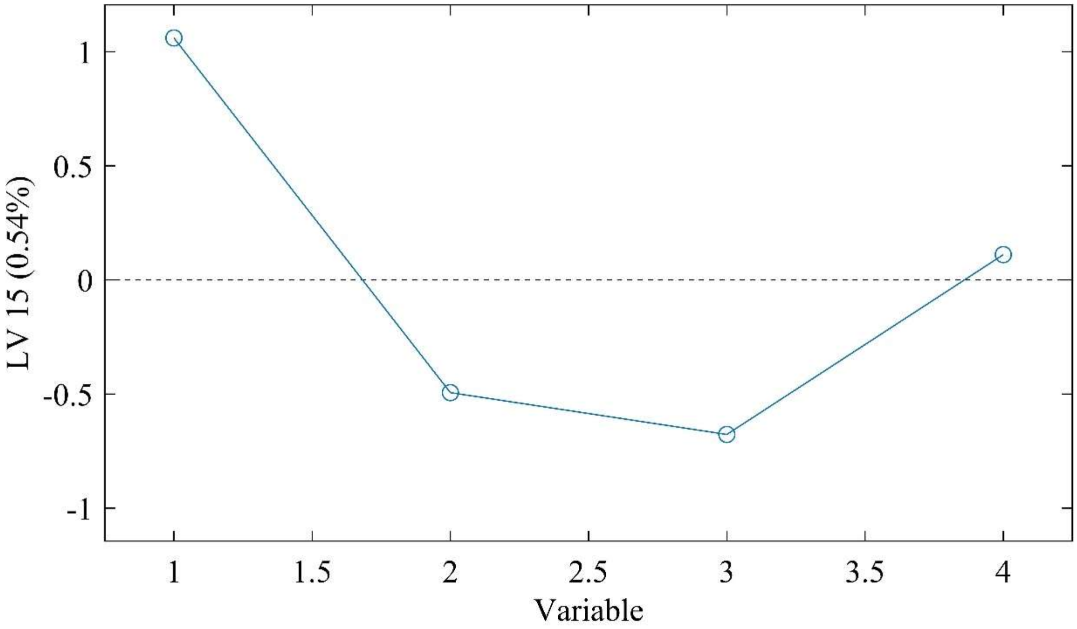

Figure B30: Y Loading vector 15 of Figure 8

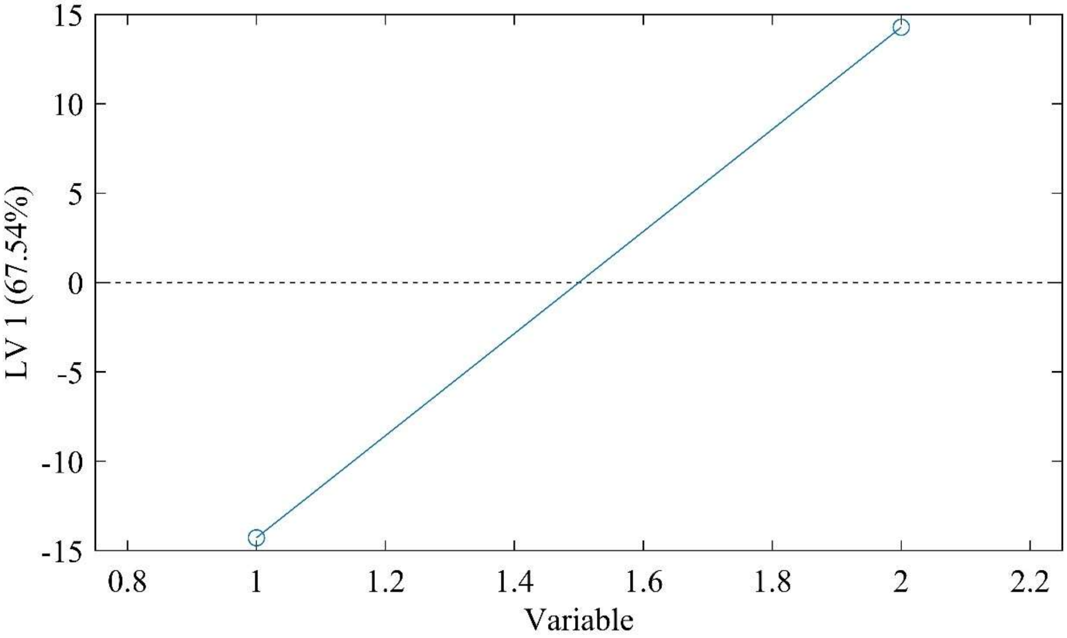

Figure B31: Y Loading vector 1 of Figure 11a

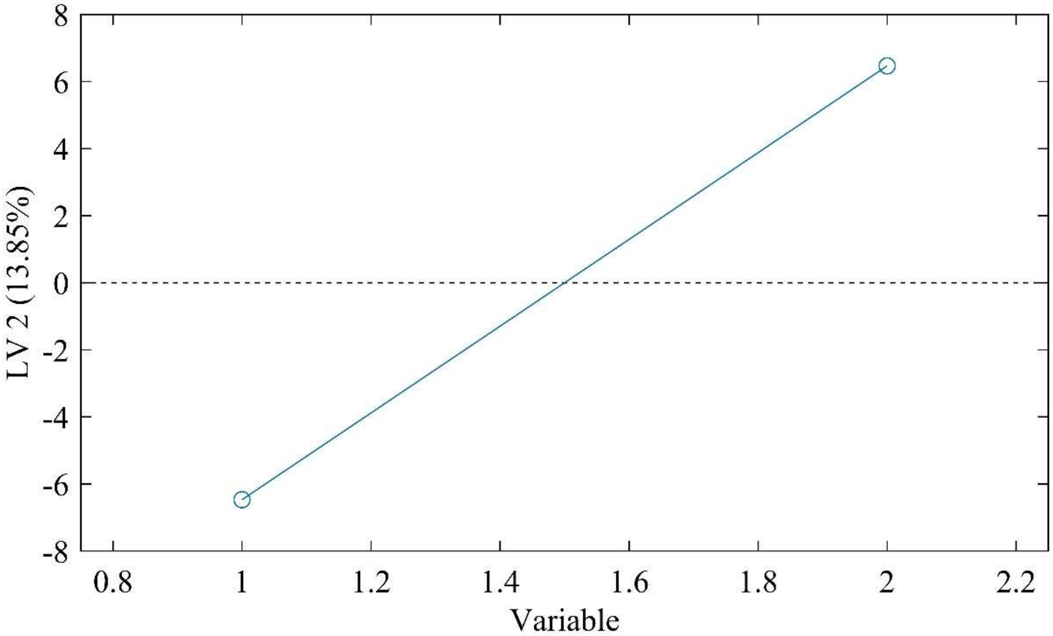

Figure B32: Y Loading vector 2 of Figure 11a

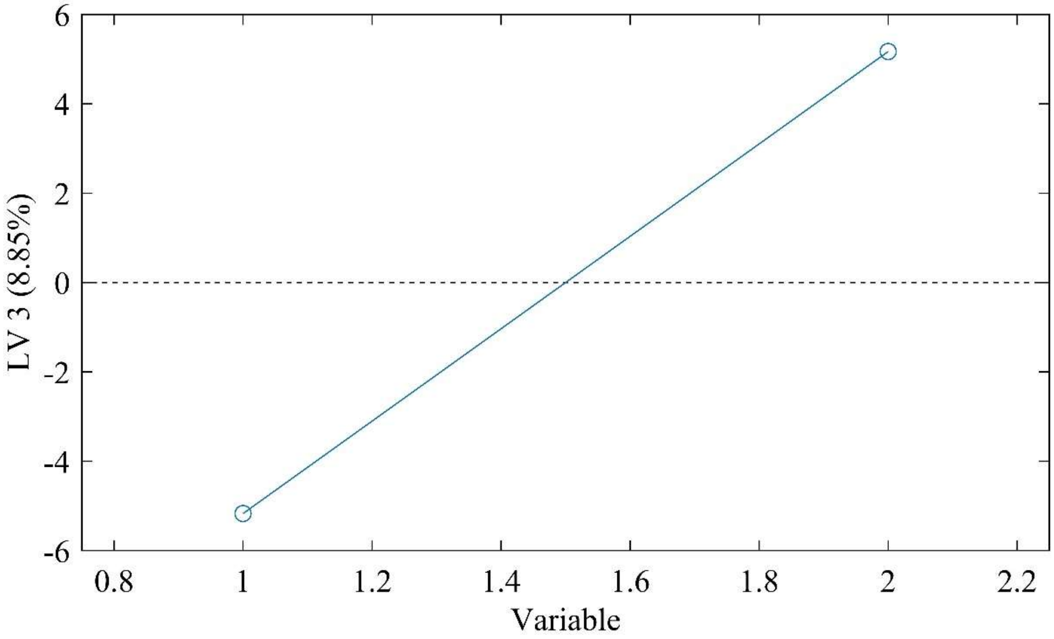

Figure B33: Y Loading vector 3 of Figure 11a

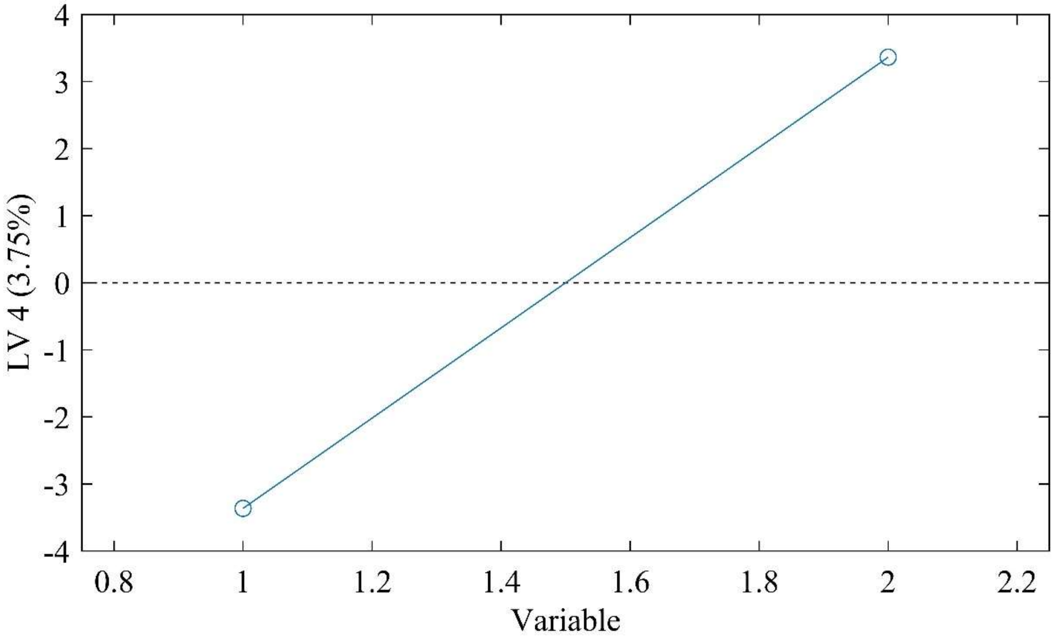

Figure B34: Y Loading vector 4 of Figure 11a

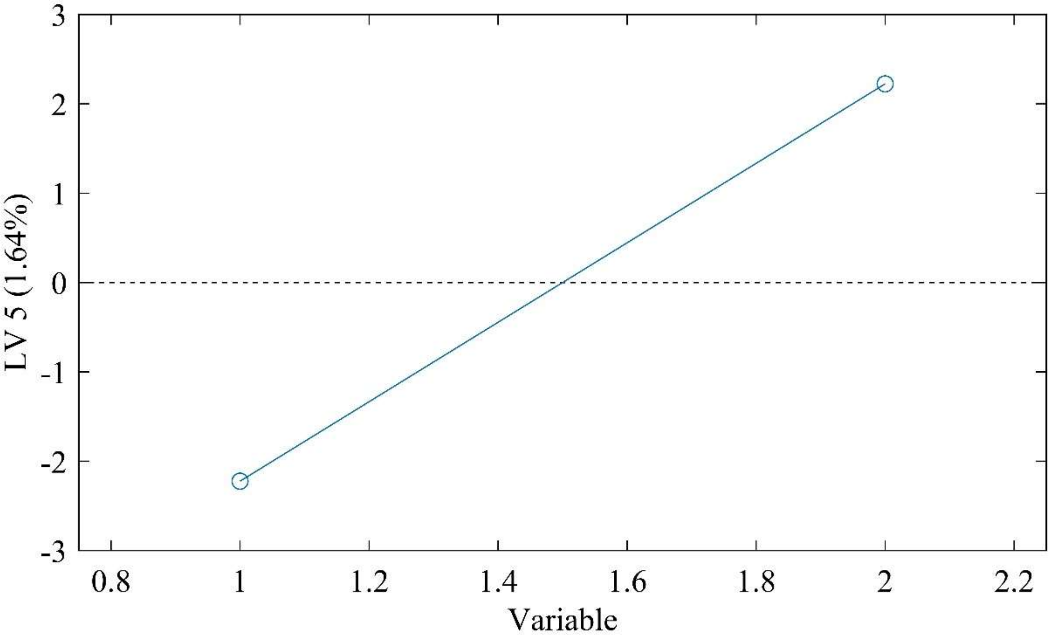

Figure B35: Y Loading vector 5 of Figure 11a

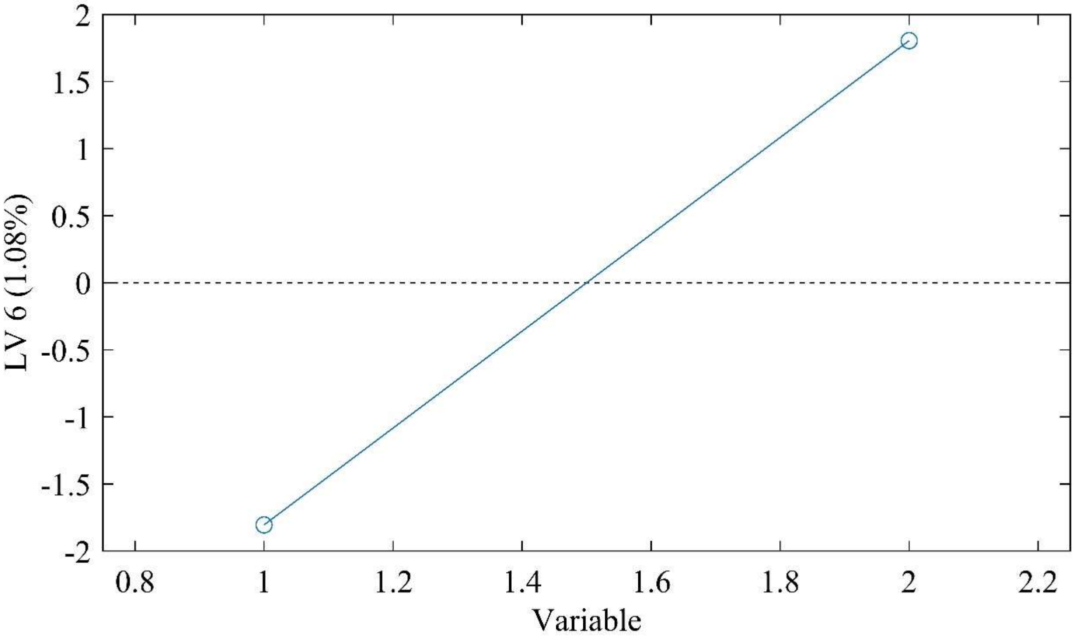

Figure B36: Y Loading vector 6 of Figure 11a

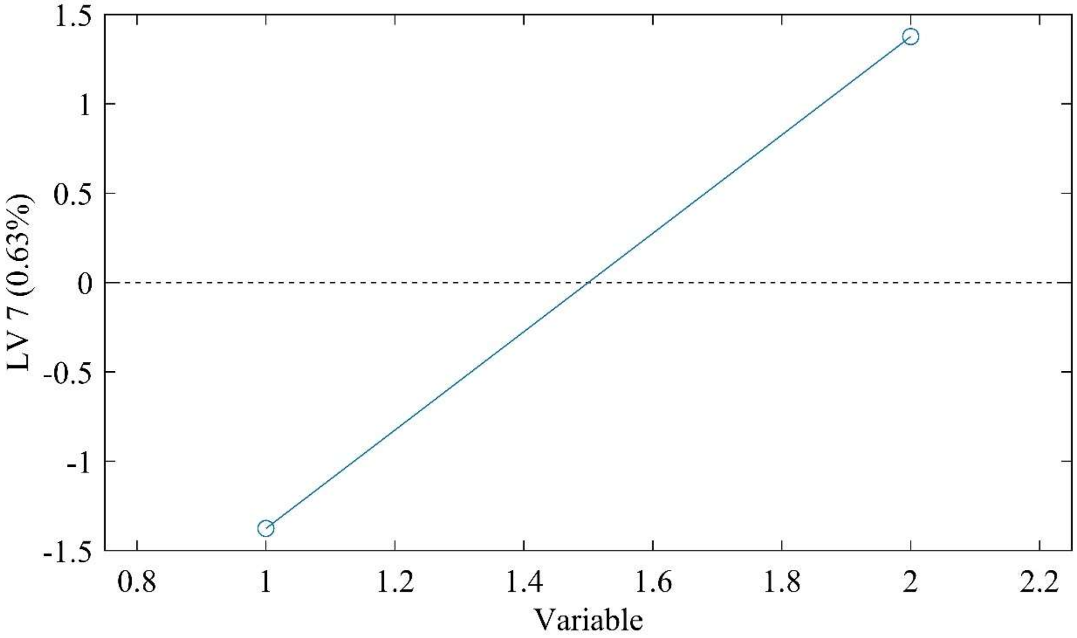

Figure B37: Y Loading vector 7 of Figure 11a

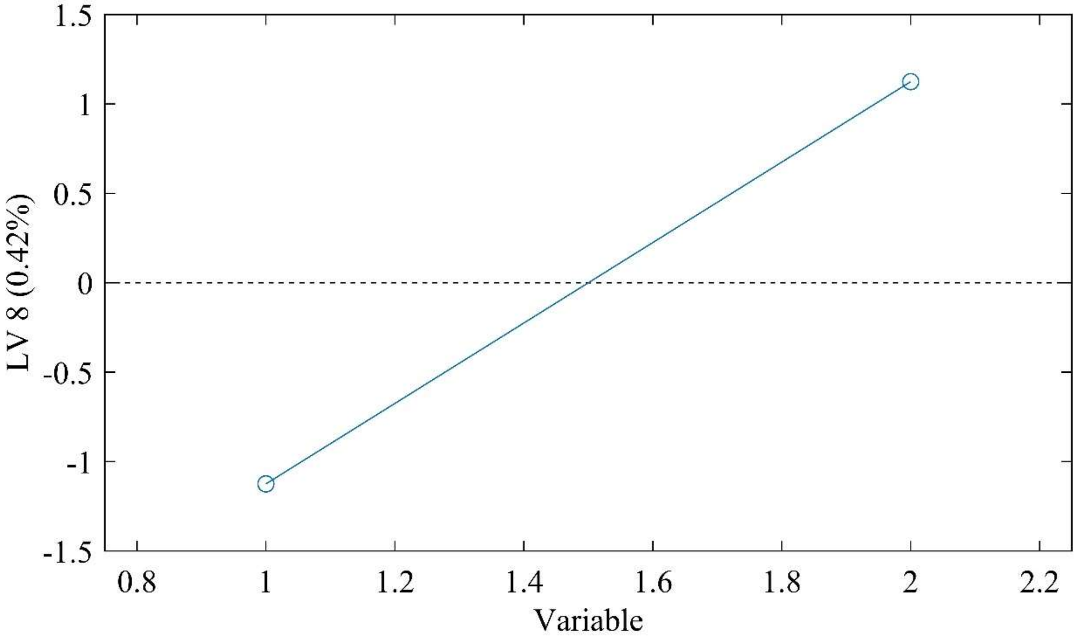

Figure B38: Y Loading vector 8 of Figure 11a

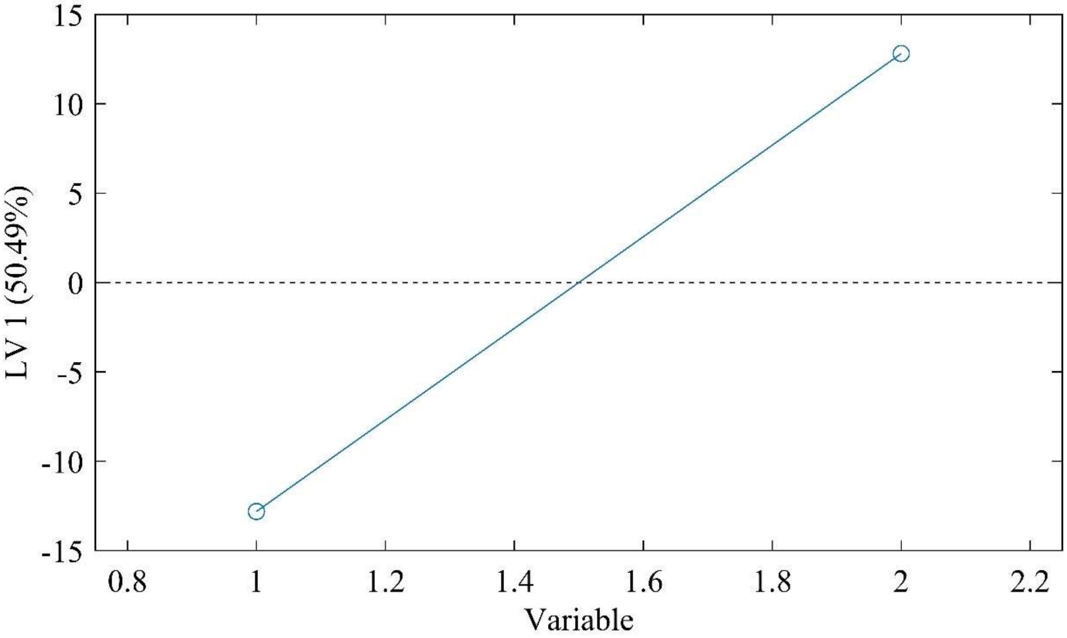

Figure B39: Y Loading vector 1 of Figure 12a

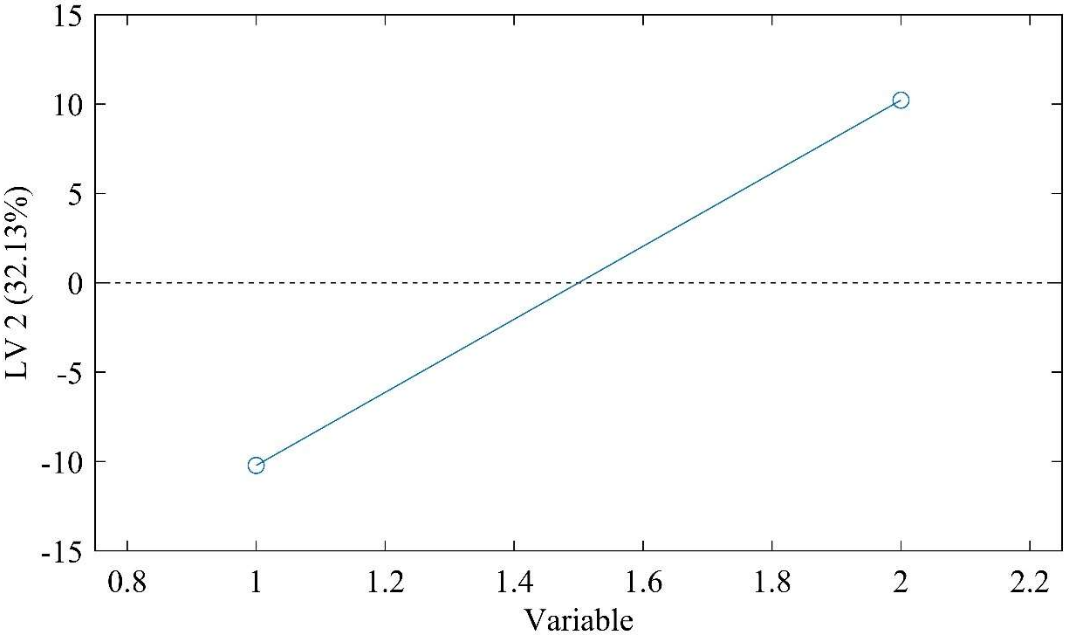

Figure B40: Y Loading vector 2 of Figure 12a

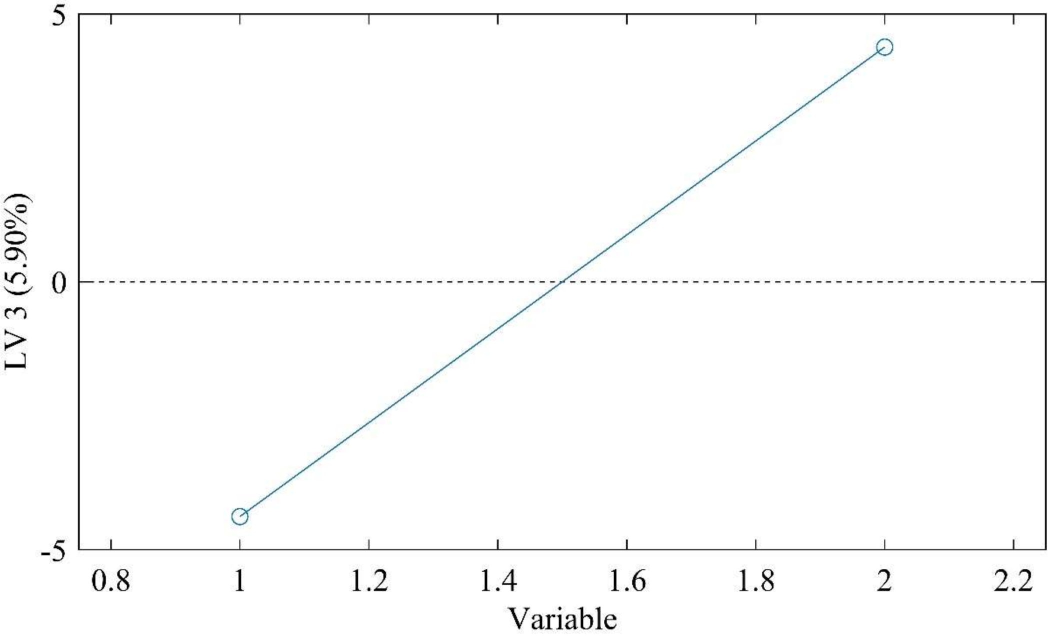

Figure B41: Y Loading vector 3 of Figure 12a

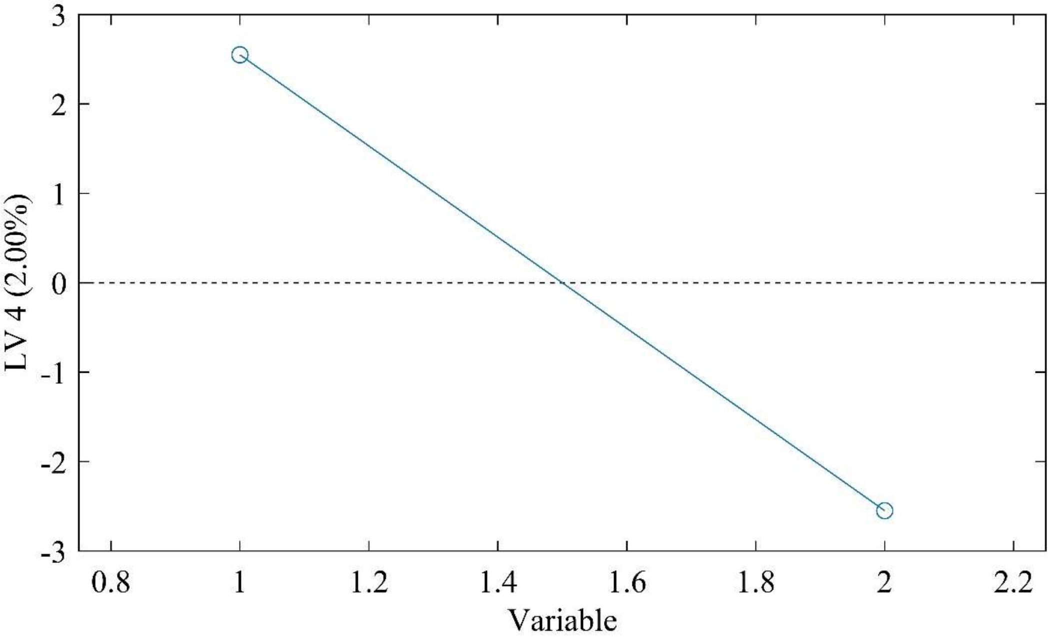

Figure B42: Y Loading vector 4 of Figure 12a

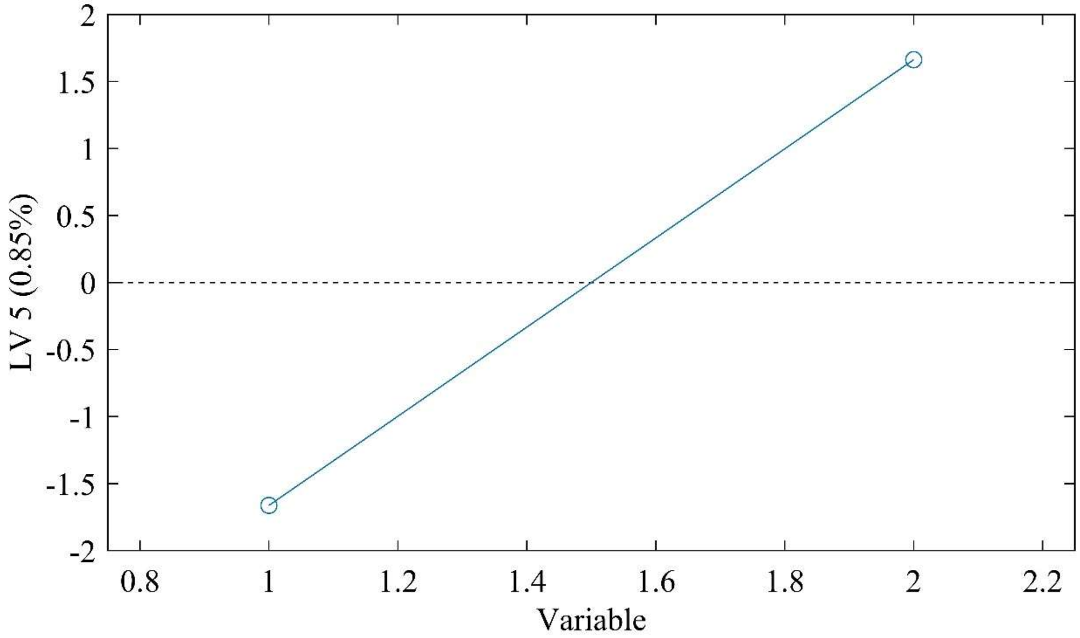

Figure B43: Y Loading vector 5 of Figure 12a

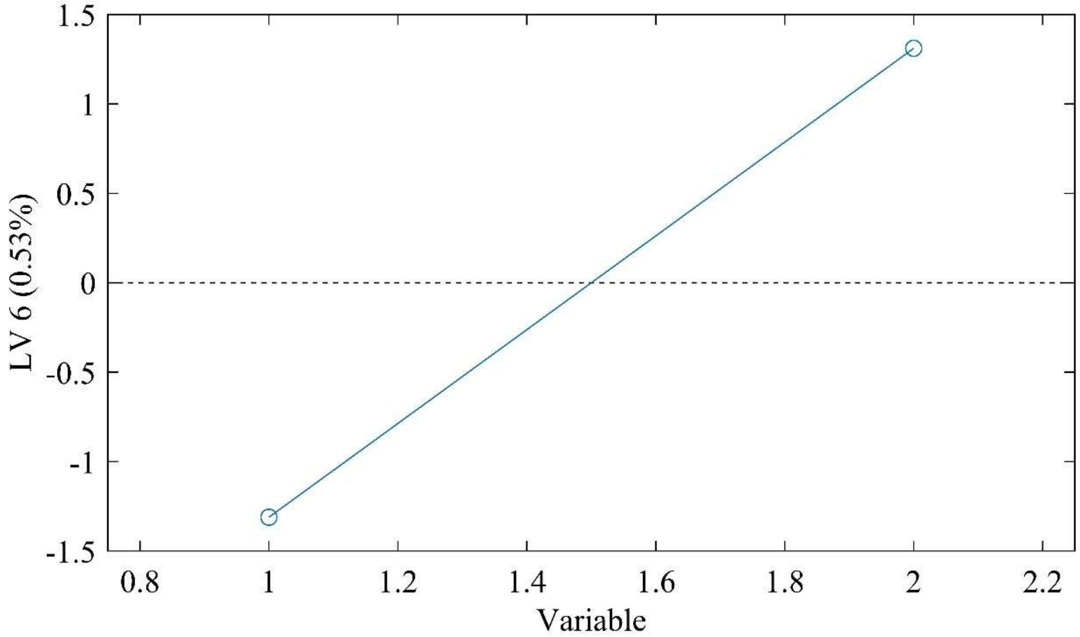

Figure B44: Y Loading vector 6 of Figure 12a

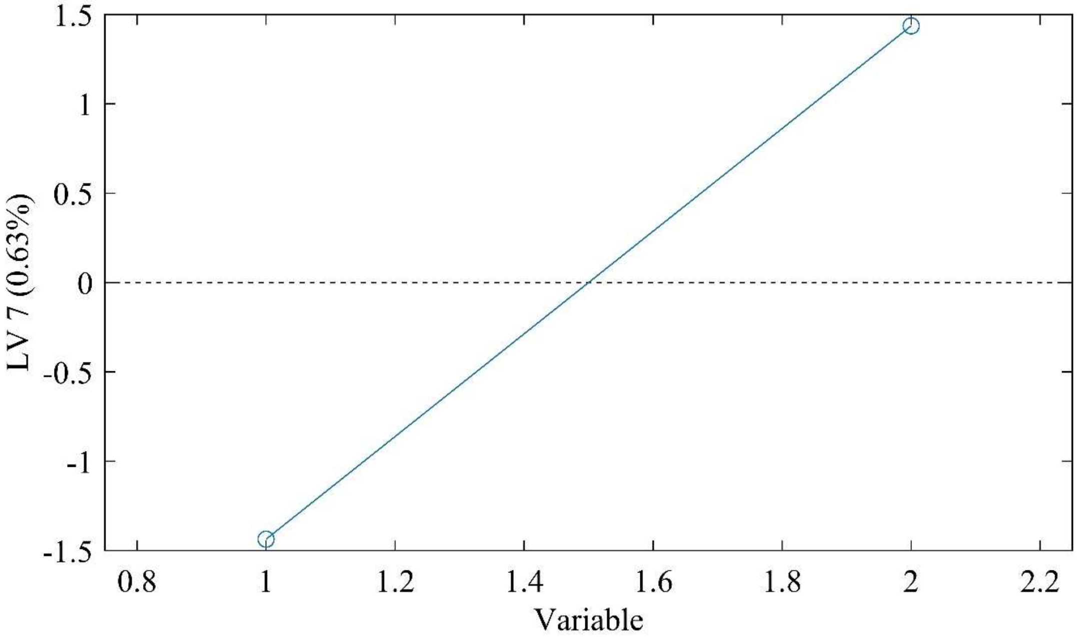

Figure B45: Y Loading vector 7 of Figure 12a

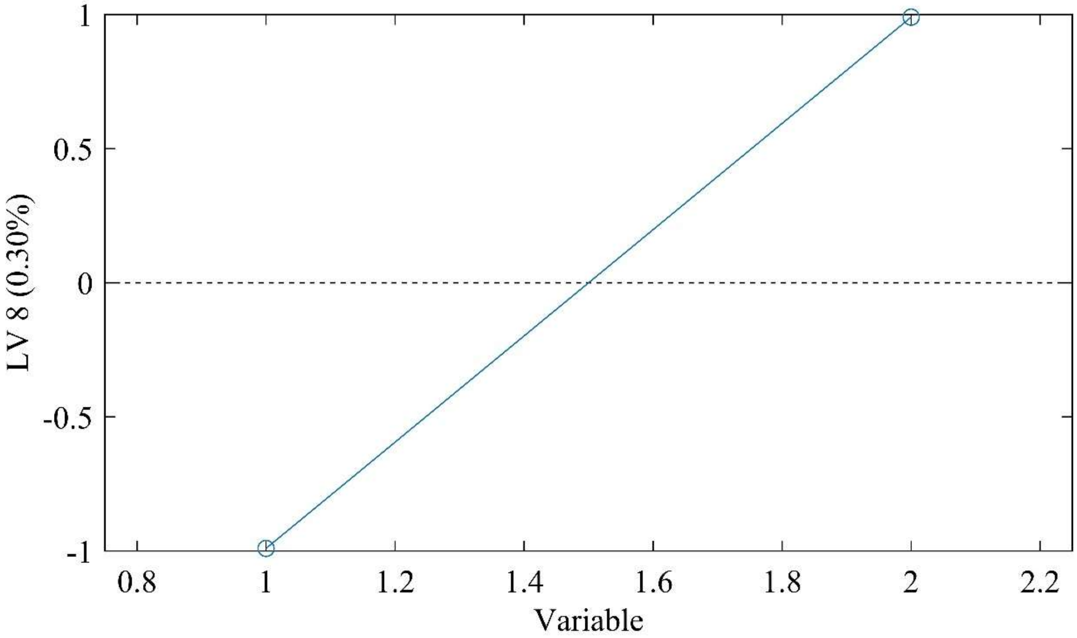

Figure B46: Y Loading vector 8 of Figure 12a

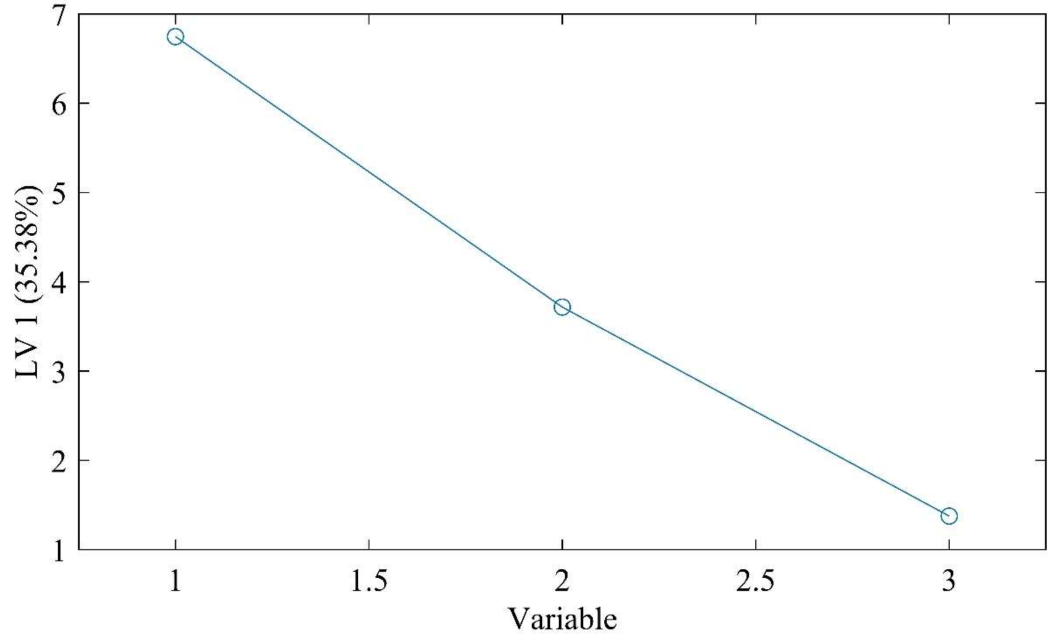

Figure B47: Y Loading vector 1 of Figure 18

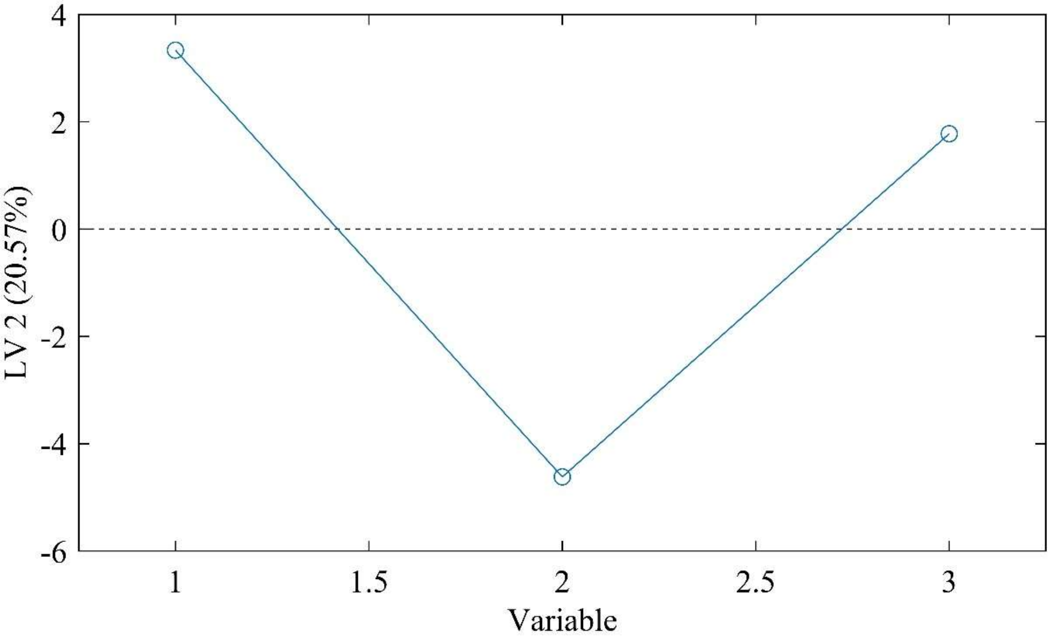

Figure B48: Y Loading vector 2 of Figure 18

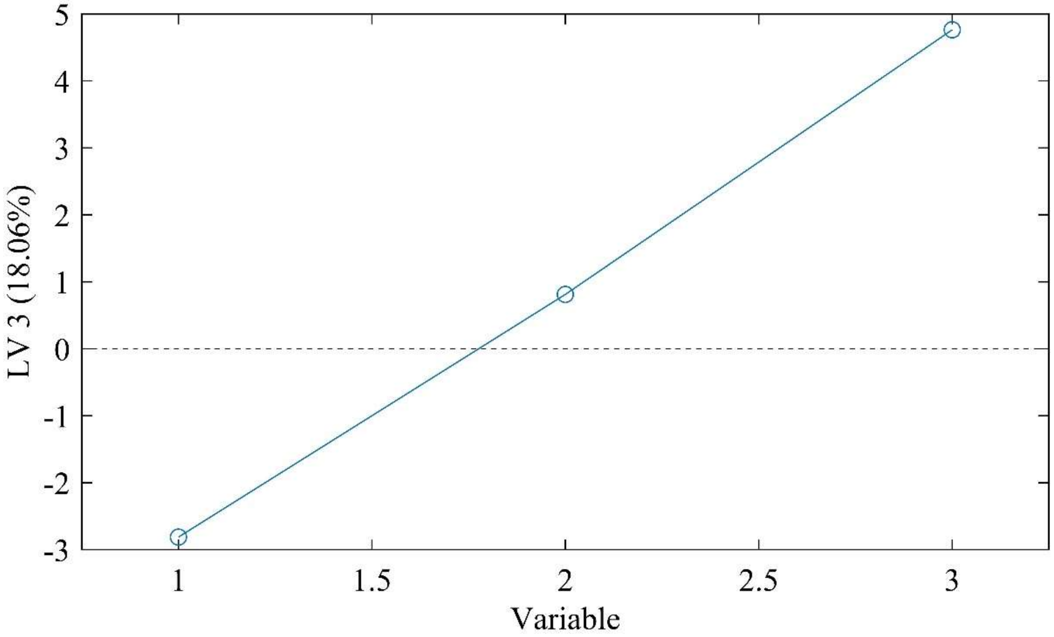

Figure B49: Y Loading vector 3 of Figure 18

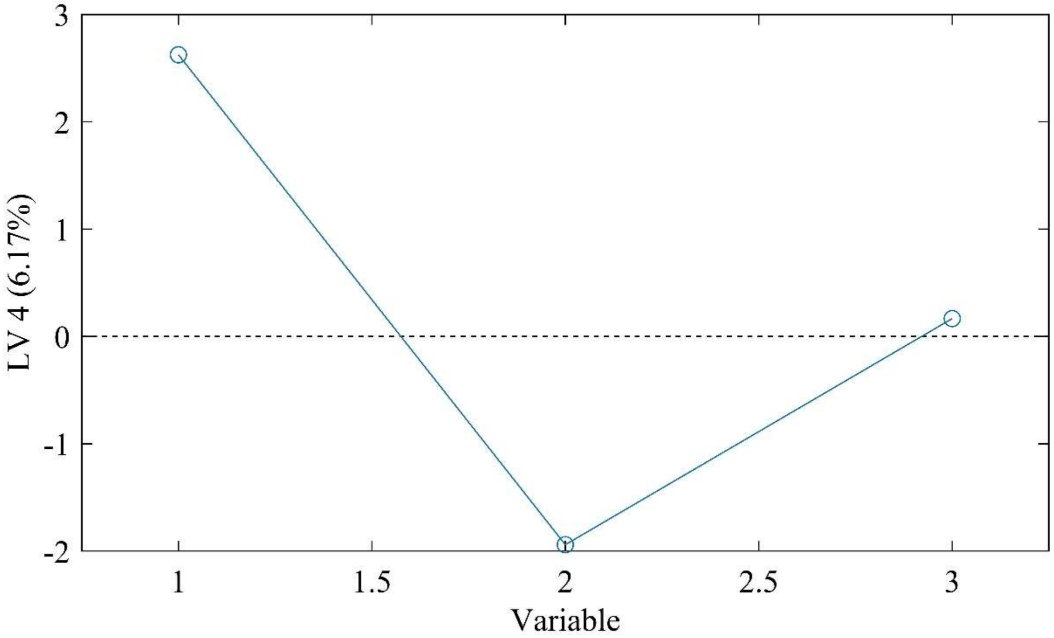

Figure B50: Y Loading vector 4 of Figure 18

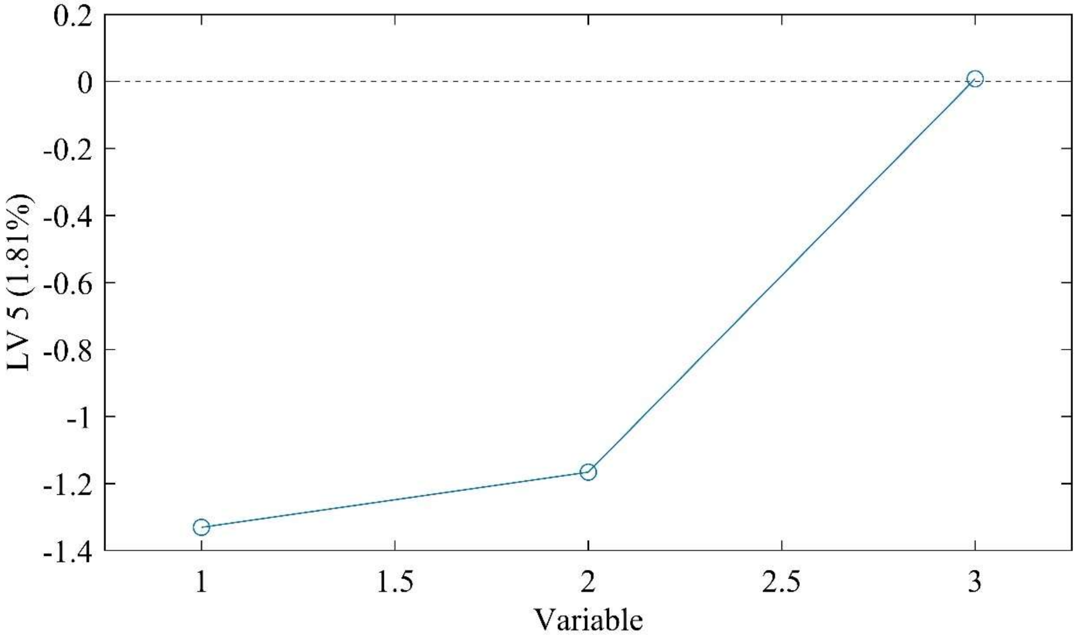

Figure B51: Y Loading vector 5 of Figure 18

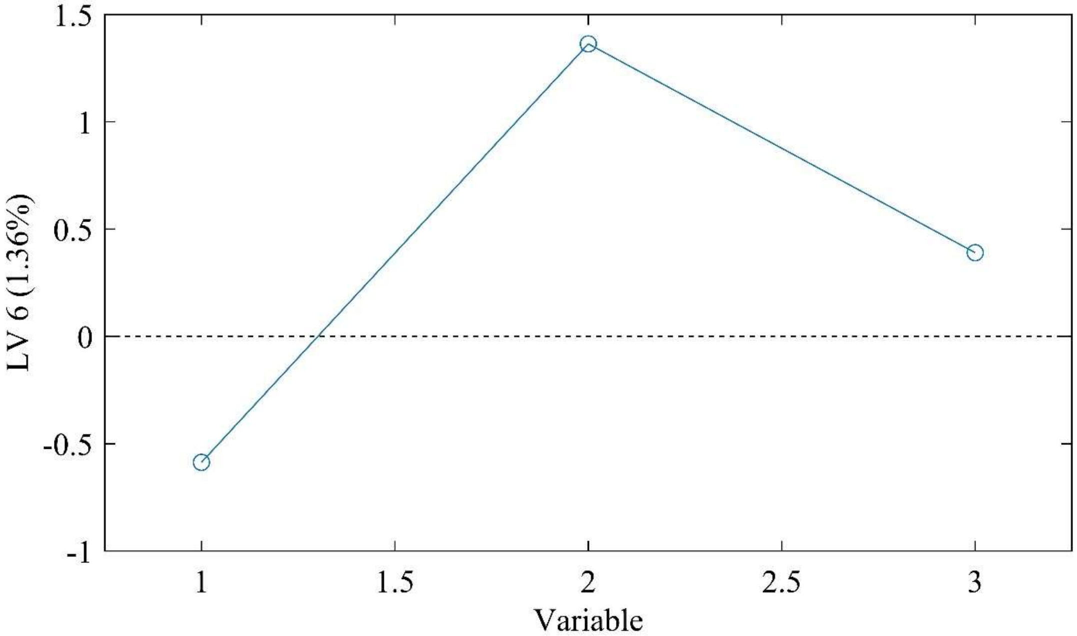

Figure B52: Y Loading vector 6 of Figure 18

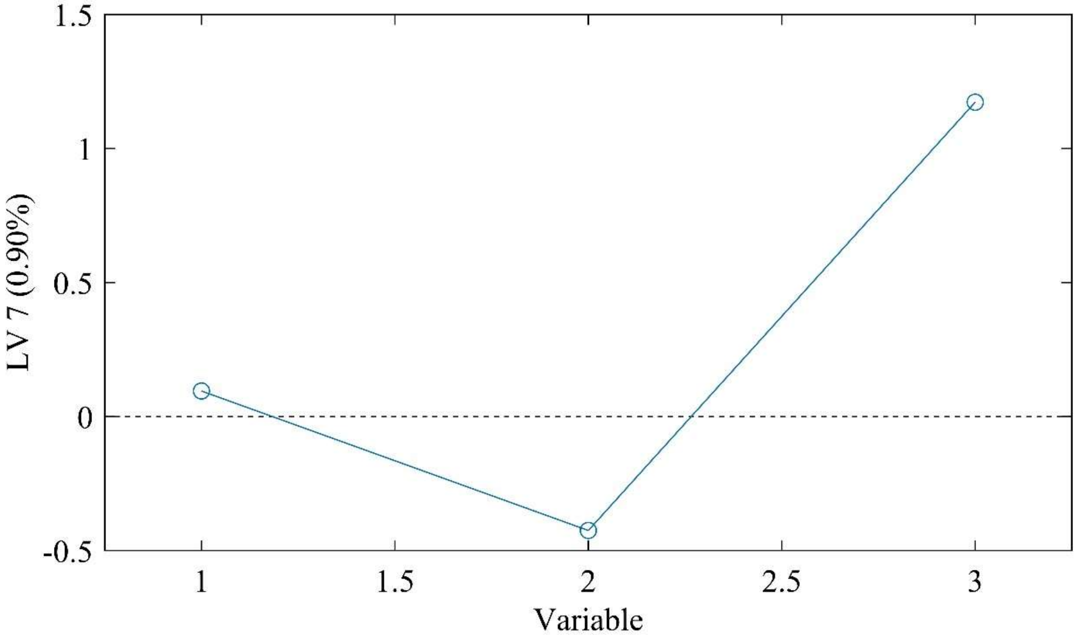

Figure B53: Y Loading vector 7 of Figure 18

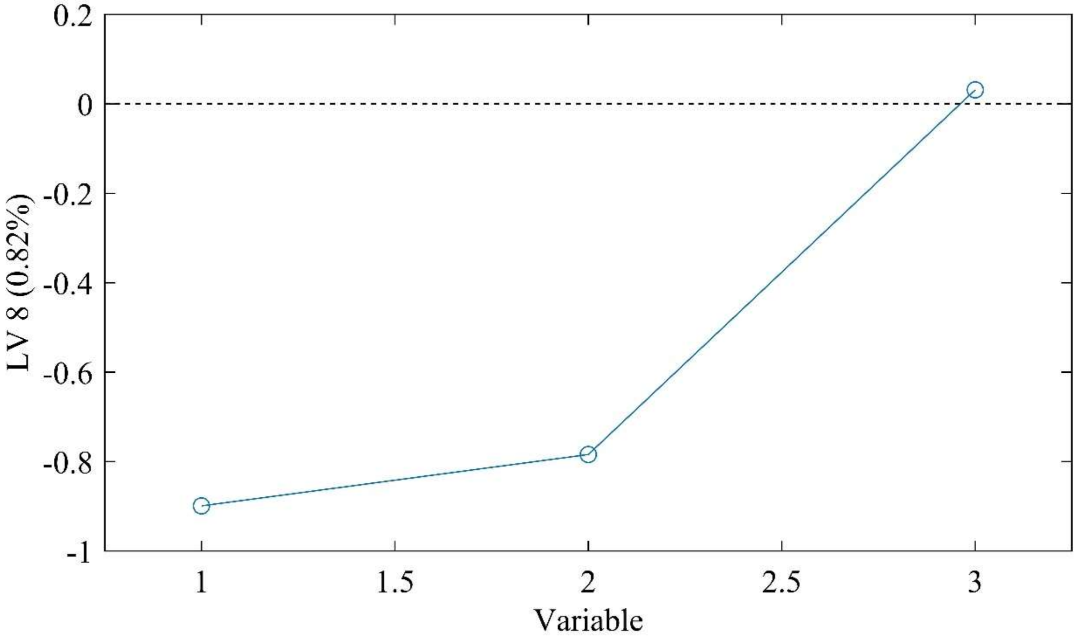

Figure B54: Y Loading vector 8 of Figure 18

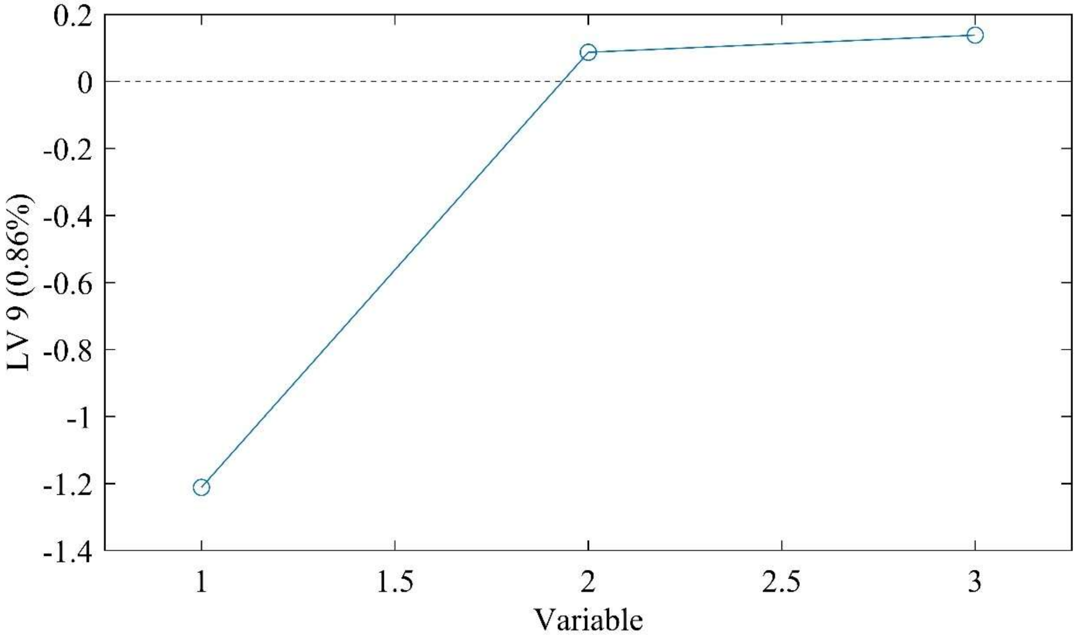

Figure B55: Y Loading vector 9 of Figure 18

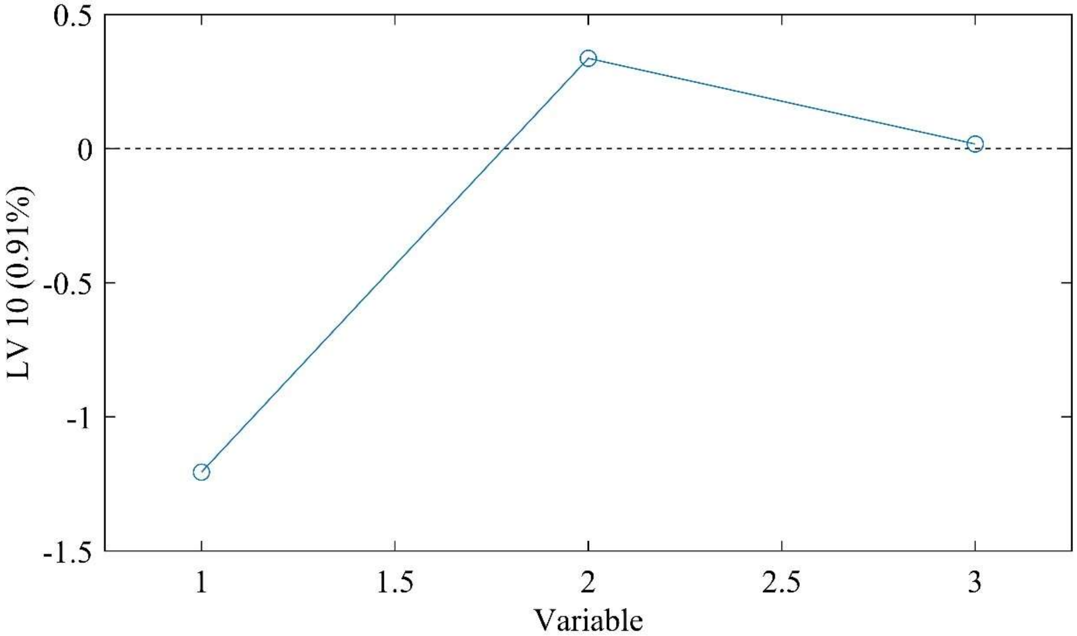

Figure B56: Y Loading vector 10 of Figure 18

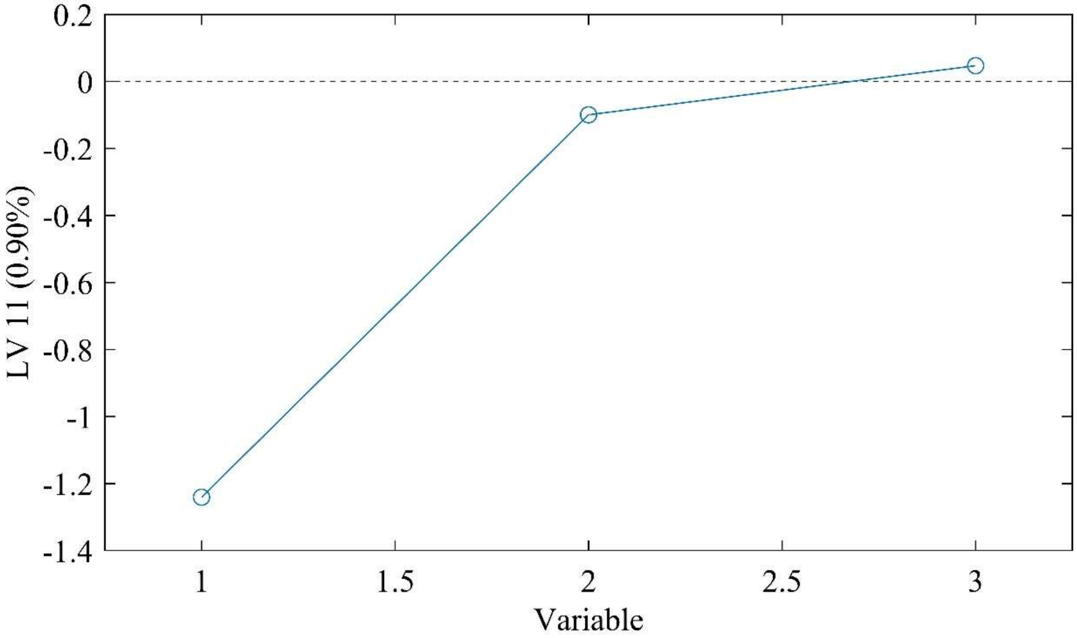

Figure B57: Y Loading vector 11 of Figure 18

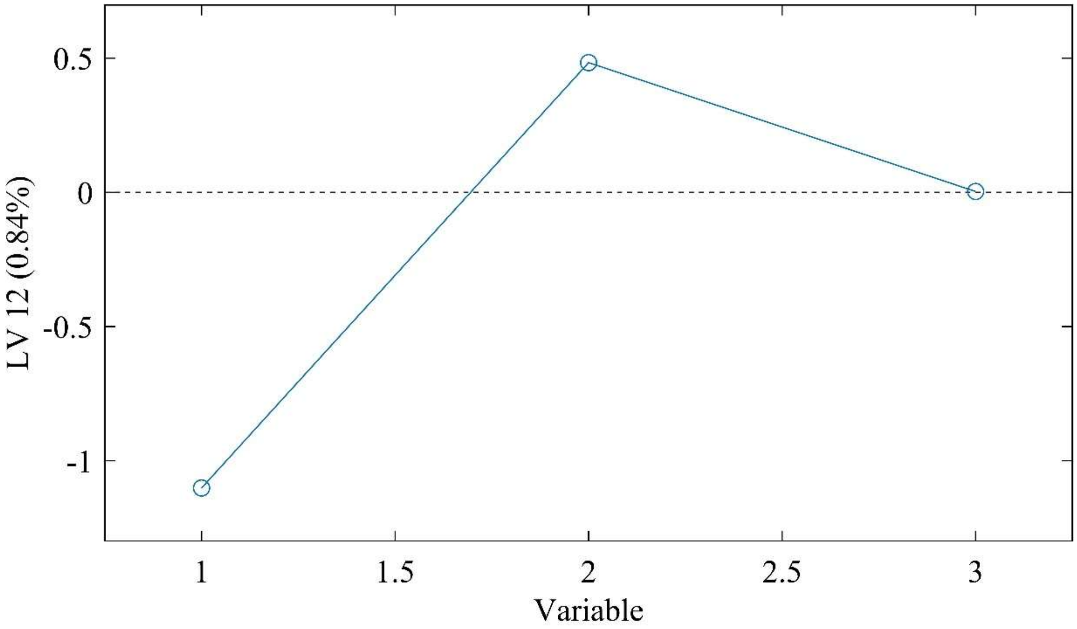

Figure B58: Y Loading vector 12 of Figure 18

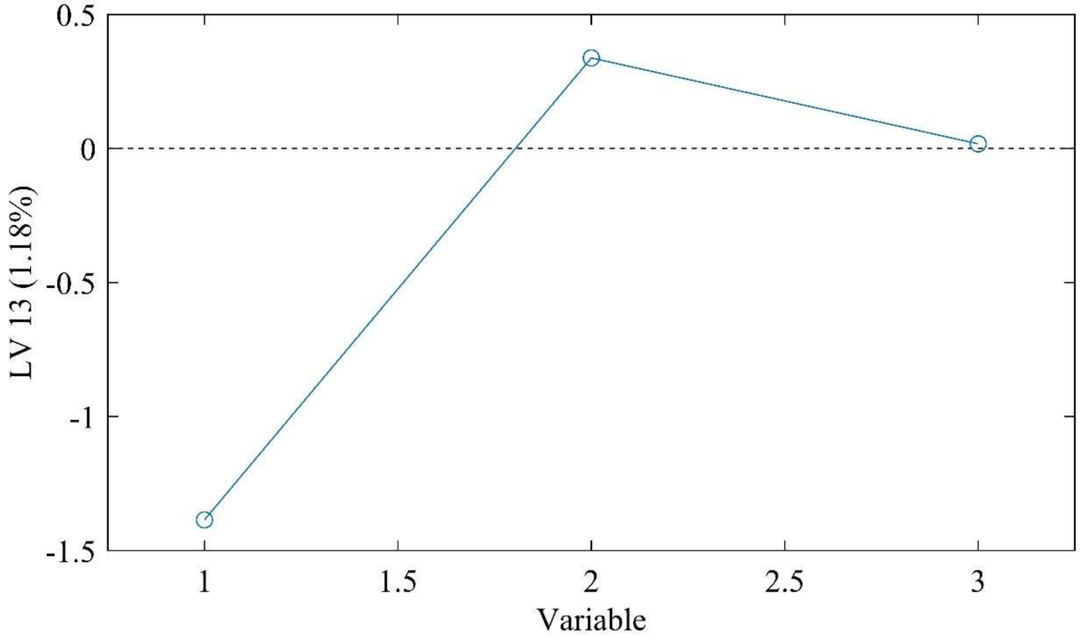

Figure B59: Y Loading vector 13 of Figure 18

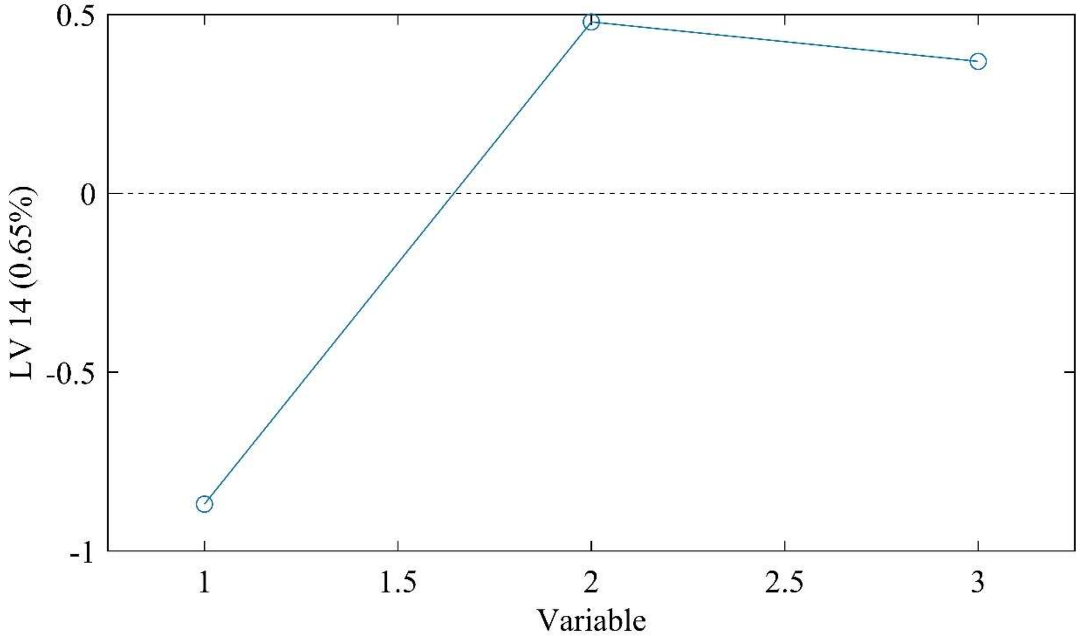

Figure B60: Y Loading vector 14 of Figure 18

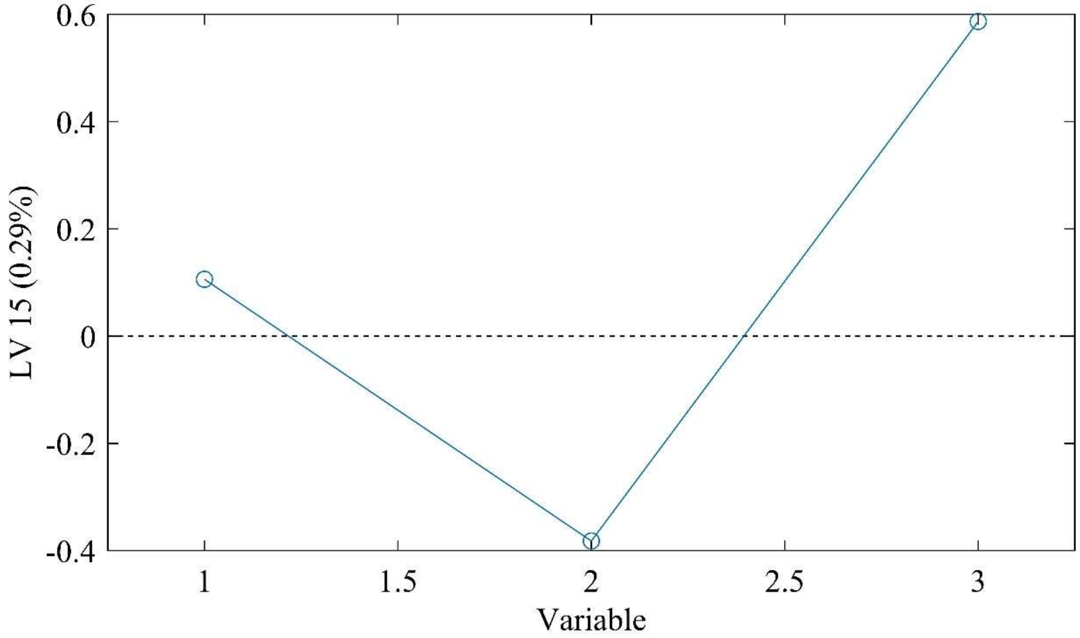

Figure B61: Y Loading vector 15 of Figure 18

APPENDIX C: FENTANYL DERIVATIVE SUB-CLASSES

Table C1: Fentanyl sub-class members listed

Class 1	**Class 2**	**Class 3**
alpha'-methyl Butyryl fentanyl	alpha-Methyl Thiofentanyl	2'-Fluoro ortho-Fluorofentanyl
alpha-Methyl Butyryl fentanyl	alpha-Methylfentanyl	2'-Fluoro, ortho-fluoro-cis-3-methyl Fentanyl
beta-methyl Acetyl fentanyl	beta-Hydroxyfentanyl	2'-Fluoro, ortho-fluoro-trans-3-methyl Fentanyl
2,2,3,3-Tetramethyl-Cyclopropyl fentanyl	beta-Hydroxythiofentanyl	3'-Fluoro ortho-Fluorofentanyl
2,3-Seco-Fentanyl	beta-Methyl Fentanyl	4'-Fluoro, ortho-fluoro-cis-3-methyl Fentanyl
2-Furanyl fentanyl	2'-Fluorofentanyl	4'-Fluoro, ortho-fluoro-trans-3-methyl Fentanyl
2'-Methyl Acetyl fentanyl	2'-Methyl Fentanyl	4'-Fluoro, para-fluoro-cis-3-methyl Fentanyl
3-Furanyl fentanyl	3-Methyl butyryl fentanyl	4'-Fluoro, para-fluoro-trans-3-methyl Fentanyl
3'-Methyl Acetyl fentanyl	3'-Methyl Fentanyl	4-Fluorobutyrylfentanyl
4'-Methyl acetyl fentanyl	3-Methylfentanyl	Despropionyl 2'-fluoro ortho-Fluorofentanyl
4-Phenyl fentanyl	4'-Fluorofentanyl	Despropionyl meta-Fluorofentanyl
Acetyl fentanyl	Acetyl norfentanyl	Despropionyl para-Fluorofentanyl
Acrylfentanyl	Acetyl-alpha-methyl fentanyl	Despropionyl-2-fluorofentanyl
Benzodioxole fentanyl	Butyryl norfentanyl	meta-Fluoro Acrylfentanyl
Benzyl Acrylfentanyl	cis-3-Methyl Butyryl fentanyl	meta-Fluoro Valeryl fentanyl
Benzylfentanyl	cis-3-methyl Norfentanyl	meta-Fluorobutyryl fentanyl
beta'-Phenyl fentanyl	cis-3-Methyl Thiofentanyl	meta-Fluorofentanyl
Butyryl fentanyl	Cyclopropyl norfentanyl	meta-Fluoroisobutyryl fentanyl
cis-Isofentanyl	Despropionyl meta-Methylfentanyl	o-Fluorofentanyl
Crotonyl fentanyl	Despropionyl ortho-Methylfentanyl	ortho-Fluoro Acrylfentanyl
Cyclobutyl fentanyl	Despropionyl para-Methylfentanyl	ortho-Fluorobutyryl fentanyl
Cyclohexyl fentanyl	Fentanyl	ortho-Fluoroisobutyryl fentanyl
Cyclopentenyl fentanyl	Fentanyl meta methylphenyl analog	para-Chloro Acrylfentanyl
Cyclopentyl fentanyl	Fentanyl meta tolyl analog	para-Fluoro Acrylfentanyl

Class 1	Class 2	Class 3
Cyclopropyl fentanyl	Fentanyl ortho tolyl acetyl analog	para-Fluoro Crotonyl fentanyl
Ethoxyacetyl fentanyl	Fentanyl ortho tolyl analog	para-Fluoro Cyclopentyl fentanyl
Fentanyl Carbamate	Fentanyl para methylphenyl analog	para-Fluoro Cyclopropyl fentanyl
Fentanyl meta methylphenyl acetyl analog	Fentanyl para tolyl acetyl analog	para-Fluoro Tetrahydrofuran fentanyl
Fentanyl methyl acetyl analog	Fentanyl propyl analog	para-Fluoro Valeryl fentanyl
Fentanyl Methyl Carbamate	Furanylethyl fentanyl	para-Fluoroacetyl fentanyl
Fentanyl ortho methylphenyl acetyl analog	Isobutyryl norfentanyl	para-Fluorobutyryl fentanyl
Fentanyl propyl acetyl analog	meta-Methyl Acetyl fentanyl	para-Fluoroisobutyryl fentanyl
Furanyl norfentanyl	meta-Methyl Cyclopropyl fentanyl	p-Fluorofentanyl
Heptanoyl fentanyl	meta-Methylfentanyl	
Hexanoyl fentanyl	N-(3-ethylindole) Norfentanyl	
Isobutyryl fentanyl	Norfentanyl	
Isovaleryl fentanyl	ortho-Methyl Acetyl fentanyl	
meta-Fluoro Furanyl fentanyl	ortho-Methyl Acrylfentanyl	
meta-Fluoro Methoxyacetyl fentanyl	ortho-Methyl Cyclopropyl fentanyl	
meta-Methoxy Furanyl fentanyl	ortho-Methylfentanyl	
meta-Methyl Furanyl fentanyl	para-Methyl Acetyl fentanyl	
meta-Methyl Methoxyacetyl fentanyl	para-Methyl Acrylfentanyl	
Methacrylfentanyl	para-Methyl Cyclopentyl fentanyl	
Methoxyacetyl fentanyl	para-Methyl Cyclopropyl fentanyl	
Methoxyacetyl norfentanyl	para-Methylfentanyl	
N,N-Dimethylamido-despropionyl fentanyl	Thiofentanyl	
N-benzyl Furanyl norfentanyl	Trans-3-methyl Norfentanyl	
N-Benzyl meta-fluoro Norfentanyl	trans-3-Methyl Thiofentanyl	
N-benzyl para-fluoro Cyclopropyl norfentanyl		
N-benzyl para-fluoro norfentanyl		
N-Benzyl para-fluoro Norfentanyl		
N-Benzyl phenyl norfentanyl		
N-methyl Cyclopropyl norfentanyl		
N-methyl Norfentanyl		
ortho-Fluoro Furanyl fentanyl		

Class 1	
ortho-Methoxy Furanyl fentanyl	
ortho-Methoxy-Butyryl fentanyl	
ortho-Methyl Furanyl fentanyl	
ortho-Methyl Methoxyacetyl fentanyl	
para-Chloro Cyclobutyl fentanyl	
para-Chloro Cyclopentyl fentanyl	
para-Chloro Cyclopropyl fentanyl	
para-Chloro Furanyl fentanyl	
para-Chloro Furanyl fentanyl 3-furancarboxamide	
para-Chloro Methoxyacetyl fentanyl	
para-Chloro Valeryl fentanyl	
para-Chlorobutyryl fentanyl	
para-Chlorofentanyl	
para-Chloroisobutyryl fentanyl	
para-Fluoro Furanyl fentanyl	
para-Fluoro Furanyl fentanyl 3-furancarboxamide isomer	
para-Fluoro Methoxyacetyl fentanyl	
para-Hydroxy Butyryl fentanyl	
para-Methoxy Acrylfentanyl	
para-Methoxy Furanyl fentanyl	
para-Methoxy Methoxyacetyl fentanyl	
para-Methoxy Valeryl fentanyl	
para-Methoxy-Butyrylfentanyl	
para-Methoxyfentanyl	
para-Methyl Furanyl fentanyl	
Phenoxyacetyl fentanyl	
Phenyl fentanyl	
Pivaloyl fentanyl	
Senecioylfentanyl	
Tetrahydrofuran fentanyl	
Tetrahydrofuran fentanyl 3-tetrahydrofurancarboxamide isomer	
Tetrahydrothiophene fentanyl	
Thienyl fentanyl	
Thiofuranyl fentanyl	
Thiophene fentanyl 3-thiophenecarboxamide	
Tigloyl fentanyl	
Valeryl fentanyl	