

MULTIVARIATE STATISTICAL ANALYSIS OF SPECTRAL REFLECTANCE IN
COLEOPTERA ELYTRA

A Thesis

Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Science

with a

Major in Statistics

in the

College of Graduate Studies

University of Idaho

by

Sarah Elizabeth-Marie Herberger

May 2014

Major Professor: Bahman Shafii, Ph. D.

Authorization to Submit Thesis

This thesis of Sarah Elizabeth-Marie Herberger, submitted for the degree of Master of Science with a Major in Statistical Science and titled “Multivariate Statistical Analysis of Spectral Reflectance in Coleoptera Elytra,” has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date _____
Bahman Shafii, Ph.D.

Committee
Members: _____ Date _____
Stephen P. Cook, Ph.D.

_____ Date _____
Christopher J. Williams, Ph.D.

Department
Administrator: _____ Date _____
Christopher J. Williams, Ph.D.

Discipline’s
College Dean: _____ Date _____
Paul Joyce, Ph.D.

Final Approval and Acceptance

Dean of the College
of Graduate Studies: _____ Date _____
Jie Chen, Ph.D.

Abstract

Coleoptera, commonly known as beetles, comprises 40% of all insects. Spectral readings of the elytra, the hardened outer wing, may be used to identify Coleoptera. Utilizing normal mixture models, eighteen peak wavelengths were identified across taxonomic groups and genders creating a multivariate structure. Multivariate procedures including principal component and discriminant analyses were employed to differentiate taxonomic groups and genders. The first three axes of the principal component analysis provided a clustering of genus and gender for a subset of taxonomic groups. The nearest neighbor discriminant analysis with proportional priors gave a misclassification rate of 5.2%. Internal bootstrap validation of the discriminant model yielded an average error rate of 3.5%. An external cross validation of the same model, conducted on independent samples resulted in an average misclassification of 6.5%. Given the low misclassification rate, multivariate statistical approaches are recommended for analysis of spectral reflectance in Coleoptera and other similar insect groups.

Acknowledgments

I would like to thank my graduate committee members, Drs. Bahman Shafii, Stephen Cook, and Christopher Williams for all the time and effort they gave me; without their dedication this thesis would not have been possible. I especially would like to thank Dr. Shafii, my advisor, for his plethora of knowledge.

I would like to acknowledge the University of Idaho Entomological Museum for the 390 specimens provided. I would also like to recognize Frank Merickel for the expertise he provided in specimen selection, without Frank or the Museum this project would not have been possible.

I would also like to recognize Dr. Alistair Smith at The University of Idaho for access to the newly calibrated spectrometer. Finally, I would like to thank Dr. Thomas Bitterwolf of The University of Idaho Chemistry Department for help in deciphering near-infrared spectrum range.

Dedication

I would like to dedicate this to my parents, Paul and Janis Herberger, who have always been there for me and continually encouraged me to reach for the stars. I want to thank my grandparents who have lead by example, and shown me that hard work can be rewarded. I would also like to dedicate this thesis to Dr. Bahman Shafii, who has been a counselor, mentor, and inspiration.

Table of Contents

Authorization to Submit Thesis	ii
Abstract	iii
Acknowledgments.....	iv
Dedication	v
List of Figures	ix
List of Tables	xii
Chapter 1	1
Introduction.....	1
Literature Review	5
Chapter 2	7
Material and Methods	7
Source and description of data	7
Statistical Analysis	10
Finite Mixture Models (FMM)	10
Principal Component analysis (PCA).....	12
Multivariate Discriminate Analysis	14
Linear discriminate analysis	14
Bayesian Nearest Neighbor or K-Nearest Neighbor.....	14
Validation	16

Internal Validation	16
External Validation	16
Chapter 3	18
Results and Discussion.....	18
Finite Mixture Model (FMM)	18
Principal Component Analysis (PCA)	22
Multivariate Discriminate Analysis.....	29
Linear discriminant analysis	29
Uniform Bayesian Prior Discriminant Analysis	32
Proportional Bayesian Prior Discriminant Analysis.....	34
Validation	36
Internal Validation	36
External Validation.....	37
Chapter 4	39
Conclusions	39
References	41
Appendix A	44
Appendix B.....	46
Appendix C.....	50
Appendix D	52

Appendix E..... 64

List of Figures

Figure 1. This diagram shows the experimental set-up.....	9
Figure 2. A generic PCA example	13
Figure 3. Normal mixture curves fit to female Lucanidae: <i>Lucanus capreolus</i>	19
Figure 4. Peak locations based on the Finite Mixture Model	20
Figure 5. The PCA scree plot.....	22
Figure 6. The 95% prediction ellipse displays PCA axis one vs PCA axis two	24
Figure 7. The 95% prediction ellipse displays PCA axis two vs PCA axis three.	25
Figure 8. The 95% prediction intervals separating male and female Cerambycidae: <i>Desmocerus piperi</i> (DE).....	26
Figure 9. Principal component loadings by the wavelengths.....	27
Figure 10. The heat map of the correlation matrix.....	28
Figure 11. The Heat Map of the Linear Discriminant Function	31
Table 4. Uniform prior discriminate analysis misclassification.....	33
Figure 12. The distribution of species misclassification rate for the internal bootstrap	37
Figure 13. The distribution of species misclassification rate for the external bootstrap	38
Figure B1. Shepard diagram badness of fit by dimension	48
Figure B2. The 95% prediction ellipse displays	

MDS dimension one vs MDS dimension two.....	49
Figure D1. Finite mixture model for Cerambycidae: Callidium sp. (CA) female	53
Figure D2. Finite mixture model for Cerambycidae: Callidium sp. (CA) male	53
Figure D3. Finite mixture model Cerambycidae: Desmocerus piperi (DE) female	54
Figure D4. Finite mixture model for Cerambycidae: Desmocerus piperi (DE) male	54
Figure D5. Finite mixture model for Buprestidae: Dicerca tenebrica (DI) female.....	55
Figure D6. Finite mixture model for for Buprestidae: Dicerca tenebrica (DI) male	55
Figure D7. Finite mixture model for Lucanidae: Lucanus capreolus (LC) female.....	56
Figure D8. Finite mixture model for Lucanidae: Lucanus capreolus (LC) male.....	56
Figure D9. Finite mixture model for Lucanidae: Lucanus mazama (LM) female.....	57
Figure D10. Finite mixture model Lucanidae: Lucanus mazama (LM) male	57
Figure D11. Finite mixture model for Buprestidae:	
Melanophila atropurpurea (ME) female	58
Figure D12. Finite mixture model for Buprestidae:	
Melanophila atropurpurea (ME) male.....	58
Figure D13. Finite mixture model for Buprestidae:	
Buprestis lyrata Casey (PC) female	59
Figure D14. Finite mixture model for Buprestidae:	
Buprestis lyrata Casey (PC) male	59
Figure D15. Finite mixture model for Cerambycidce:	

Prionus californicus (PR) female 60

Figure D16. Finite mixture model for Cerambycidae:

Prionus californicus (PR) male 60

Figure D17. Finite mixture model for Cerambycidae

Spondylis upiformis (SP) female 61

Figure D18. Finite mixture model for Cerambycidae

Spondylis upiformis (SP) male 61

Figure D19. Finite mixture model for Trogossitidae:

Temnocheila chlorodia (TE) female 62

Figure D20. Finite mixture model for Temnocheila

chlorodia (TE) chlorodia male 62

Figure D21. Finite mixture model Buprestidae:

Trachykele blondeli blondeli (TR) female 63

Figure D22. Finite mixture model for Buprestidae:

Trachykele blondeli blondeli (TR) male 63

List of Tables

Table 1. List of Coleoptera taxa.....	8
Table 2. Detailed outline of the 18 variables generated from FMM.....	21
Table 3. Linear discriminant analysis misclassification results.....	30
Table 4. Uniform prior discriminate analysis misclassification.....	33
Table 5. Proportional prior discriminate analysis misclassification of individual species.....	35
Table A1. The first 20 observations by individual insect of relative reflectance values at peak values of R1-R18 selected by the FMM.	45
Table B1. The first 20 observations by individual insect of relative reflectance values at peak values of R1-R18 selected by the FMM from the validation database	51
Table E1. Correlation coefficients among variables R1 through R18 generated using FMM	65

Chapter 1

Introduction

Insects are one of the most abundant, diverse, and necessary life forms on earth. They play an integral role in pollination, degradation of waste, maintenance of pests, and medicine. They are also used as model organisms for conducting a variety of scientific research. The order Coleoptera makes up over 50% of known insect species, with 350,000 species of Coleoptera having been formally described. Estimates on the number of Coleoptera species range from 600,000 to 30 million (Seago et al. 2009). Coleoptera can be found in every terrestrial climate in the world with species diversity often increasing in tropical locations (Vigneron et al. 2006). The majority of Coleoptera species are undescribed, even when using conservative estimates. With such a low percentage of species having been described, it can be difficult for an Entomologist to correctly identify an insect. Taking into account human error, the probability that one will incorrectly identify a Coleoptera is higher than what is considered acceptable in a research setting. However, the ability to rapidly differentiate Coleoptera taxonomic groups with little to no error in this modern age should be within our means.

The methods used for identifying Coleoptera species are usually difficult and inaccurate. The classification is typically derived from antennal, tarsi, mouthparts (labial and maxillary palpi), ventral characters (sterna, pleura, coxae), and other morphological characteristics (Choate 1999). There is a high potential for misclassification that can occur in the process of identification. For example, long-horned beetles which do not have long antennae, snout beetles which do not have snouts, ground beetles that live in trees, or aquatic beetles that are never in the vicinity of water (Choate 1999). Morphology of an insect has to be painstakingly analyzed in order to identify them accurately, i.e. antenna measured, veins on

wings analyzed, carapace shape diagrammed, etc. Typically, Coleoptera are viewed under a dissecting microscope where traits can be analyzed. Memorizing or locating references for morphology and then applying that knowledge for the process of identification can be very time consuming. Coupled with human error, and the ever expanding number of described species, this may lead to misclassifications. For example, one of the taxonomic groups chosen for this study, the genus *Callidium* within the family Cerambycidae, has been viewed by three different entomological experts with each one identifying it differently.

One of the most accurate ways to differentiate Coleoptera species is through their coloring. In fact, entomologists have created more than 30 different terms that are used to describe the color brown (Seago et al. 2009). With such a strong emphasis on color, the most distinguishable coloration is often seen in the hard outer wing, or elytra, of Coleoptera. The elytra typically have a relatively uniform coloration with the most frequent colors seen being blue or green (Piszter 2010). The elytra are composed of chitin, with elements such as carbon, hydrogen, nitrogen, oxygen, calcium, and magnesium present to achieve a particular coloration (Piszter 2010). Elytral coloration is exposed to some of the strongest evolutionary pressures (Piszter 2010) which include, but are not limited to crypsis, aposematic, sexual signals, polarized signaling (for conspecific communication), thermoregulation, and confusion of depth perception of predators (Seago et al. 2009). Coloration in Coleoptera has been observed to change during development or as a result of environmental conditions (Seago et al. 2009). Elytral coloration has been shown to vary along geographical gradients (Kawakami et al. 2013). This is typically referred to as thermal melanism, where low elevations or temperatures have high production of dark pigments. Elytral coloration has also been shown to correlate with humidity, rainfall, temperature, and isolation (Kawakami et al. 2013).

All of the specimens selected for this project are wood borers or predators of wood boring insects. Wood-boring beetles, or Woodborers, are considered pests in trees and some wooden structures. The mandibles of these species are specifically designed for wood consumption. When trees are dying, they release certain chemicals into the air which attract wood-boring beetles to the specific location. When trees are cut down, for industrial purposes, they release chemicals signaling their death, attracting wood boring beetles. The destruction typically stops when pheromones are released from insects in the tree, letting other insects know that the tree is full. Wood-boring beetles are considered a pest by most people because of their tendency to infest trees and wooden structures.

Eggs of woodborers are typically laid in the crevices of bark or under the bark. Larvae will fully develop in the tree. Individuals typically bore under the bark of trees and often times into the xylem. Adults typically live for a short period of time, long enough to reproduce and begin the life cycle again.

The taxonomic groups selected from the University of Idaho William Barr Entomology Museum for this study included species in the families: Cerambycidae *Callidium* sp., Cerambycidae *Desmocerus piperi* Webb, Buprestidae *Dicerca tenebrica*, Lucanidae *Lucanus capreolus*, Lucanidae *Lucanus mazama*, Buprestidae *Melanophila atropurpurea*, Buprestidae *Buprestis lyrata* Casey, Cerambycidae *Prionus californicus* Motschulsky, Cerambycidae *Spondylis upiformis* Mannerheim, Trogossitidae *Temnocheila chlorodia*, and Buprestidae *Trachykele blondeli blondeli*. Under most museum conditions, beetles have been shown to retain their color (Seago et al. 2009). Previous research has indicated that near infrared reflectance can be used for rapid identification of wheat pests (Dowell et al. 1999;

Vigneron et al. 2006). Slight variations of color allow one to distinguish between closely related species, as well as genders within the same species (Vigneron et al. 2006).

The objectives of this study was to differentiate Coleoptera taxonomic groups, as listed above, in the William Barr entomology museum at the University of Idaho, based upon the spectrometer readings. It was also intended to differentiate the gender of the aforementioned taxonomic groups using the same methodology.

Literature Review

Many previous studies have arbitrarily chosen regions of spectrometer readings before taking any readings, which solely remained in or very close to the visible light range (Thrope 2002; Vigneron et al. 2006). Human vision is restricted to only 400 – 700 nm wavelengths, while many other organisms use different or wider ranges to communicate (Thrope 2002). Thrope (2002) advocates that wavelengths should be chosen free of biological context, and that they should be non-overlapping, in unequal ranges, with gaps existing between selections. However, other researchers maintain that the selection of spectrometer wavelengths should take into account the conspecifics method of communication, as well as the data, and not rely on visual ability. There are many wavelengths within which beetles communicate, ranging from the ultraviolet to the infrared. For example, the ultraviolet spectrum has been studied in beetles, typically for the use of mate or conspecific recognition and communication (Seago et al. 2009). Also, some beetles also communicate through the use of hormones and can be detected in the near-infrared range.

A wide assortment of research has used spectrometer readings. One common use of a spectrometer is identifying the chemicals present in a reaction (Barber et al. 1963). Another use is the tracking of certain isotopes through a food web, particularly the detection of ^{13}C in bacteria and invertebrates to identify carbon pathways (Hall 1995). Other uses of spectrometers have been extraterrestrial, acquiring the measurements of temperature, composition, and atmosphere of planets and stars (Lillie 1972). Spectrometers have also been used in the measurement of barrower bird color and prediction of their visual system evolution (Endler et al. 2005).

Coloptera elytral color and composition is thought to be highly related to phylogenetic relationships, habitat adaptation, sexual selection, and other interactions (Thrope 2002). Spectrometer readings classify color as well as chemical bonds between molecules in readings from 350 to 2500 nm. Dowell, et al. (1999) has shown that near-infrared reflectance, 400 to 1700 nm, can be used for the distinction of eleven laboratory-raised Coloptera wheat pest species on controlled diets. However, further research needs to be conducted with a wider spectral range using different species on an uncontrolled diet and location. This would ensure correctly classifying species beyond the controlled settings.

Various univariate and multivariate statistical techniques have been previously used to analyze spectrometer data. The most common approach is the principal component analysis (Thrope 2002; Endler et al. 2005; Rousu, et al. 2013). Other methods include least-square regression, discriminant analysis function, analysis of variance, students-t test and canonical correlation (Dowell et al. 1999; Cook et al. 2010; Rousu 2013). Spectrometer readings generate multiple peaks across the admissible range of wavelengths and hence, create a multivariate response structure. In this study, multivariate methods including principal component analysis, as well as parametric and nonparametric discriminant analyses were used to differentiate Coloptera taxonomic groups and genders based on the spectrometer readings of their coloration. Internal and external cross validations were carried out using the bootstrap simulation technique.

Chapter 2

Material and Methods

Source and description of data

The insect order Coleoptera was selected because of their overwhelming commonality and unique identifying body parts, such as large elytra. Specifically, wood boring Coleoptera specimens were selected from the William F. Barr Entomological Museum (College of Agricultural and Life Sciences, University of Idaho, Moscow, Idaho) , controlling for the location and year within a taxa. The collections at the William F. Barr Entomological Museum date back to 1893 and the holdings are of significant historical as well as scientific importance. It is a substantial regional and national resource for specimens from the intermountain west, in addition to containing a worldwide representation of select taxa. Given its breadth of specimens, the museum provided a unique opportunity to examine several families of wood borers. Table 1 provides the taxa, year, collection location, number of individuals and abbreviations for the species used in this study.

Table 1. List of Coleoptera taxa used with year collected, location collected, number of individual specimens measured and respective abbreviations

Family: Genus Species	Year	Location	Number of individuals	Abbreviation
Cerambycidae: <i>Callidium</i> sp.	1990	Clark mountain	18	CA
Cerambycidae: <i>Desmocerus piperi</i>	1963	lost trail pass Idaho	18	DE
Buprestidae: <i>Dicerca tenebrica</i>	1954	bear creek camp 10 min north of Leslie	20	DI
Lucanidae: <i>Lucanus capreolus</i>	2006	Camden AR	7	LC
Lucanidae <i>Lucanus mazama</i>	2006	Kanal Utah	22	LM
Buprestidae: <i>Melanophila atropurpurea</i>	2012	I 84 rest stop nearest to Utah border	18	ME
Buprestidae: <i>Buprestis lyrata</i>	1982	5 min west of paradise pt. Palouse range ID	24	PC
Cerambycidae: <i>Prionus californicus</i>	2008	Parma research center	27	PR
Cerambycidae: <i>Spondylis upiformis</i>	1976	3.4 miles west of clarkia Idaho	19	SP
Trogossitidae: <i>Temnocheila chlorodia</i>	1977		26	TE
Buprestidae: <i>Trachykele blondeli blondeli</i>	1966	Marion county Oregon	10	TR

The data collection was carried out in two separate dark room laboratories in order to control the lighting. The first room was located in the College of Agricultural and Life Science (operated under Professor Stephen Cook), while the second room was located at the College of Natural Resources (operated under Professor Alistair Smith). Specimens were enclosed in an area painted with Krylon Ultra-Flat black paint. This paint was chosen because it does not register on the spectral instrument used and, therefore, provides a null background for the desired readings.

Each insect was attached through the left, front wing with a standard insect mounting pin. A description of the insect's scientific name, collection date and location was attached below the insect. Spectrometer readings of insects were collected with a FieldSpec® Pro Full Range model, with the spectral acquisition range of 350 to 2500 nanometers (nm). This instrument has a resolution of 3 nm at 700 nm, and 10 nm at 1400 nm and 2100 nm (ASD Inc. 2012).

Fiber optics, connected to the spectrometer, were maintained at a 3 centimeter distance from the specimens and manipulated through a pistol grip control affixed at a 90° angle to the specimens target area. Each specimen was sequentially illuminated across a spectrum of 400 to 700 nm. The light source used was a Smith Vector Corp Photographic Light Model 750-SG outfitted with a full spectrum light bulb and placed at a 45° angle, one meter away from the specimens. The experimental setup is presented in figure 1.

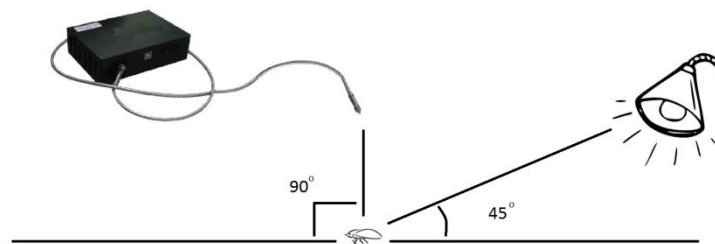


Figure 1. This diagram shows the experimental set-up of light, specimen & spectrometer. This set up was used to reduce the direct light from the source while still fully illuminating the specimen.

Specimen's elytral spectral relative reflectance (%) was recorded at each wavelength (nm). The relative reflectance was the percentage of a white 8° hemispherical spectral reflectance factor for SRT-99-050. After every third spectrometer reading, the hemispherical

spectral reflectance factor was recorded. This allowed base reference points for the data, and ensured that the machine's calibration remained constant.

Each specimen was measured three times with the spectrometer. The instrument recording software (R^3), itself, averaged three additional shots for each of these observations. Following data collection, the three manual observations per specimen were averaged, effectively giving one spectral data point based on nine spectrometer readings. This was intended to reduce any potential measurement errors. Eleven taxa were measured, and each included approximately the same number of male and female specimens. Replication (individuals per taxa) ranged from a minimum of three to a maximum of 12, for a total of 210 insects. An overall multispectral database was subsequently created from these specimens which encompassed reflectance measurements of 2150 wavelengths.

Statistical Analysis

Finite Mixture Models (FMM)

In order to approximate the multi-modal spectral data series, finite mixture models were used, assuming normal distribution components. Finite mixture models have the general form of $\sum_{i=1}^q p_i f_i(x_j)$. Assuming a normal distribution model basis, the finite mixture model becomes:

$$\begin{aligned} \text{pr}(x_j) &= \sum_{i=1}^q p_i \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right) e^{\left(\frac{-(x_{ij}-\mu_i)^2}{2\sigma_i^2} \right)} = \\ & p_1 \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \right) e^{\left(\frac{-(x_{1j}-\mu_1)^2}{2\sigma_1^2} \right)} + \dots + p_q \left(\frac{1}{\sqrt{2\pi\sigma_q^2}} \right) e^{\left(\frac{-(x_{qj}-\mu_q)^2}{2\sigma_q^2} \right)} \end{aligned} \quad (1)$$

where μ_i is the i^{th} component peak (mean and mode), σ_i is the associated standard deviation, x_{ij} represents the univariate observed response of the i^{th} component at the j^{th} wavelength, and p_i is the proportion accounted for by the i^{th} mixture component, where $0 \leq p_i \leq 1$, and $\sum_{i=1}^q p_i = 1$, satisfying the necessary conditions for a complete probability distribution. The above represents the univariate method for identifying the multiple peaks in the original wavelength data. Finite mixture models have been previously used to describe and compare other biological responses, such as the length distributions of mountain white fish (Shafii et al. 2010). Also, Royle and Link (2005) created a Gaussian mixture model of Anuran call surveys to predict species abundance.

Procedure FMM in SAS 9.3 was used to fit a varying number of normal curves mixture model components to the data for 22 separate taxa and gender groups within the data, i.e. 11 species, both male and female. Following adequate model estimation, the wavelengths at the corresponding model component peaks, μ_i , were chosen as the basis for further analysis. This provided a way to reduce the number of wavelengths from 2150 down to a more manageable database where false positives were less likely to occur.

Principal Component analysis (PCA)

PCA takes multidimensional data sets and reduces their dimensions by determining one or more linear combinations of the variables (wavelengths) that account for the largest variation in the data. In our case, wavelength $x_1, x_2, x_3, \dots, x_p$, selected by the FMM, may potentially be correlated and can lack the ability to show any underlying data structure by themselves. PCA can help define potential unobserved latent variables in the data by reducing the inherent dimensions of the problem through a centering of the data origin to \bar{x} and subsequently rotating the data using:

$$z_1 = A(x_p - \bar{x}) \quad (2)$$

where A is an orthogonal matrix of coefficients, and z_1 is x_1 rotated. The rotation is done such that $(z_1, z_2, z_3, \dots, z_p)$ are uncorrelated to one another and the covariance matrix of $(z_1, z_2, z_3, \dots, z_p)$ will be defined as:

$$S_z = ASA' = \begin{bmatrix} s_{z_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_{z_p}^2 \end{bmatrix} \quad (3)$$

$s_{z_p}^2$ is the Eigen vector, λ_p , where λ_1 has the largest variance and λ_p has the smallest variance (Rao 1964).

For example, as seen in figure 2, the original variables X and Y are compressed to Axis 1 = $a_1X + b_1Y$, and Axis 2 = $a_2X + b_2Y$. Axes 1 and 2 are Eigen vectors; a and b are the loadings of the respective variables. The purpose of PCA is to retain as few axes as possible while explaining the majority of the data variability among taxa (Morrison 1976; James and McCulloch 1990). The axes represent lurking variables and the coefficients (loadings) of these axes convey information on the relative importance of individual variables.

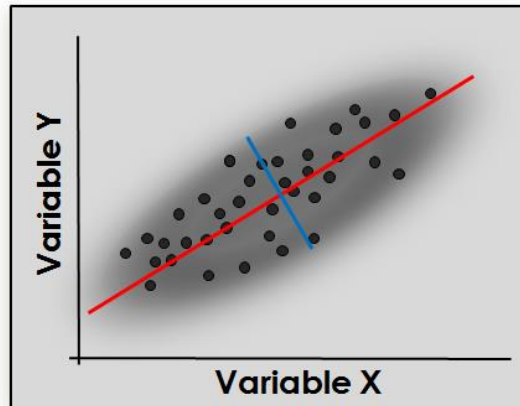


Figure 2. A generic PCA example. Axis 1 = $a_1X + b_1Y$, Axis 2 = $a_2X + b_2Y$. Axis 1 & 2 are Eigen vectors; a and b are the loadings of the variables.

Principal component analysis is widely used in ecology (James and McCulloch 1990). The ecological applications vary from differentiating infested oak trees (Colemana et al. 2011) to the color differences seen in anoles. Another example of the use of principal component analysis is to classify wine using spectrometer data of trace elements (Taylor et al. 2003). The SAS procedure PRINCOMP was used for PCA estimation, based on the underlying correlation matrix.

An additional ordination technique, multidimensional scaling, based on a dissimilarity matrix, was also attempted in order to distinguish underlying dimensions that could clarify the similarities or dissimilarities of the data. However, the results were not satisfactory. A brief description of the technique as well as the results appears in appendix B.

Multivariate Discriminate Analysis

Linear discriminate analysis

Normal discriminate analysis is typically of the form

$$L_k = (2\pi)^{-\frac{k}{2}} |V_k|^{-\frac{1}{2}} \exp(-0.5M_k) \quad (4)$$

With L_k representing the likelihood that an individual belongs to species k and V_k the variance-covariance matrix obtained from the k^{th} species. M_k is interpreted as the Mahalanobis distance given by:

$$M_k = (d - x_k)' V_k^{-1} (d - x_k) \quad (5)$$

The Mahalanobis distance measures the distance between the data response vector, d , and a known vector of responses from the k^{th} species, x_k (Lachenbruch 1964).

Multidimensional normal discriminate analysis has aided in the identification of insects prior to this study. For example, the identification of the Africanized honey bees in the U.S. based upon the insect's characteristics (Daly and Balling 1978).

Bayesian Nearest Neighbor or K-Nearest Neighbor

The nearest neighbor rule was first introduced by Fix and Hodges in 1951.

Subsequently, a nearest neighbor discriminate analysis was proposed for selecting the K^{th} nearest points using the distance function:

$$M_k = (x_m - x_n)' S_{pl}^{-1} (x_m - x_n) \quad (6)$$

where S_{pl}^{-1} is the inverse of the pooled sample variance-covariance matrix from the defined sample, x_m is a data point of interest, and x_n is all other data points. The purpose of this technique is to classify each x_m , using the k points nearest to x_n . That is, if the majority of

the k points belong to group 1, assign x_m to group 1; otherwise, assign x_m to another group, etc. K-Nearest Neighbor Discriminate Analysis, or Non-parametric Discriminate Analysis, dispenses with the need to make probabilistic assumptions for likelihood determinations.

Previously, Bayesian discriminate analysis has been used for classification of other biotic factors. For example, the yellow starthistle growth patterns are based on landscape characteristics (Shafii et al. 2004).

Bayesian discriminant analysis modifies (4) through the addition of a prior assumption on group assignments. A base model for this would use a uniform or uninformed prior for discriminate analysis following:

$$p(\text{insect is in } k^{\text{th}} \text{ species} | d) = \frac{q_k L_k}{\sum q_k L_k} \quad (7)$$

This will produce a probability between 0 and 1 with q_k as the prior probability, where k represents the number of species as follows:

$$q_k = \frac{1}{k} \quad (8)$$

Alternatively, a proportional prior for discriminate analysis can be defined as the proportion of observations from each group in the input data set (Hinich 1979). When data are balanced, the proportions for each group are the same so this method becomes equivalent to the uniform prior and hence, may be only relevant when the groups are unbalanced. It can be seen in (7), where we know that q_k is the prior probability for species k as follows:

$$q_k = \text{number of individuals in species } k / \text{total number of individuals} \quad (9)$$

The nearest neighbor methods aid in the prediction of species based upon multivariate spectrometer readings. Procedure DISCRIM in SAS 9.3 was used for all discriminant analyses estimations and validations.

Validation

Internal Validation

Bootstrap is a resampling technique, with replacement, that is done when one is unsure about the behavior of the target population (Efron 1976). By randomly selecting a subsample (X_i^*) from the sample (X_i), a new sample is produced which is selected from a known population. By analyzing the relationship between the sample and subsample, conclusions can be drawn about the actual population. Gathering data on the population would require a census, which would be impractical for a subject such as Coleoptera due to the number of possible individuals. Bootstrap procedures, therefore, provide a practical means of assessing the differences between Coleopteran species and the analyses carried out in this study.

An internal bootstrap of proportional discriminate analysis was performed through bootstrap sampling with replacement and data splitting. Data splitting was completed using a proportion of the database to create the model, while the remainder of the data was utilized for validation. The bootstrap sample, X_i^* , was selected from the data X_i at predefined proportions of sex and species in the database. For each bootstrap sample, two types of misclassification were possible: omission (Type I), or commission (Type II). An error of omission occurred when an observation, X_i^* , was classified outside of its true type, while an error of commission occurred when an observation was placed in the wrong type. Confidence intervals, means, and standard deviations were created from $B = 5000$ bootstrap simulations.

External Validation

A new independent database was created from 180 insects of the same species that were not previously sampled. External validation was carried out using these data and the

same methodology as the internal validation, that is, a bootstrap of discriminate analysis assuming a proportional prior. Unlike the first database, however, the insects chosen for inclusion were not controlled for location or year. This validation provided a robust confirmation of the adequacy of the estimated discriminant model.

Chapter 3 Results and Discussion

Finite Mixture Model (FMM)

As an objective means of identifying the peaks (modes) of the spectral distributions, equation (1) was fitted separately to each species-gender combination assuming spectral reflectance values were proportional to their probability of being observed.

The number of normal curve components was allowed to vary and were ultimately estimated from the distribution of the data. The final number of components ranged from 3 to 8 distributions per species - gender group. Thus, each of the 22 groups had a different set of fitted normal curves. The peaks (e.g. the means) were selected from the normal curves as a technique to quantifying the strongest wavelengths in the spectrum. The set of peak bands from each spectrum could then be used as a basis for comparing species-gender combinations. An example for the female Lucanidae: *Lucanus capreolus* data set is given in Figure 3. In that case, six peaks were identified and ranged from 977 nm to 2133 nm. The estimated fits for the 22 distributions can be viewed in appendix D.

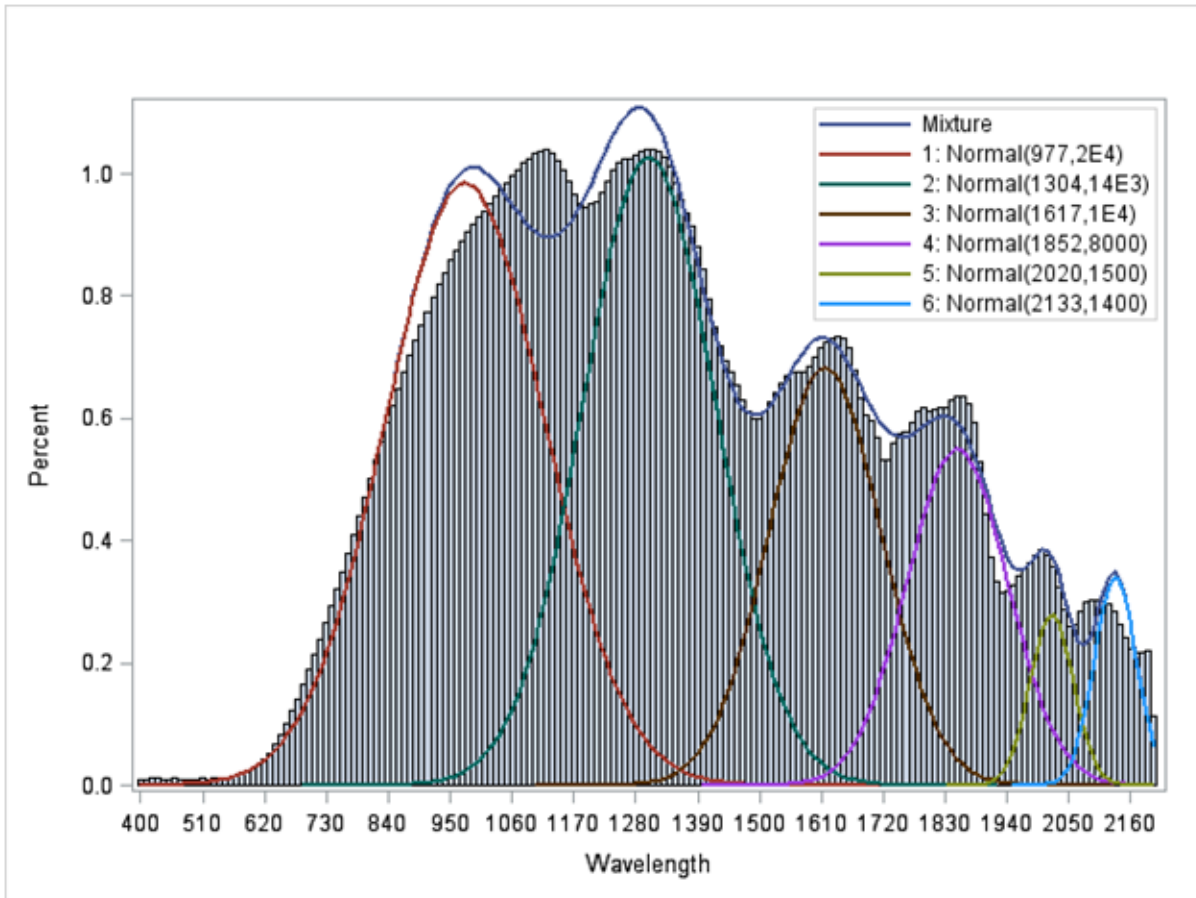


Figure 3. Example fit of normal curves fitted to the female Lucanidae: *Lucanus capreolus* distribution.

Overall, a large number of peaks were identified. To assess any commonalities among the 22 species-gender combinations, peak placement in relation to the wavelength was graphed (Figure 4).

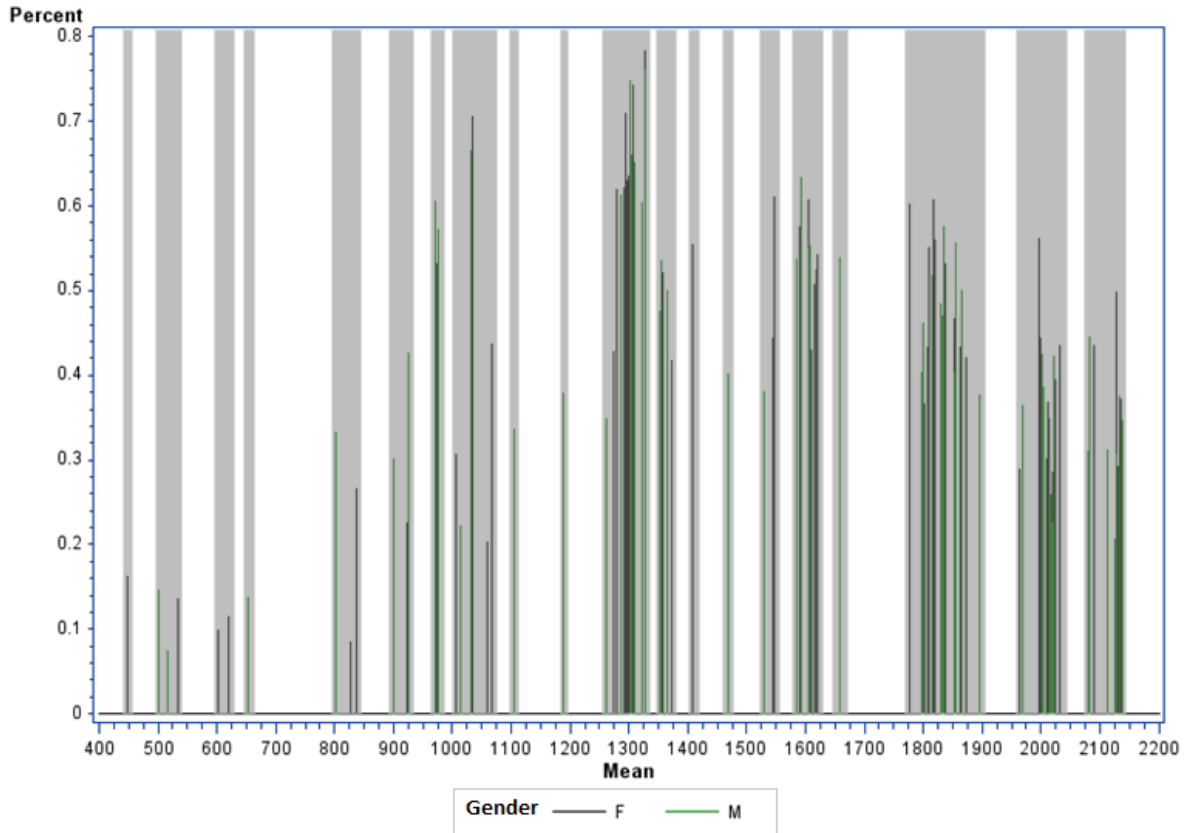


Figure 4. Peak locations based on the Finite Mixture Model in relative reflectance (percent) by wavelength. The green lines are male and the black lines are female. The lines are representing the relative reflectance at peak locations as identified by equation (1). The grey shaded area is emphasizing the aggregation of the 18 peak observations.

From figure 4, it was determined that the peaks showed some aggregation and hence, it led to the creation of 18 common peaks ($R_1, R_2, R_3, \dots, R_{18}$), i.e. 18 different bandwidths selected as a common dataset across species. A detailed outline of the 18 variables generated from FMM procedure and their corresponding bandwidths are given in Table 2.

Table 2. Detailed outline of the 18 variables generated from FMM procedure and their corresponding bandwidths

Bandwidth	Lower Limit	Upper Limit	New Variable	Wavelength	Mean
10	440	450	R1	445	445
50	500	550	R2	525	525
60	600	660	R3	630	630
50	800	850	R4	825	825
30	900	930	R5	915	915
20	960	980	R6	970	970
125	1000	1125	R7	1063	1062.5
50	1175	1225	R8	1200	1200
80	1250	1330	R9	1290	1290
30	1350	1380	R10	1365	1365
25	1400	1425	R11	1413	1412.5
20	1460	1480	R12	1470	1470
25	1525	1550	R13	1538	1537.5
45	1580	1625	R14	1603	1602.5
25	1650	1675	R15	1663	1662.5
125	1775	1900	R16	1838	1837.5
90	1950	2040	R17	1995	1995
65	2075	2140	R18	2108	2107.5

The new 18-variable database provided a manageable number of variables for subsequent analyses. A sample of the database used is presented in Appendix A.

Principal Component Analysis (PCA)

The relationship between the 18 variables created by FMM was investigated using principle components analysis (PCA). The analysis implemented equations (2) and (3) in order to obtain the Eigen vectors or PCA axis. The first PCA axis explained 66.84% and the second PCA axis explained 19.88% of the total variability in the data. The third axis explained 10.3% of the variability, while the amount of variability explained by PCA axis 4 through 18 was less than 5%. The retention of three PCA axis, or a three dimensional space, explained 96.3% of the variability. The third axis would normally not have been considered, however, the 10.3% of the variability explained by that axis provided an increased separation between species and gender. The scree plot further detailing the first six PCA axes is given in Figure 5.

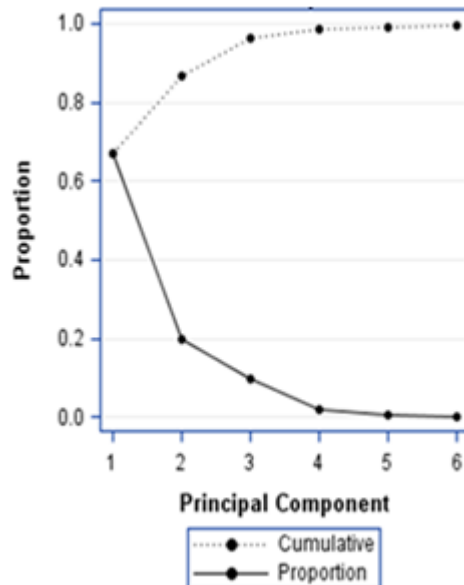


Figure 5. The PCA scree plot showing the variance explained by the first six PCA axes.

Plots of the resulting first three PCA axes, coded by species, are given in Figures 6 and 7. The ellipses represent an approximate 95% confidence region for each species, assuming bivariate normality. The separation of species seen in PCA axis two vs. PCA axis three (Figure 7) and PCA axis one vs. PCA axis two (Figure 6) is more closely related to PCA axis two. In both figures (6 and 7), LC (Lucanidae: *Lucanus capreolus*), LM (Lucanidae: *Lucanus mazama*), and PR (Cerambycidae: *Prionus californicus*) separate from the rest of the species. The angle of their respective ellipses also varies from other species in these plots.

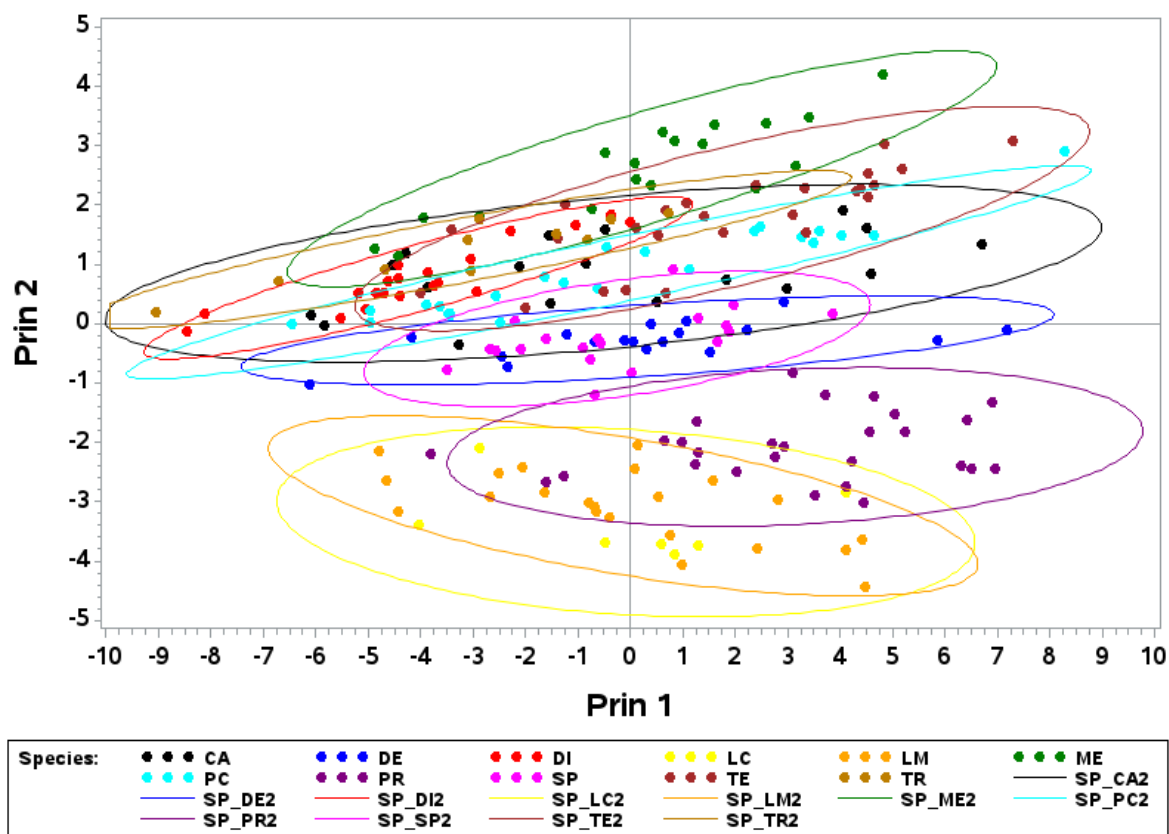


Figure 6. The 95% prediction ellipse displays PCA axis one vs PCA axis two. The points are the original data points projected into the PCA space. The abbreviations represent the following species: Cerambycidae: *Callidium* sp. (CA, SP_CA2), Cerambycidae: *Desmocerus piperi* (DE, SP_DE2), Buprestidae: *Dicerca tenebrica* (DI, SP_DI2), Lucanidae: *Lucanus capreolus* (LC, SP_LC2), Lucanidae: *Lucanus mazama* (LM, SP_LM2), Buprestidae: *Melanophila atropurpurea* (ME, SP_ME2), Buprestidae: *Buprestis lyrata* Casey (PC, SP_PC2), Cerambycidae: *Prionus californicus* (PR, SP_PR2), Cerambycidae: *Spondylis upiformis* (SP, SP_SP2), Trogossitidae: *Temnocheila chlorodia* (TE, SP_TE2), Buprestidae: *Trachykele blondeli blondeli* (TR, SP_TR2).

In Figure 7, species DE appears to be at a 90° angle to other species, particularly species PR, giving some indication that they are independent of one another. Also, LM and DI are mirror angles from one another, separated by 180 degrees, thus implying that they are negatively correlated judged by the sign of their respective PCA loading. For figures 6 and 7, the separation of species and gender were not clear when viewing all 22 groups. However, as seen in Figure 8, plotting the data separately by individual species can discern some separation by gender.

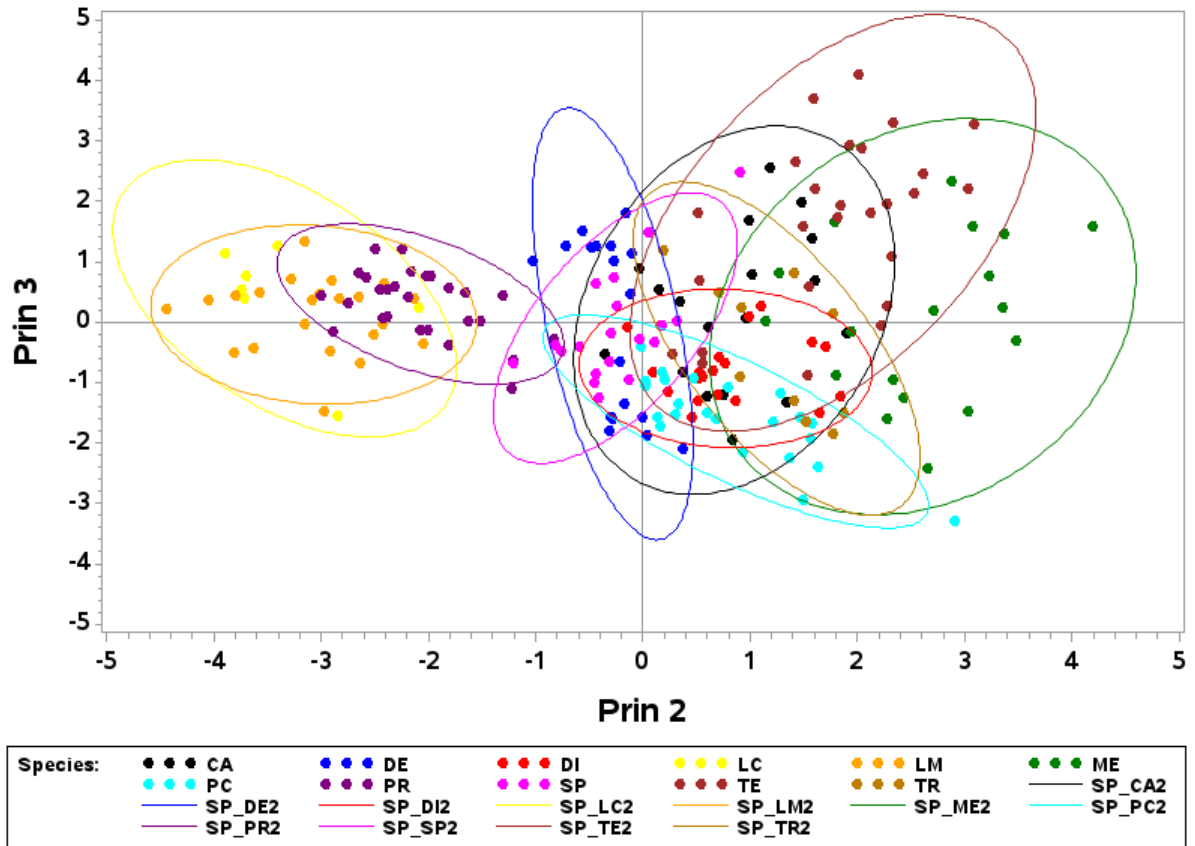


Figure 7. The 95% prediction ellipse displays PCA axis two vs PCA axis three. The points are the original data points projected into the PCA space. The ellipses are the 95% confidence interval assuming bivariate normality. The abbreviations represent the following species: Cerambycidae: *Callidium* sp. (CA, SP_CA2), Cerambycidae: *Desmocerus piperi* (DE, SP_DE2), Buprestidae: *Dicerca tenebrica* (DI, SP_DI2), Lucanidae: *Lucanus capreolus* (LC, SP_LC2), Lucanidae: *Lucanus mazama* (LM, SP_LM2), Buprestidae: *Melanophila atropurpurea* (ME, SP_ME2), Buprestidae: *Buprestis lyrata* Casey (PC, SP_PC2), Cerambycidae: *Prionus californicus* (PR, SP_PR2), Cerambycidae: *Spondylis upiformis* (SP, SP_SP2), Trogossitidae: *Temnocheila chlorodia* (TE, SP_TE2), Buprestidae: *Trachykele blondeli blondeli* (TR, SP_TR2).

In Figure 8, Cerambycidae: *Desmocerus piperi* (DE) indicates separation between male and female. The ellipse shapes are different indicating that males are better described by PCA axis 2 while the females are described by both PCA axes 2 and 3.

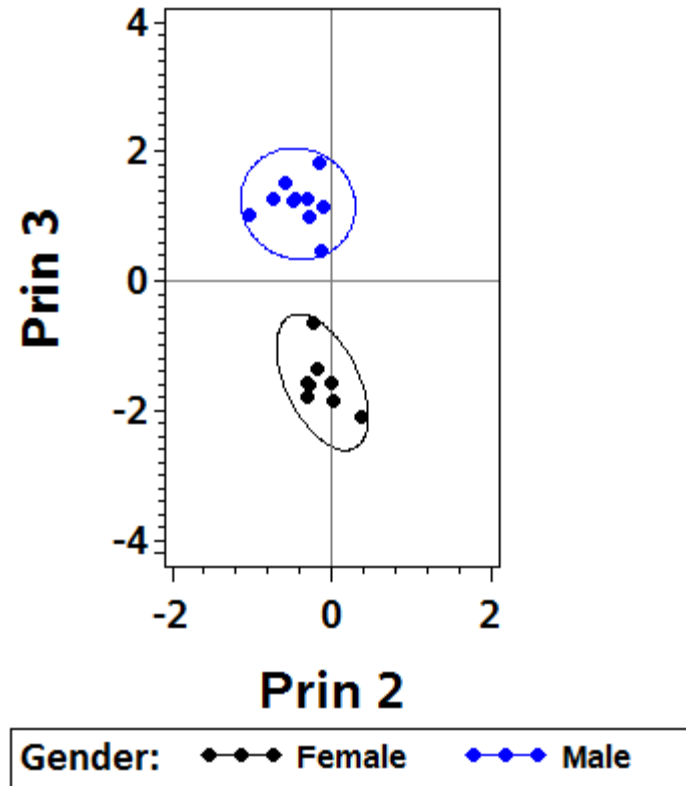


Figure 8. The 95% prediction intervals separating male and female of Cerambycidae: *Desmocerus piperi* (DE) when viewed by PCA axis two and three.

The PCA loadings for each variable by wavelength ($R_1, R_2, R_3, \dots, R_{18}$), are plotted in Figure 9. The first PCA axis (red), primarily explains the overall variability through a positive loading value across the spectrum. The second PCA axis (green) explains data variability by providing an approximate inverse to the respective loadings of the third principal component (yellow). While the true meaning of these axes is purely speculative, the inverse behavior seen between PCA axis two and three and the relationship shown in Figure 8 may indicate some gender differentiation based on these axes.

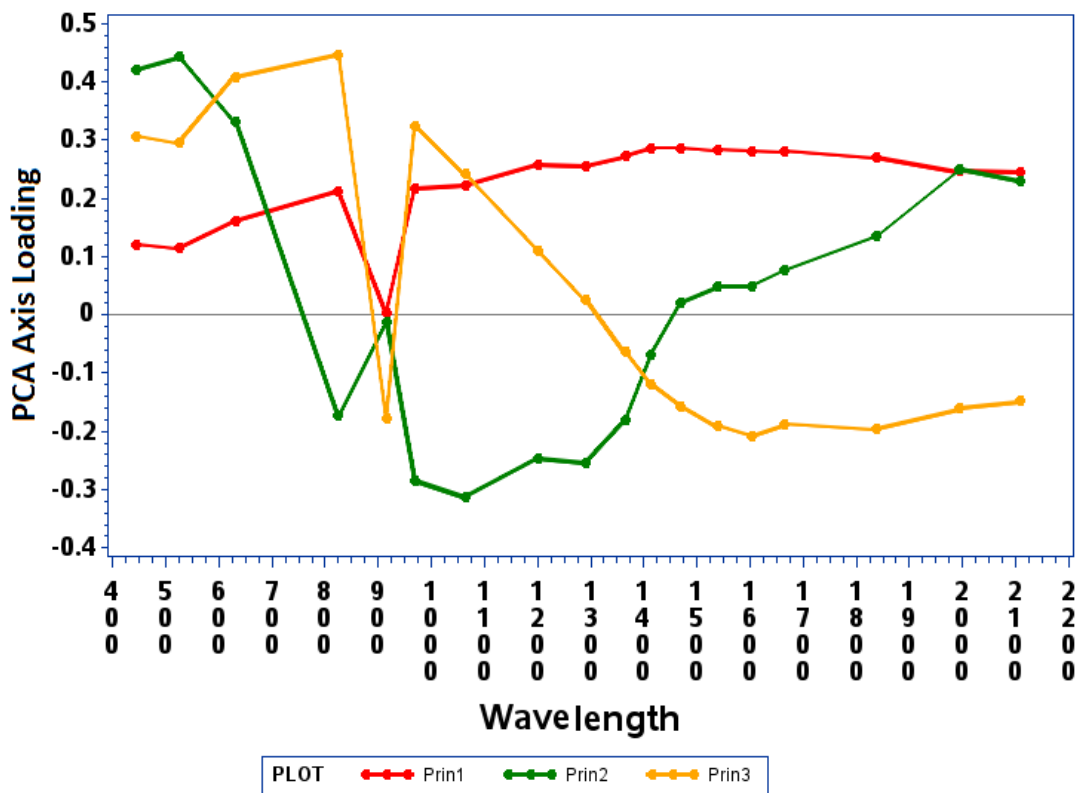


Figure 9. Principal component loadings by the wavelengths (R_1, R_2, \dots, R_{18}) is defined by principal component axis 1 (red) 2 (green) and 3 (yellow).

Figure 10 displays a heat map of the correlation matrix for the peak wavelength values. The values of correlation along the diagonal are one, or very close to one (white). This signifies that variables (peak wavelength values) close to one another are highly correlated. The lower correlation values observed between R_1, R_2, R_3 , or rather the visual spectrum, (400 to 700 nm), versus R_4 through R_{16} does not correlate with the near infrared spectrum (800 – 1800 nm). It is unexpected, however, that the visual spectrum, R_1, R_2, R_3 , is correlated to $R_{17}, & R_{18}$. The visual spectrum encompasses what humans can see with their naked eye, violet, blue, green, yellow, orange and red. Insects can sense a wider spectrum, outside of the human's capabilities, which range from ultraviolet (350 nm) to red (700 nm) (Stark and Tan 1982). The near infrared spectrum describes the bonds between molecules,

which may indicate some pheromones or the composition of chitin. The wavelengths 1654, 1560 and 1310 nm are known identifiers of beetle's chitin components (Liu et al. 2012). Chitin composes insects elytra, and the wavelengths that closely match are R_{15} (1654), and R_{13} (1560). The correlation matrix can be viewed in appendix E.

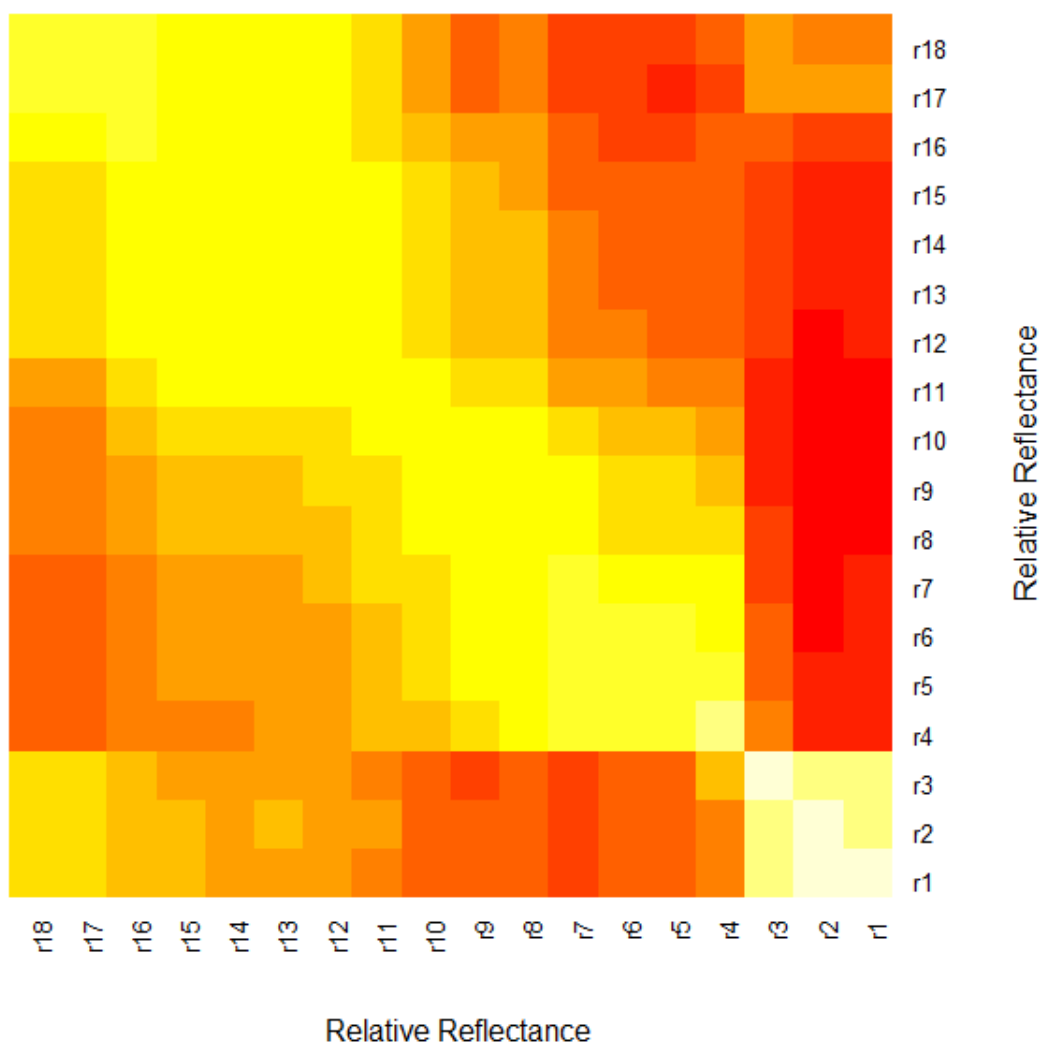


Figure 10. The heat map of the correlation matrix indicating the correlation between peak wavelength values. The wavelengths closely correlated to one another are yellow; while the lower correlation values are red. The color values are assigned based upon their z-score value.

The PCA analyses attempted to reduce the dimensions of the data while separating species and gender. While the genders were not always clearly separated from one another, the

species seem to separate. The classification of species however, required consideration of additional statistical techniques such as multivariate discriminant functions.

Multivariate Discriminate Analysis

Linear discriminant analysis

Linear discriminant analysis was used to classify each species based on the eighteen variables of $(R_1, R_2, R_3, \dots, R_{18})$ and the assumption of multivariate normality. This was completed using equations (4) and (5).

The multivariate normal discriminant analysis resulted in a misclassification rate of 4.14% of individuals incorrectly classified as the wrong species. The majority of the error originated from the comparison of species LC (Lucanidae: *Lucanus capreolus*) to LM (Lucanidae: *Lucanus mazama*), with a 27.27% misclassification rate. This misclassification might be attributed to LC having a small number of observations and the fact that LC and LM are taxonomically very similar. The misclassification between CA (Cerambycidae: *Callidium*) and TE (Buprestidae: *Trachykele blondeli blondeli*) is thought to stem from the similar blue iridescent color they share and the low sample size of TE. The small misclassification rate between PC (Buprestidae: *Buprestis lyrata*) and DI (Buprestidae: *Dicerca tenebrica*) is thought to stem from their very similar elytra. The complete classification results from the multivariate normal discriminant analysis are provided in Table 3.

Table 3. Linear discriminant analysis misclassification results of individual species. The abbreviations represent the following species: Cerambycidae: *Callidium* sp. (CA), Cerambycidae: *Desmocerus piperi* (DE), Buprestidae: *Dicerca tenebrica* (DI), Lucanidae: *Lucanus capreolus* (LC), Lucanidae: *Lucanus mazama* (LM), Buprestidae: *Melanophila atropurpurea* (ME), Buprestidae: *Buprestis lyrata* Casey (PC), Cerambycidae: *Prionus californicus* (PR), Cerambycidae: *Spondylis upiformis* (SP), Trogossitidae: *Temnocheila chlorodia* (TE), Buprestidae: *Trachykele blondeli blondeli* (TR).

Number of Observations and Percent Classified into Species												
From Species	CA	DE	DI	LC	LM	ME	PC	PR	SP	TE	TR	Total
CA	17	0	0	0	0	0	0	0	0	1	0	18
	94.44	0	0	0	0	0	0	0	0	5.56	0	100
DE	0	18	0	0	0	0	0	0	0	0	0	18
	0	100	0	0	0	0	0	0	0	0	0	100
DI	0	0	19	0	0	0	1	0	0	0	0	20
	0	0	95	0	0	0	5	0	0	0	0	100
LC	0	0	0	7	0	0	0	0	0	0	0	7
	0	0	0	100	0	0	0	0	0	0	0	100
LM	0	0	0	6	16	0	0	0	0	0	0	22
	0	0	0	27.27	72.73	0	0	0	0	0	0	100
ME	0	0	0	0	0	18	0	0	0	0	0	18
	0	0	0	0	0	100	0	0	0	0	0	100
PC	0	0	0	0	0	0	24	0	0	0	0	24
	0	0	0	0	0	0	100	0	0	0	0	100
PR	0	0	0	0	0	0	0	27	0	0	0	27
	0	0	0	0	0	0	0	100	0	0	0	100
SP	0	0	0	0	0	0	0	0	19	0	0	19
	0	0	0	0	0	0	0	0	100	0	0	100
TE	2	0	0	0	0	0	0	0	0	24	0	26
	7.69	0	0	0	0	0	0	0	0	92.31	0	100
TR	0	0	0	0	0	0	0	0	0	0	10	10
	0	0	0	0	0	0	0	0	0	0	100	100
Total	19	18	19	13	16	18	25	27	19	25	10	209
	9.09	8.61	9.09	6.22	7.66	8.61	11.96	12.92	9.09	11.96	4.78	100
Priors	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909	

In figure 11, the heat map of the linear discriminant function, the location of the highest (white) and lowest coefficients (red) of the original variables are at R_{11} , R_{12} , R_{13} , & R_{14} . It can be inferred that the majority of the information provided by the

discriminant function comes from these variables, or rather the near-infrared spectrum. One of the variables, R_{13} , contributing a higher loading is considered a wavelength identifying the chitin molecule particularly the amide II of N-H bond (Liu et al. 2012).

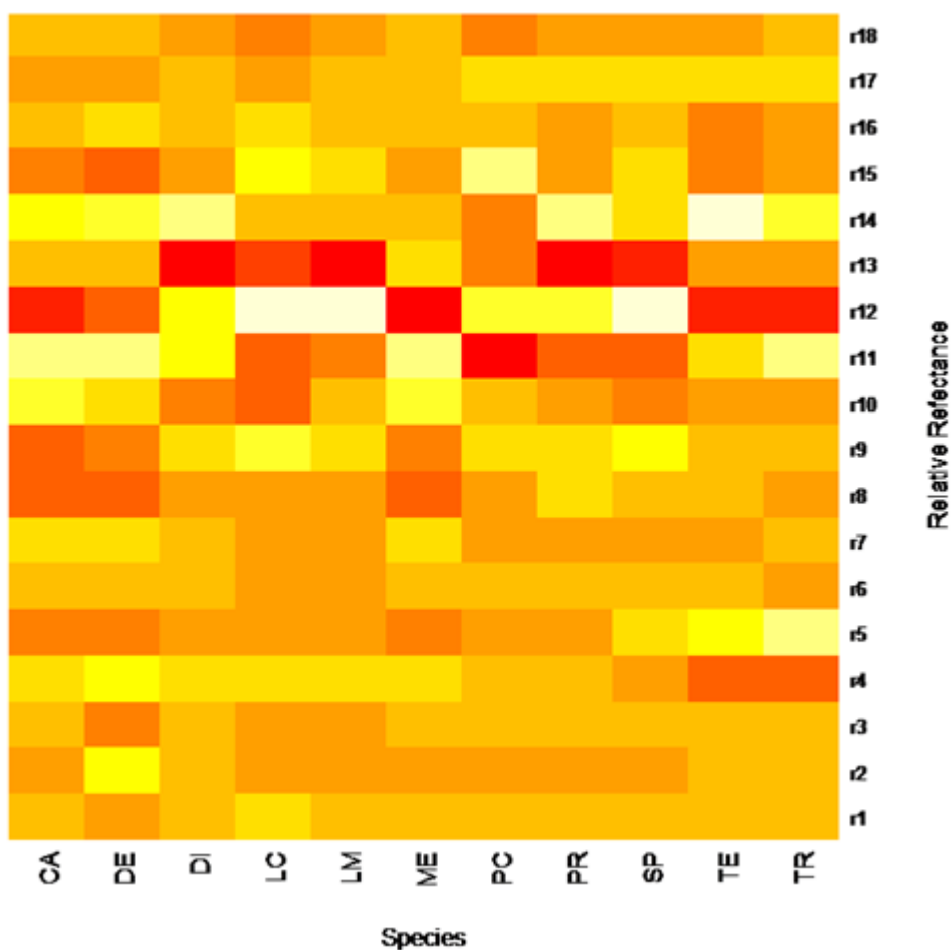


Figure 11. The Heat Map of the Linear Discriminant Function for individual Species. Correlation colors are assigned based upon their z-score value, with low z-score given red and high z-score given white or yellow. The abbreviations represent the following species: Cerambycidae: *Callidium* sp. (CA), Cerambycidae: *Desmocerus piperi* (DE), Buprestidae: *Dicerca tenebrica* (DI), Lucanidae: *Lucanus capreolus* (LC), Lucanidae: *Lucanus mazama* (LM), Buprestidae: *Melanophila atropurpurea* (ME), Buprestidae: *Buprestis lyrata* Casey (PC), Cerambycidae: *Prionus californicus* (PR), Cerambycidae: *Spondylis upiformis* (SP), Trogossitidae: *Temnocheila chlorodia* (TE), Buprestidae: *Trachykele blondeli blondeli* (TR).

The misclassification rate produced by the multivariate linear discriminant analysis is below 0.05, signifying that this model works well as a classification for the data set.

However, the underlying distribution of the wavelengths are often skewed, and hence, the

assumption of multivariate normality may not have been appropriate. Thus, additional statistical approaches were considered in an attempt to relax the assumptions of normality.

Uniform Bayesian Prior Discriminant Analysis

Bayesian priors utilize 'K' nearest neighbor for analysis. K defines the number of nearest points utilized for discriminating the species differences. $K < 3$ was considered too few points and $K > 10$ too many points. The local maxima of misclassification occurred at $K = 6$ and was chosen for subsequent analysis. At $K = 6$, the misclassification rate was 3.8% with the highest rate of misclassification occurring between LC and LM at 27.27%. The total misclassification rate is 3.8% which is below 0.05, signifying that the uniform prior provides a good classification for these data. The tabulated results of the rate of misclassification for the uniform prior discriminant analysis are given in Table 4.

Proportional Bayesian Prior Discriminant Analysis

Given the knowledge of the Coleoptera order, it becomes apparent that the species are not equally abundant. Proportional priors assume that the collections found at the University of Idaho Entomological museum are proportional to species abundance in their habitat. Equations (7) and (9) were utilized for the proportional prior Bayesian discriminant analysis.

The species misclassification rate was calculated using non-parametric K^{th} nearest neighbor at $K = 6$. The value at $K=6$ was chosen for the location of a local maxima, and for consistency with the previous method, the uniform prior. The proportional prior discriminant analysis error rate was 5.2%. While this value is very close to the misclassification values obtained under uniform priors, it is the most accurate given our knowledge about Coleoptera. The species misclassification rates are somewhat consistent with the uniform prior analysis with regard to species CA, LC, LM, and TE. The 'other' species category received several individuals accounting for the highest rate of species misclassification. The results of misclassification rates by species using proportional prior discriminant analysis are given in Table 5.

Table 5. Proportional prior discriminate analysis misclassification of individual species. The abbreviations represent the following species: Cerambycidae: *Callidium* sp. (CA), Cerambycidae: *Desmocerus piperi* (DE), Buprestidae: *Dicerca tenebrica* (DI), Lucanidae: *Lucanus capreolus* (LC), Lucanidae: *Lucanus mazama* (LM), Buprestidae: *Melanophila atropurpurea* (ME), Buprestidae: *Buprestis lyrata* Casey (PC), Cerambycidae: *Prionus californicus* (PR), Cerambycidae: *Spondylis upiformis* (SP), Trogossitidae: *Temnocheila chlorodia* (TE), Buprestidae: *Trachykele blondeli blondeli* (TR).

Number of Observations and Percent Classified into Species													
From Species	CA	DE	DI	LC	LM	ME	PC	PR	SP	TE	TR	Other	Total
CA	18 100	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	18 100
DE	0 0	18 100	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	18 100
DI	0 0	0 0	20 100	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	20 100
LC	0 0	0 0	0 0	1 14.29	1 14.29	0 0	0 0	0 0	0 0	0 0	0 0	5 71.43	7 100
LM	0 0	0 0	0 0	0 0	19 86.36	0 0	0 0	0 0	0 0	0 0	0 0	3 13.64	22 100
ME	0 0	0 0	0 0	0 0	0 0	18 100	0 0	0 0	0 0	0 0	0 0	0 0	18 100
PC	0 0	0 0	0 0	0 0	0 0	0 0	24 100	0 0	0 0	0 0	0 0	0 0	24 100
PR	0 0	0 0	0 0	0 0	0 0	0 0	0 0	27 100	0 0	0 0	0 0	0 0	27 100
SP	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	19 100	0 0	0 0	0 0	19 100
TE	1 3.85	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	24 92.31	0 0	1 3.85	26 100
TR	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	10 100	0 0	10 100
Total	19 9.09	18 8.61	20 9.57	1 0.48	20 9.57	18 8.61	24 11.48	27 12.92	19 9.09	24 11.48	10 4.78	9 4.31	209 100

Validation

Internal Validation

Further exploration into the proportional prior method was completed using a bootstrap simulation for the purpose of model validation. The proportional prior was used because it more accurately described the underlying population. The bootstrap simulation was created using 5,000 separate samples selected with replacement and data splitting. The data was split with 60% of the database used in creating the model, while 40% of the data was used for validation. Each selection generated a species misclassification rate based upon a proportional prior discriminant analysis.

The distribution of misclassification by the proportional prior discriminant analysis bootstrap is given in Figure 12. The distribution can be approximated with a normal curve that has a mean of 0.0348 and a standard deviation of 0.011. The standard deviation is rather low, indicating that a majority of the data is within a small range of the mean. The fifth percentile error is 0.025 and the ninety-fifth percentile is 0.067. The median is located at 0.0341, which indicates that the skewness is low. The low skewness is another indicator that the mean and median agree, and that the normal curve is a reasonable approximation of the data in this case. The range of misclassification is low in value, validating the use of the proportional prior for this data set.

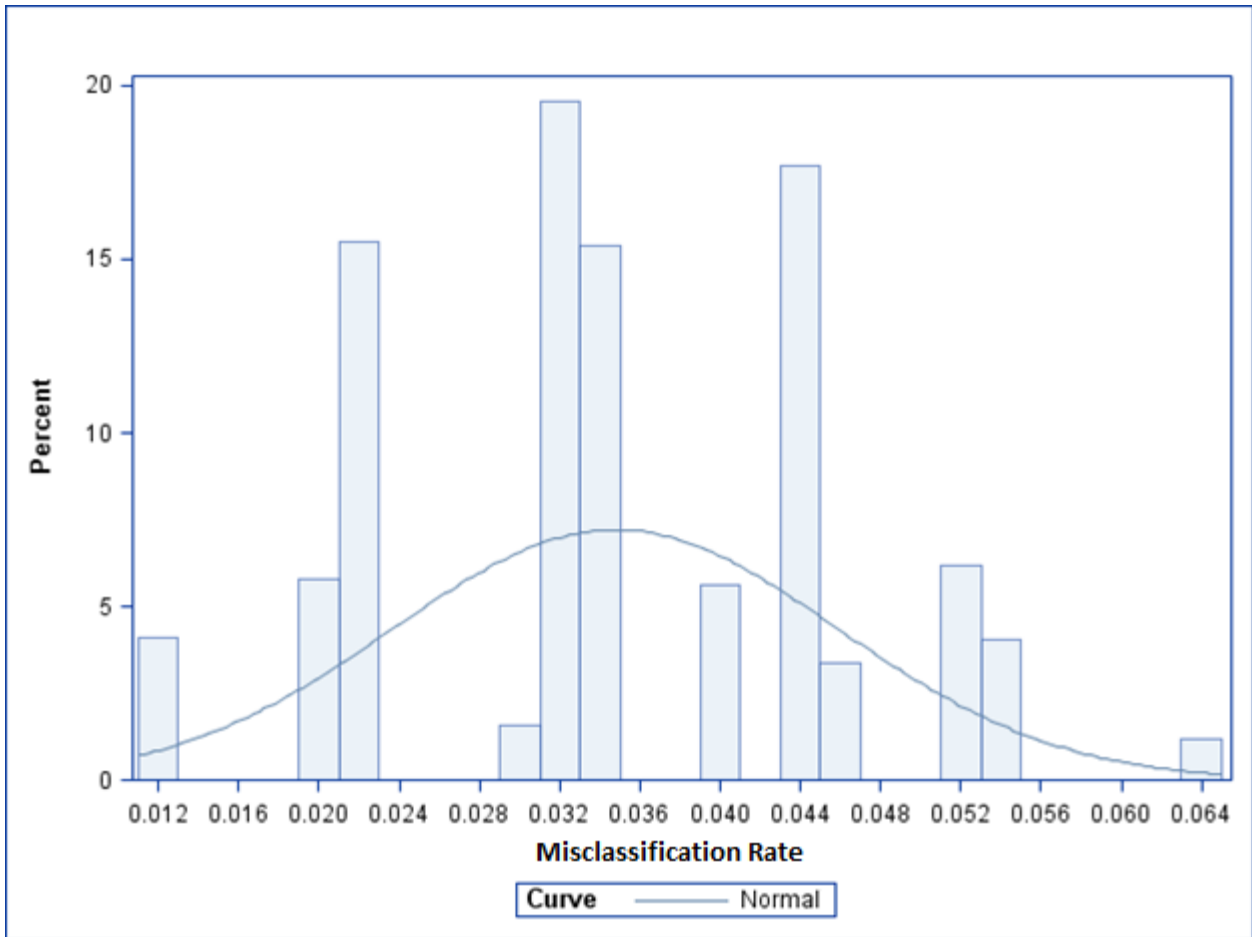


Figure 12. The distribution of species misclassification rate for the internal bootstrap is described using a normal approximation. Species misclassification rate has a mean of 0.0348 and a standard deviation of 0.011.

External Validation

External validation was performed on a new data set independent from the original database. The new data contained 187 insects of the previous taxa groups, not controlled for by location or year collected (Appendix F). The distribution of misclassification had a fifth percentile of 0% and the ninety-fifth percentile of 11.95%. The actual validation bootstrap distribution is given in Figure 13.

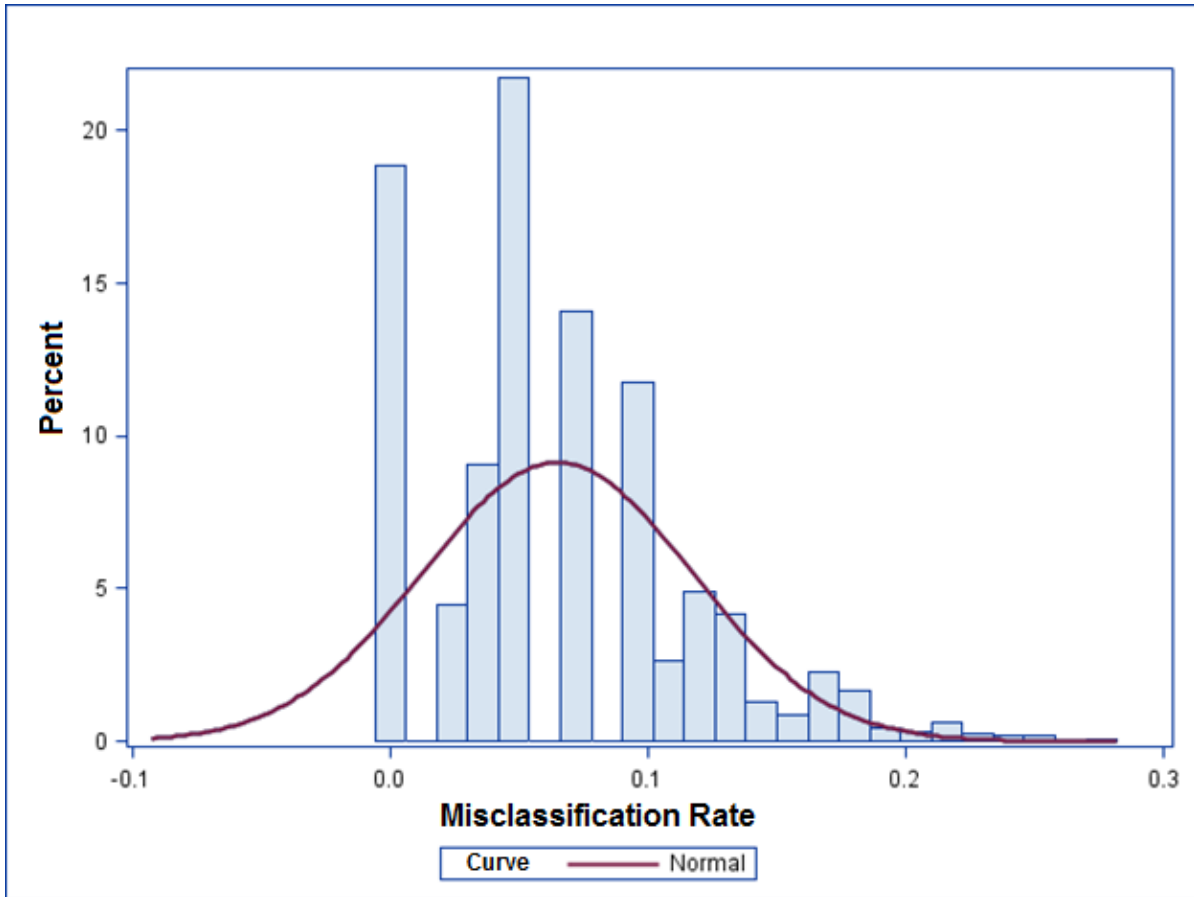


Figure 13. The distribution of species misclassification rate for the external bootstrap is described using a normal approximation. The mean species misclassification rate is 0.0646 or 6.46% and the standard deviation is 0.0278 or 2.7%.

The mean misclassification was 0.0646 and the median misclassification was 0.0455. The amount of skewness was 0.02 which is low in value. The misclassification rate between LC (Lucanidae: *Lucanus capreolus*) and LM (Lucanidae: *Lucanus mazama*) decreased to 14%, which might imply that the misclassification rate is dependent on sample size. The standard deviation is 0.027 which is low in value, so the data are centered near the mean. Overall, given that the specimens were not lab reared, geographically controlled or collected on the same year, the small misclassification rate of the proportional prior discriminant analysis provides an effective way to correctly classify these Coleoptera species.

Chapter 4 Conclusions

Replicated samples of eleven species of wood boring beetles were selected from William Barr Entomology Museum at the University of Idaho for potential differentiation of their taxonomic group and gender based on spectral reflectance readings. The methodology used for correctly identifying Coleoptera species typically relies on morphology of the individual species. In this study, however, spectroscopy on elytra composition of the insects was utilized for the purpose of separation of their species and gender. Specifically, the analyses focused on the visual and near-infrared spectrum to differentiate species and gender. Spectrometer readings generated for each species-gender group were fitted to normal distribution mixture models to identify multiple peak reflectance wavelengths of prominence for further statistical analyses. Principal component and discriminant analyses were subsequently used to assess the differentiation of taxonomic groups and genders based on spectral reflectance. The principal component ordination technique clearly grouped Coleoptera by taxonomic groups, while the linear discriminant analysis, under an assumption of multivariate normality, provided a distinct classification of taxonomic groups and provided a low rate of misclassification error. The assumption of normality was subsequently relaxed using a nonparametric nearest neighbor discriminant analysis, which resulted in very highly accurate classification of Coleoptera species. Further internal and external validation of the nearest neighbor discriminant model confirmed the results of low species misclassification error rates.

Given the low error rates of misclassification, the multivariate statistical approaches outlined in this study are recommended for analysis of spectral reflectance in Coleoptera and

other similar insect groups. However, it is noted that further research in this area should consider using a larger number of individual insects, as well as increasing the number of species analyzed. Also, extrapolation of results has to be practiced cautiously due to varying sensitivity of spectroscopy equipment. If practically feasible, utilizing insects from multiple museums is highly recommended. Incorporation of other Coleoptera attributes such as developmental stage, length, pheromones present, location and collection date might further improve the resolution of the classification techniques. Finally, it is recommended to obtain additional spectrometer readings in the ultraviolet spectrum because while insects are able to see in that spectrum, it may contain markings that are invisible to the human eye.

References

- Barber, Michael, J. Cuthbert, J. Farren, J.W. Linnett. "A Mass Spectrometer for Studying Chemical Reactions." *Proceedings of The Royal Society A Mathematical, Physical & Engineering Sciences*. 274. (1963): 285-292.
- Choate, Paul. "Introduction to the Identification of Beetles (Coleoptera)." *Dichotomous Keys to Some Families of Florida Coleoptera*. (1999): 23-33.
- Coleman, Tom. N.E. Grulke, M. Daly, C. Godinez, S.L. Schilling, P.J. Riggan, S.J. Seybold. "Coast live oak, *Quercus agrifolia*, susceptibility and response to goldspotted oak borer, *Agrilus auroguttatus*, injury in southern California." *Forest Ecology and Management*. 261. (2011): 1862-1865.
- Cook, Stephen, K.S. Humes, R. Hruska, G. Fraley, C.J. Williams. "Pre-visual Detection of Two Conifer-Infesting Adelgid Species in North America Forests." *International Journal of Forestry Research*. 2010. (2010): 1-8.
- Daly, Howell, and S.S. Balling. "Identification of Africanized Honeybees in the Western Hemisphere by Discriminant Analysis." *Journal of the Kansas Entomological Society*. 51.4 (1978): 857-869.
- Dowell, Floyd, J.E. Throne, D. Wang, and J.E. Baker. "Identifying Stored-Grain Insects Using Near-Infrared Spectroscopy." *Journal of Economic Entomology*. 92.1 (1999): 165-169.
- Efron, Bradley. "The 1977 Rietz Lecture Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics*. 7.1 (1979): 1-26.
- Endler, John, D.A. Westcott, J.R. Madden, and T. Robson. "Animal Visual Systems and the Evolution of Color Patterns: Sensory Processing Illuminates Signal Evolution." *Society for the Study of Evolution*. 59.8 (2005): 1795-1818.
- Hall, Robert. "Use of a Stable Carbon Isotope Addition to Trace Bacterial Carbon through a Stream Food Web." *Journal of the North American Benthological Society*, Vol. 14, No. 2 (Jun., 1995): 269-277.
- Hinich, Melvin. "Dichotomous Variable Regression Coefficients and Discriminant Weights." *Political Methodology*. 6.1 (1979): 5-9.
- James, Frances, and C.E. McCulloch. "Multivariate Analysis in Ecology and Systematics: Panacea or Pandora's Box?" *Ecologic Systems*. 21. (1990): 129-166.

- Kawakami, Yasuko, K. Yamazaki, and K. Ohashi. "Geographical variations of elytral color polymorphism in *Cheilomenes sexmaculata* (Fabricius) (Coleoptera: Coccinellidae)." *Entomological Science*. 16. (2013): 235-242.
- Kruskal, Joseph. "Nonmetric Multidimensional Scaling: A Numeric Method." *Psychometrika*. 29.2 (1964): 115-129.
- Lachenbruch, Peter, and M. Goldstein. "Discriminant Analysis." *Biometrics*. 35.1 (1979): 69-85.
- Lillie, Charles, R.C. Bohlin, M.R. Molnar, C.A. Barth, A.L. Lane. "Mariner 9 Ultraviolet Spectrometer Experiment: Stellar Observations." *Science*, New Series, Vol. 175, No. 4019 (1972): 321-322.
- Liu, Shaofang, J. Sun, L. Yu, C. Zhang, J. Bi, F. Zhu, M. Qu, C. Jiang, Q. Yang. "Extraction and Characterization of Chitin from the Beetle *Holotrichia parallela* Motschulsky." *Molecules*. 17. (2012): 4604-4611.
- Morrison, Donald. *Multivariate Statistical Methods*. 3rd Ed. New York: McGraw Hill Inc.
- Piszter, Gabor, K. Kertesz, Vertesy, Z. Zofia, Laszlo, Z. Balint, L.P. Biro. "Color based discrimination of chitin–air nanocomposites in butterfly scales and their role in conspecific recognition." *Hungarian Natural History Museum*. (2010): 3(1). 78-83.
- Rao, Radhakrishna. "The Use and Interpretation of Principal Component Analysis in Applied Research." *Indian Journal of Statistics*. 26.4 (1964): 329-358.
- Rousu, Juho, D.D. Agranoff, O. Sodeinde, J. Shawe-Taylor, D. Fernandez-Reyes. (2013) "Biomarker Discovery by Sparse Canonical Correlation Analysis of Complex Clinical Phenotypes of Tuberculosis and Malaria." *PLoS Comput Biol* 9(4): 1-10.
- Royle, Andrew, and W.A. Link. "A General Class of Multinomial Mixture Models for Anuran Calling Survey Data." *Ecology*. 86.9 (2005): 2505-2512.
- SAS online documentation 9.3 Copyright © 2012 SAS Institute Inc., Cary, NC, USA.
- Seago, Ainsley, B. Parrish, J. Vigneron, and T.D. Schultz. "Gold bugs and beyond: a review of iridescence and structural colour mechanisms in beetles (Coleoptera)." *Journal of The Royal Society Interface*. 6. (2009): S165-S184.
- Shafii, Bahman, W. J. Price, C. Holderman, C. Gidley, and P. J. Anders. 2010. Modeling fish length distribution using a mixture technique. *Applied Statistics in Agriculture*, W. Song and G. L. Gadbury (Eds). Kansas State University, Manhattan, Kansas, pp. 2-11.

- Shafii, Bahman, W.J. Price, T.S. Prather, L.L., Lass, and D.C. Thill. "Using landscape characteristics as prior information for Bayesian classification of yellow starthistle." *Weed Science*. 52. (2004): 948-953.
- Shepard, Roger. "The Analysis of Proximities Multidimensional Scaling with an Unknown Distance Function." *Psychometrika*. 27. (1962): 125-139.
- Stark, Williams, and K.P. Tan. "Ultraviolet light: photosensitivity and other effects on the visual system." *Photochem. Photobiol.* (1982): 371–380.
- Thorpe, Roger. "Analysis of Color Spectra in Comparative Evolutionary Studies: Molecular Phylogeny and Habitat Adaptation in the St. Vincent Anole (*Anolis trinitatis*)."
Systematic Biology, Vol. 51, No. 4 (2002): 554-569.
- Taylor, Viven, H.P. Longerrich, and J. Greenough. "Multielement Analysis of Canadian Wines by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) and Multivariate Statistics." *J. Agric. Food Chem.* 51.4 (2003): 856–860.
- Vigeneron, Jean Pol, M. Rassart, C. Vandembem, V. Lousse, O. Deparis, L.P. Biró, László. D. Dedouaire, A. Cornet, P. Defrance. "Spectral filtering of visible light by the cuticle of metallic woodboring beetles and microfabrication of a matching bioinspired material," *Physical Review E*, 73, No. 4 (2006): 1-8.

Appendix A

A subset (first 20 individual insects) with R1-R18 variables created from the FMM using the original data

Table A1. The first 20 observations by individual insect of relative reflectance values at peak values of R1-R18 selected by the FMM. The abbreviation used in the subset data represent the following species: Cerambycidae: *Callidium* sp. (CA), Cerambycidae: *Desmocerus piperi* (DE).

Species	Sex	Rep	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18
CA	M	CA11	0.21	0.19	0.19	0.25	0.29	0.31	0.37	0.43	0.46	0.47	0.45	0.43	0.43	0.45	0.46	0.43	0.37	0.31
CA	M	CA12	0.22	0.19	0.18	0.29	0.37	0.42	0.53	0.63	0.70	0.71	0.69	0.67	0.68	0.69	0.70	0.67	0.58	0.52
CA	M	CA13	0.18	0.17	0.18	0.22	0.25	0.27	0.26	0.31	0.35	0.35	0.34	0.31	0.31	0.33	0.33	0.35	0.29	0.24
CA	M	CA14	0.15	0.13	0.14	0.18	0.20	0.22	0.25	0.31	0.34	0.35	0.34	0.33	0.33	0.34	0.35	0.33	0.29	0.25
CA	M	CA15	0.20	0.19	0.18	0.25	0.30	0.33	0.37	0.45	0.51	0.53	0.51	0.49	0.49	0.51	0.51	0.46	0.41	0.36
CA	M	CA16	0.09	0.07	0.07	0.13	0.16	0.19	0.26	0.33	0.39	0.40	0.40	0.39	0.39	0.41	0.41	0.39	0.33	0.29
CA	M	CA17	0.07	0.06	0.06	0.10	0.13	0.16	0.21	0.27	0.30	0.31	0.31	0.29	0.30	0.31	0.31	0.29	0.25	0.21
CA	M	CA18	0.08	0.07	0.07	0.12	0.16	0.19	0.22	0.28	0.33	0.34	0.32	0.30	0.30	0.31	0.31	0.28	0.22	0.18
CA	F	CA21	0.13	0.10	0.09	0.17	0.22	0.26	0.33	0.40	0.45	0.46	0.45	0.43	0.43	0.44	0.44	0.49	0.41	0.38
CA	F	CA22	0.13	0.11	0.09	0.22	0.30	0.35	0.46	0.57	0.65	0.66	0.64	0.61	0.62	0.63	0.63	0.58	0.49	0.46
CA	F	CA23	0.12	0.10	0.11	0.20	0.26	0.30	0.39	0.45	0.52	0.52	0.48	0.46	0.45	0.46	0.46	0.45	0.37	0.34
CA	F	CA24	0.12	0.10	0.09	0.19	0.27	0.33	0.42	0.53	0.61	0.63	0.60	0.57	0.57	0.58	0.58	0.51	0.41	0.37
CA	F	CA25	0.22	0.21	0.22	0.35	0.43	0.49	0.63	0.66	0.72	0.73	0.69	0.66	0.66	0.68	0.69	0.67	0.57	0.51
CA	F	CA26	0.15	0.11	0.10	0.24	0.34	0.40	0.53	0.62	0.71	0.72	0.68	0.65	0.65	0.66	0.66	0.62	0.52	0.48
CA	F	CA27	0.14	0.12	0.11	0.23	0.34	0.41	0.57	0.67	0.79	0.81	0.76	0.73	0.73	0.75	0.74	0.69	0.58	0.53
CA	F	CA28	0.20	0.15	0.18	0.33	0.43	0.50	0.59	0.74	0.84	0.87	0.84	0.80	0.80	0.81	0.82	0.76	0.62	0.56
CA	F	CA29	0.16	0.13	0.14	0.23	0.29	0.33	0.41	0.46	0.50	0.51	0.49	0.46	0.46	0.46	0.46	0.49	0.44	0.42
CA	F	CA10	0.06	0.05	0.04	0.13	0.19	0.24	0.32	0.39	0.47	0.48	0.45	0.43	0.42	0.43	0.42	0.37	0.31	0.27
DE	F	DE21	0.05	0.04	0.04	0.09	0.14	0.18	0.26	0.35	0.43	0.44	0.42	0.40	0.39	0.41	0.40	0.36	0.28	0.22
DE	F	DE22	0.07	0.08	0.06	0.16	0.24	0.30	0.41	0.54	0.63	0.65	0.62	0.57	0.57	0.60	0.58	0.53	0.42	0.35
DE	F	DE23	0.08	0.09	0.07	0.14	0.23	0.29	0.43	0.57	0.69	0.71	0.67	0.62	0.62	0.64	0.61	0.53	0.41	0.33
DE	F	DE24	0.07	0.07	0.05	0.13	0.22	0.28	0.40	0.53	0.64	0.66	0.62	0.58	0.57	0.59	0.57	0.48	0.35	0.29
DE	F	DE25	0.06	0.06	0.05	0.13	0.22	0.28	0.41	0.55	0.65	0.66	0.62	0.57	0.57	0.59	0.58	0.51	0.38	0.31
DE	F	DE26	0.11	0.10	0.08	0.19	0.28	0.35	0.48	0.63	0.74	0.76	0.73	0.68	0.68	0.71	0.70	0.62	0.49	0.39
DE	F	DE27	0.07	0.07	0.05	0.12	0.19	0.25	0.36	0.47	0.60	0.61	0.56	0.53	0.53	0.54	0.52	0.43	0.34	0.29

Appendix B

Multidimensional Scaling (MDS)

Multidimensional scaling was used to find potential relationships between the variables in order to make distinctions between the taxa groups. The data on which MDS operates are the elements of the dissimilarity matrix among all pairs of objects using the Euclidian distance matrix. The Euclidian distance matrix is an n by n matrix with distances between each subject. Let D_{ij} be the dissimilarity between objects i and j , and let δ_{ij} be the Euclidean distance between objects i and j in the ordination space. The objective is to produce an ordination such that:

$$D_{ij} < D_{kl} \Rightarrow \delta_{ij} \leq \delta_{kl} \text{ for all } i, j, k, l$$

If any given pair of objects has dissimilarity less than some other pair, then the first pair should be no further apart in the ordination than the second pair (Shepard 1962). A scatter plot of ordination distances, against dissimilarities, is known as a Shepard diagram and can be viewed in figure B1. Badness of fit (Kruskal's Stress Test), S , is defined as

$$S = \sqrt{\frac{\sum_{i=2}^n \sum_{j=1}^{i-1} (\delta_{ij} - \hat{\delta}_{ij})^2}{\sum_{i=2}^n \sum_{j=1}^{i-1} \delta_{ij}^2}} \quad (10)$$

Stress decreases as the rank-order agreement between distances and dissimilarities improves (Kruskal 1964). The aim is therefore to find the ordination with the lowest possible stress.

In figure 14, two dimensions explain 90% of the data, and in figure 15 the two dimensions are graphed using Euclidian distance to create separation based on ranks. In figure 15, the separation generated between species and gender created by MDS was not deemed adequate for the purpose of this study.

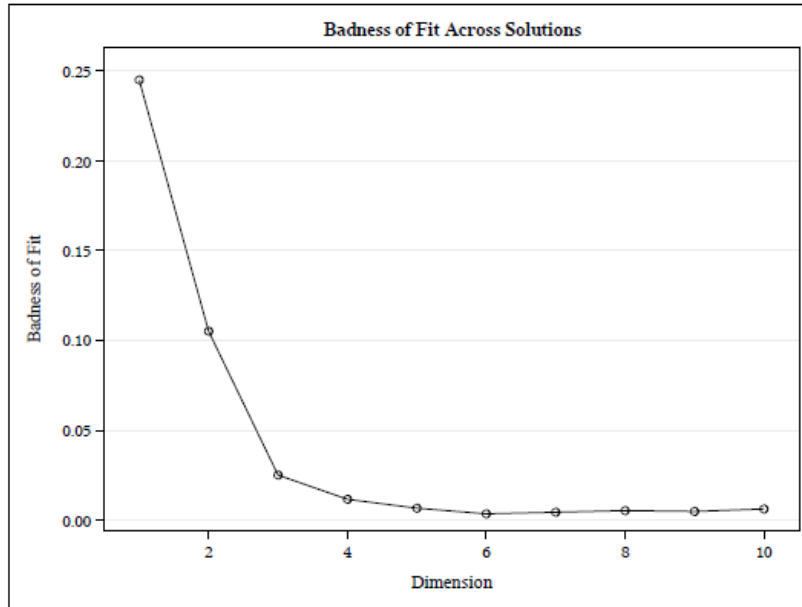


Figure B1. Shepard diagram badness of fit by dimension; the first dimension has a badness of fit of 24%, two dimensions 11% and three dimensions 2.5%.

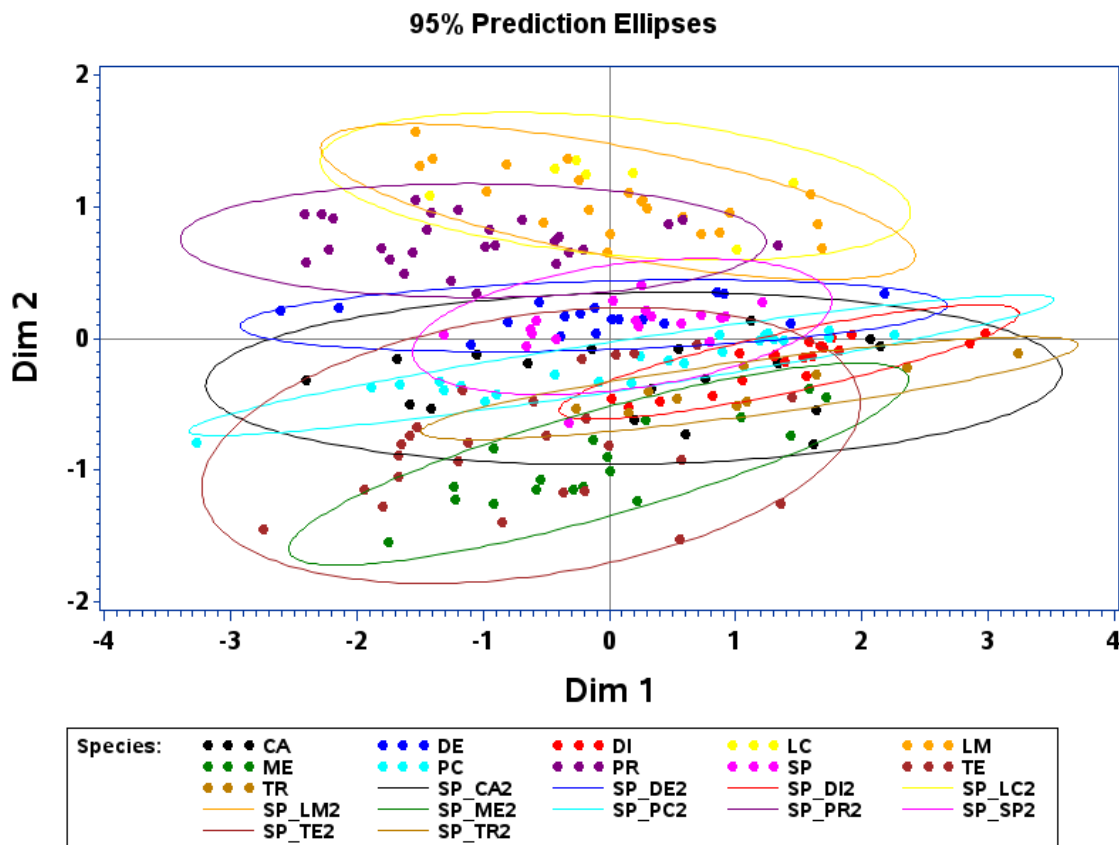


Figure B2. The 95% prediction ellipse displays MDS dimension one vs MDS dimension two. Euclidian distance was used to create species separation based on rank distance. However, as seen in this figure the species are not differentiating. The abbreviations represent the following species: Cerambycidae: *Callidium* sp. (CA, SP_CA2), Cerambycidae: *Desmocerus piperi* (DE, SP_DE2), Buprestidae: *Dicerca tenebrica* (DI, SP_DI2), Lucanidae: *Lucanus capreolus* (LC, SP_LC2), Lucanidae: *Lucanus mazama* (LM, SP_LM2), Buprestidae: *Melanophila atropurpurea* (ME, SP_ME2), Buprestidae: *Buprestis lyrata* Casey (PC, SP_PC2), Cerambycidae: *Prionus californicus* (PR, SP_PR2), Cerambycidae: *Spondylis upiformis* (SP, SP_SP2), Trogossitidae: *Temnocheila chlorodia* (TE, SP_TE2), Buprestidae: *Trachykele blondeli blondeli* (TR, SP_TR2).

Appendix C

A subset (first 20 individual insects) with R1-R18 variables created from the FMM using the external Validation data

Table B1. The first 20 observations by individual insect of relative reflectance values at peak values of R1-R18 selected by the FMM from the validation database. The abbreviation used in the subset data represent the following species: Cerambycidae: *Callidium* sp. (CA).

Spec	Sex	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18
CA	1	0.02	0.04	0.07	0.12	0.06	0.03	0.03	0.10	0.15	0.13	0.18	0.17	0.16	0.14	0.12	0.09	0.09	0.06
CA	1	0.01	0.01	0.04	0.11	0.07	0.04	0.03	0.13	0.19	0.17	0.23	0.23	0.20	0.18	0.15	0.12	0.11	0.09
CA	1	0.02	0.04	0.07	0.14	0.08	0.04	0.03	0.12	0.17	0.15	0.20	0.20	0.18	0.16	0.14	0.11	0.10	0.07
CA	1	0.01	0.03	0.06	0.11	0.06	0.03	0.02	0.09	0.14	0.13	0.17	0.17	0.15	0.13	0.11	0.09	0.08	0.06
CA	1	0.01	0.01	0.02	0.06	0.04	0.02	0.02	0.08	0.12	0.11	0.14	0.14	0.13	0.12	0.10	0.07	0.06	0.05
CA	1	0.02	0.06	0.11	0.19	0.10	0.05	0.04	0.14	0.20	0.18	0.23	0.22	0.20	0.18	0.15	0.12	0.12	0.08
CA	1	0.02	0.04	0.08	0.14	0.08	0.04	0.03	0.11	0.15	0.14	0.18	0.18	0.16	0.14	0.13	0.11	0.10	0.08
CA	1	0.01	0.03	0.05	0.11	0.06	0.03	0.02	0.10	0.15	0.14	0.18	0.18	0.16	0.14	0.12	0.09	0.08	0.06
CA	1	0.02	0.04	0.07	0.13	0.08	0.04	0.03	0.12	0.17	0.16	0.20	0.20	0.18	0.16	0.14	0.11	0.10	0.07
CA	1	0.01	0.01	0.02	0.09	0.06	0.03	0.03	0.10	0.15	0.14	0.18	0.18	0.16	0.15	0.13	0.10	0.09	0.07
CA	1	0.01	0.03	0.05	0.10	0.06	0.04	0.03	0.12	0.18	0.16	0.21	0.21	0.19	0.17	0.15	0.12	0.11	0.08
CA	1	0.02	0.06	0.10	0.16	0.09	0.05	0.03	0.12	0.17	0.15	0.20	0.20	0.18	0.16	0.14	0.10	0.10	0.07
CA	1	0.01	0.02	0.04	0.09	0.06	0.03	0.02	0.09	0.14	0.13	0.17	0.17	0.15	0.14	0.12	0.09	0.08	0.06
CA	2	0.01	0.02	0.03	0.09	0.06	0.04	0.03	0.11	0.18	0.16	0.20	0.20	0.18	0.16	0.13	0.10	0.09	0.07
CA	2	0.02	0.04	0.06	0.12	0.07	0.04	0.03	0.10	0.15	0.13	0.18	0.17	0.16	0.14	0.12	0.10	0.09	0.07
CA	2	0.01	0.03	0.05	0.10	0.06	0.03	0.03	0.10	0.15	0.13	0.17	0.17	0.15	0.13	0.11	0.08	0.07	0.05
CA	2	0.01	0.03	0.05	0.13	0.08	0.04	0.03	0.13	0.20	0.18	0.23	0.23	0.20	0.18	0.15	0.12	0.11	0.08
CA	2	0.02	0.04	0.07	0.13	0.07	0.04	0.03	0.11	0.15	0.14	0.18	0.17	0.15	0.14	0.12	0.09	0.08	0.06
CA	2	0.01	0.02	0.02	0.07	0.05	0.03	0.02	0.09	0.13	0.12	0.16	0.16	0.14	0.13	0.11	0.09	0.09	0.07
CA	2	0.01	0.03	0.05	0.13	0.08	0.04	0.03	0.13	0.19	0.17	0.23	0.23	0.21	0.18	0.16	0.12	0.11	0.09
CA	2	0.02	0.04	0.07	0.13	0.07	0.04	0.03	0.11	0.16	0.14	0.18	0.18	0.16	0.14	0.12	0.09	0.08	0.06
CA	2	0.01	0.03	0.05	0.13	0.08	0.04	0.03	0.12	0.19	0.17	0.22	0.22	0.19	0.17	0.15	0.12	0.10	0.08
CA	2	0.01	0.02	0.04	0.08	0.05	0.03	0.02	0.08	0.13	0.12	0.15	0.15	0.14	0.12	0.10	0.07	0.06	0.05
CA	2	0.01	0.02	0.03	0.09	0.06	0.03	0.03	0.10	0.17	0.15	0.19	0.19	0.16	0.15	0.12	0.09	0.08	0.06
CA	2	0.02	0.04	0.08	0.13	0.08	0.04	0.03	0.11	0.17	0.15	0.20	0.19	0.17	0.15	0.13	0.09	0.08	0.06

Appendix D

Finite Mixture Model distributions

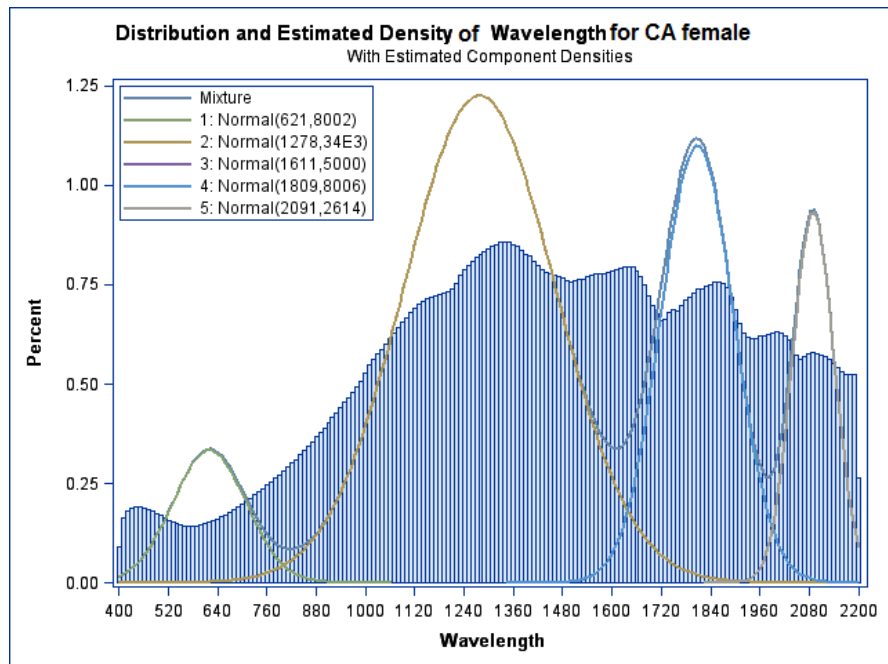


Figure D1. Finite mixture model for Cerambycidae: *Callidium* sp. (CA) female with four normal curves.

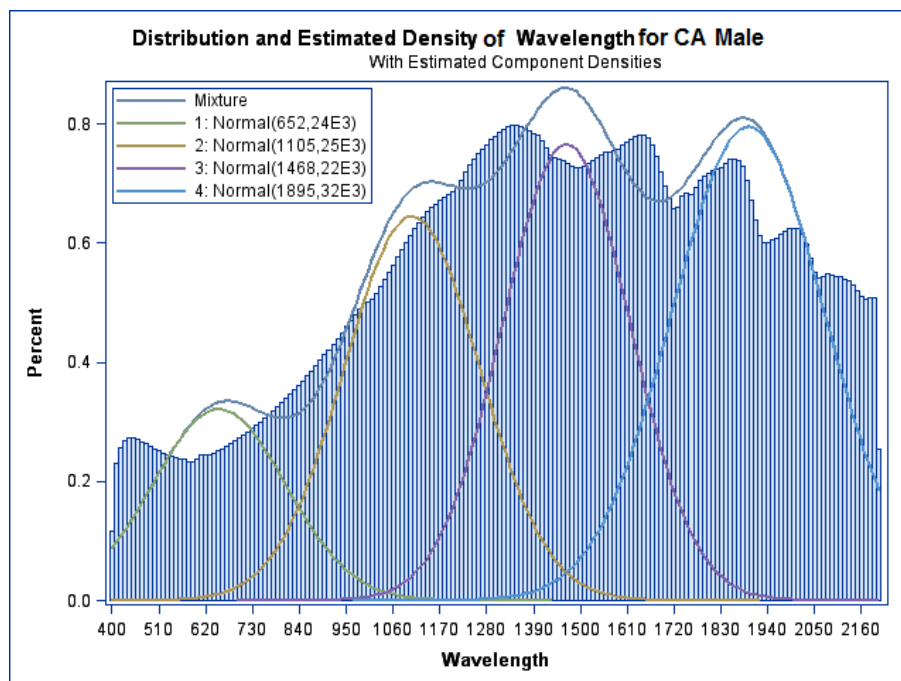


Figure D2. Finite mixture model for Cerambycidae: *Callidium* sp. (CA) male with four normal curves.

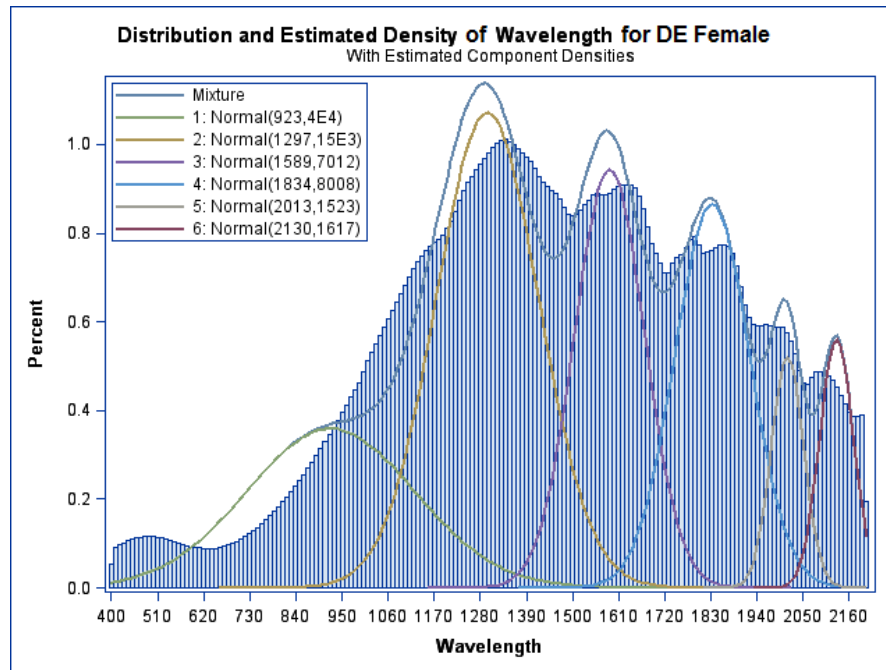


Figure D3. Finite mixture model Cerambycidae: *Desmocerus piperi* (DE) female with six normal curves.

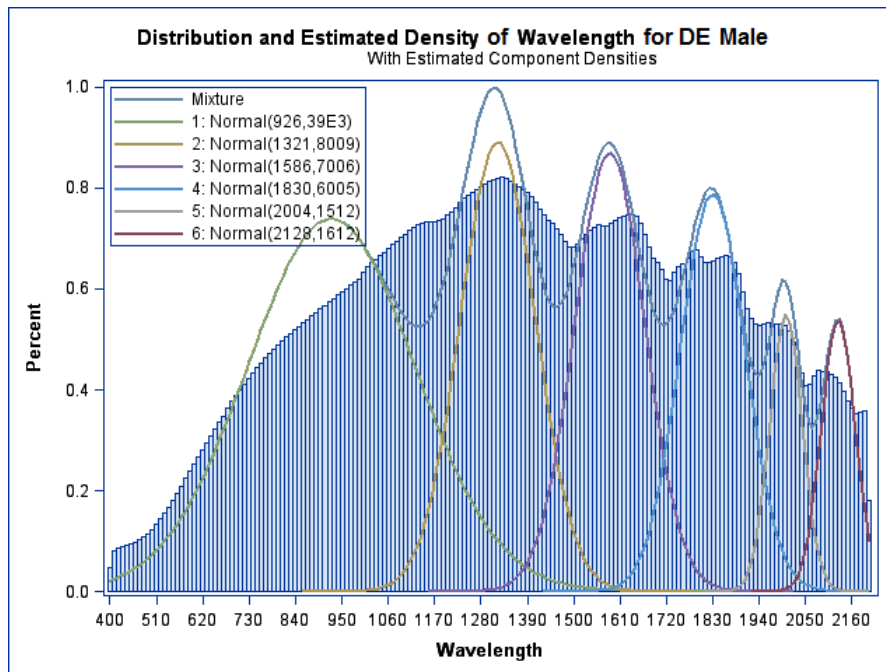


Figure D4. Finite mixture model for Cerambycidae: *Desmocerus piperi* (DE) male with six normal curves.

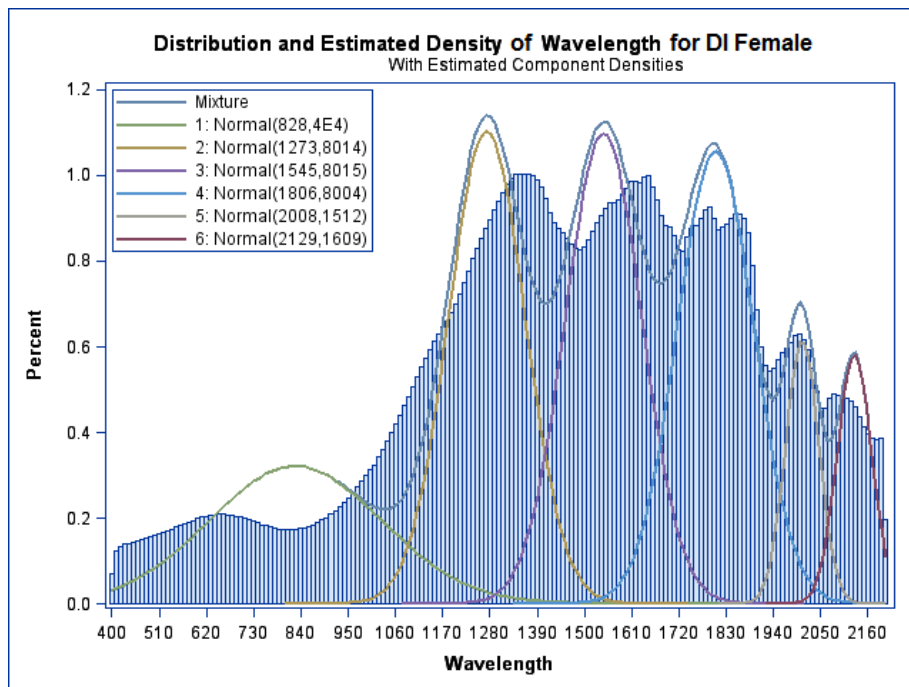


Figure D5. Finite mixture model for Buprestidae: *Dicerca tenebrica* (DI) female with six normal curves.

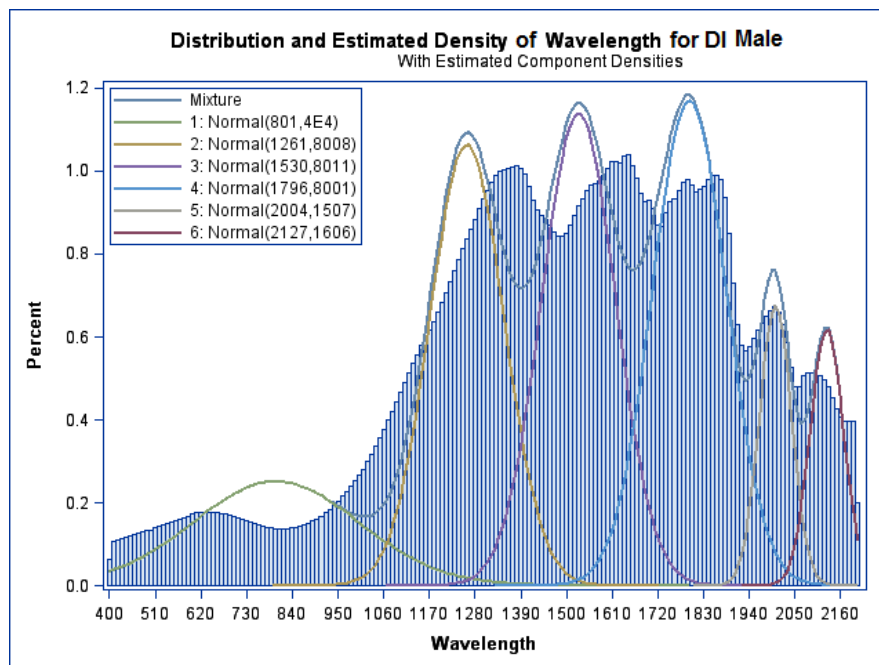


Figure D6. Finite mixture model for Buprestidae: *Dicerca tenebrica* (DI) male with six normal curves.

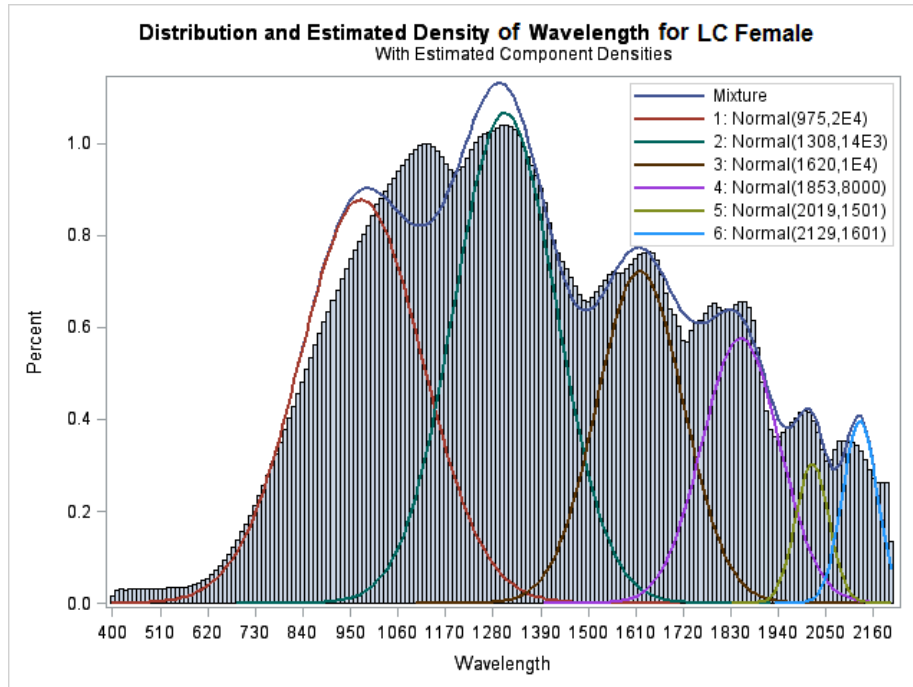


Figure D7. Finite mixture model for Lucanidae: *Lucanus capreolus* (LC) female with six normal curves.

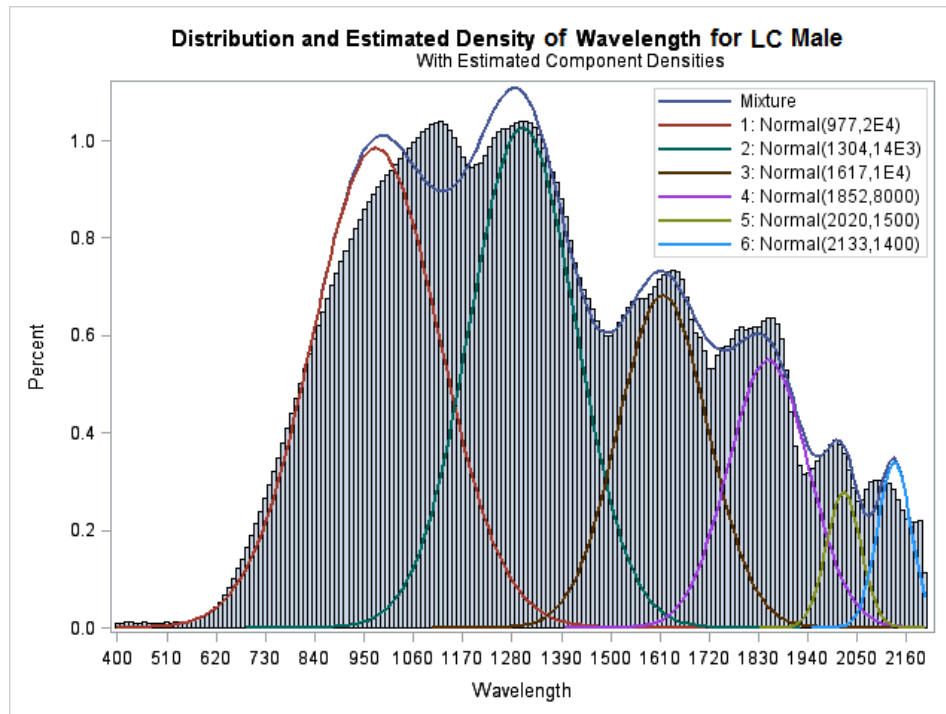


Figure D8. Finite mixture model for Lucanidae: *Lucanus capreolus* (LC) male with six normal curves.

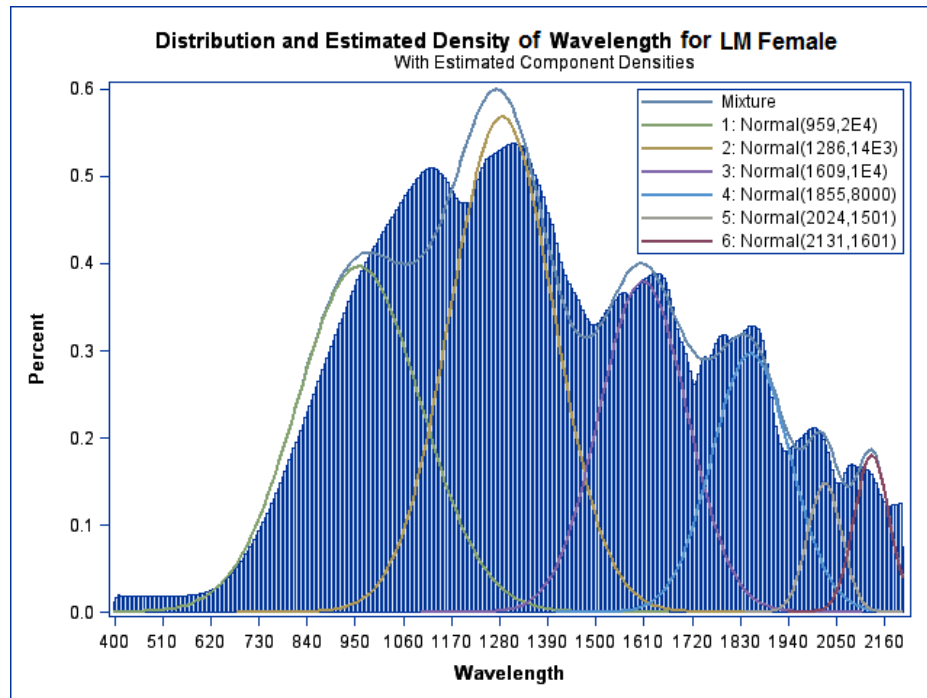


Figure D9. Finite mixture model for Lucanidae: *Lucanus mazama* (LM) female with six normal curves.

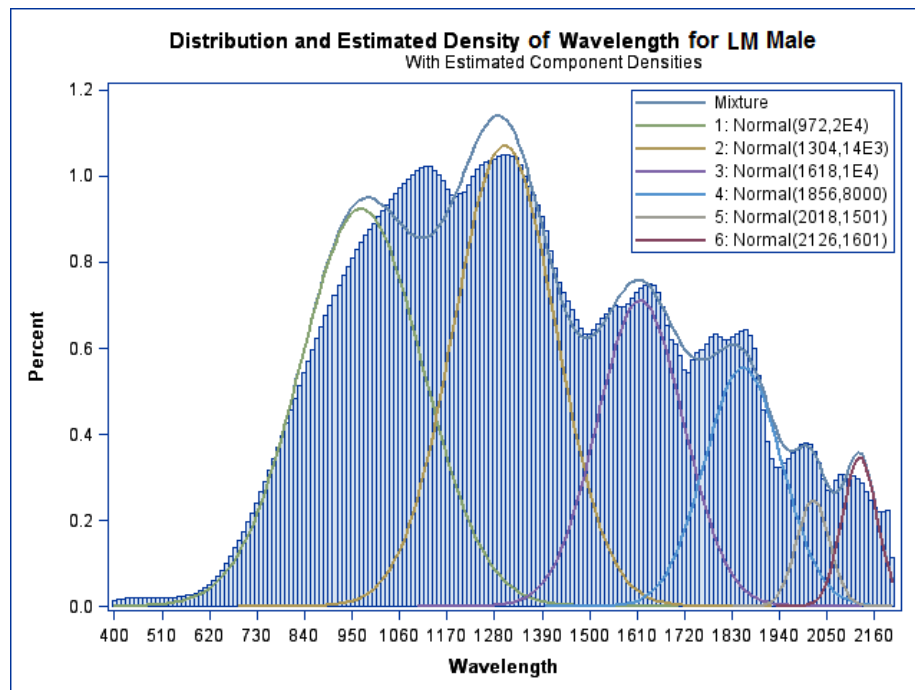


Figure D10. Finite mixture model Lucanidae: *Lucanus mazama* (LM) male with six normal curves.

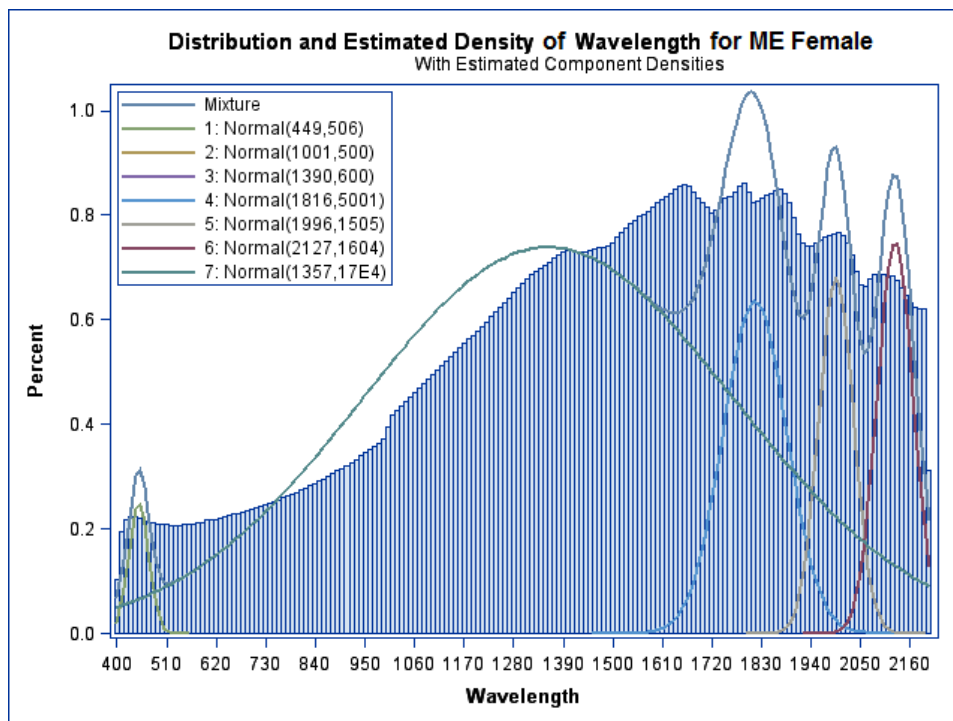


Figure D11. Finite mixture model for Buprestidae: *Melanophila atropurpurea* (ME) female with five normal curves.

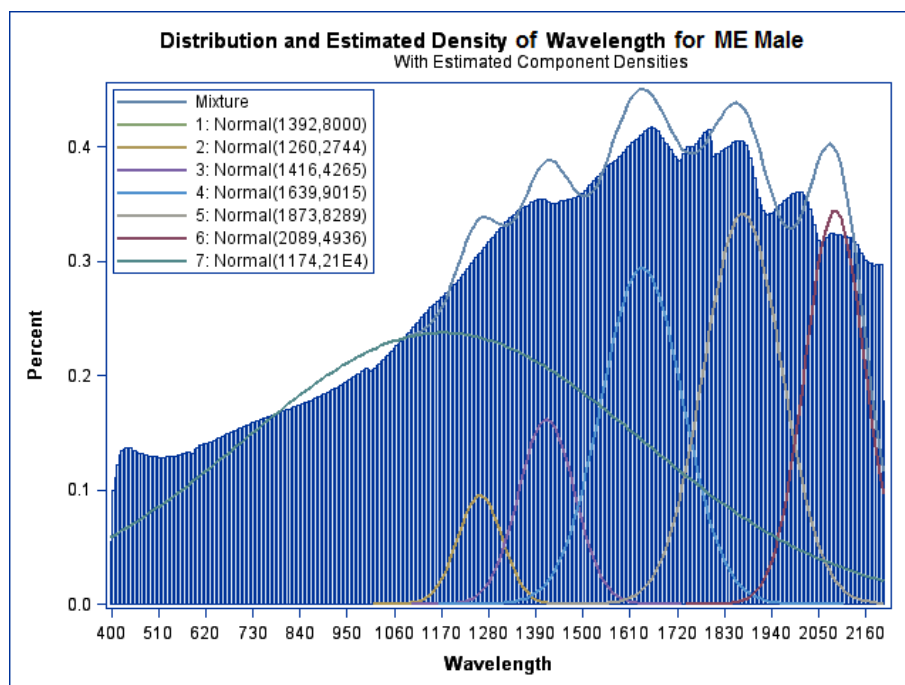


Figure D12. Finite mixture model for Buprestidae: *Melanophila atropurpurea* (ME) male with six normal curves.

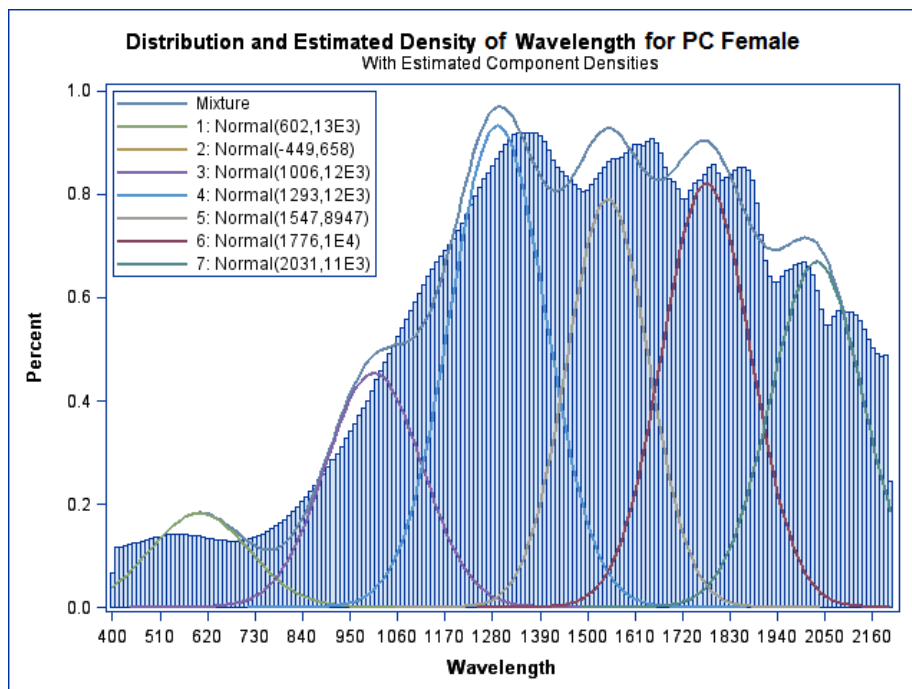


Figure D13. Finite mixture model for Buprestidae: *Buprestis lyrata* Casey (PC) female with six normal curves.

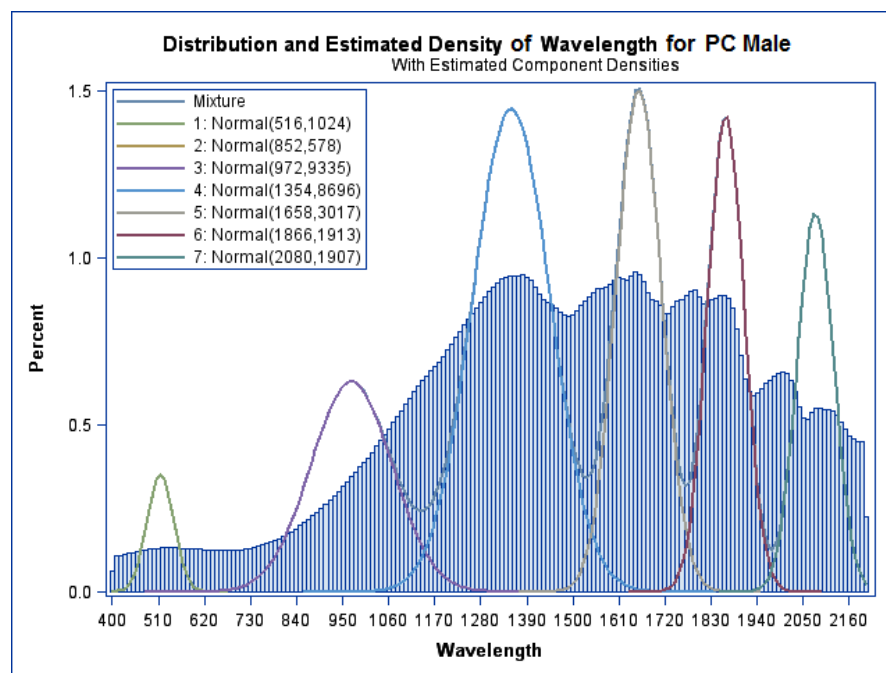


Figure D14. Finite mixture model for Buprestidae: *Buprestis lyrata* Casey (PC) male with six normal curves.

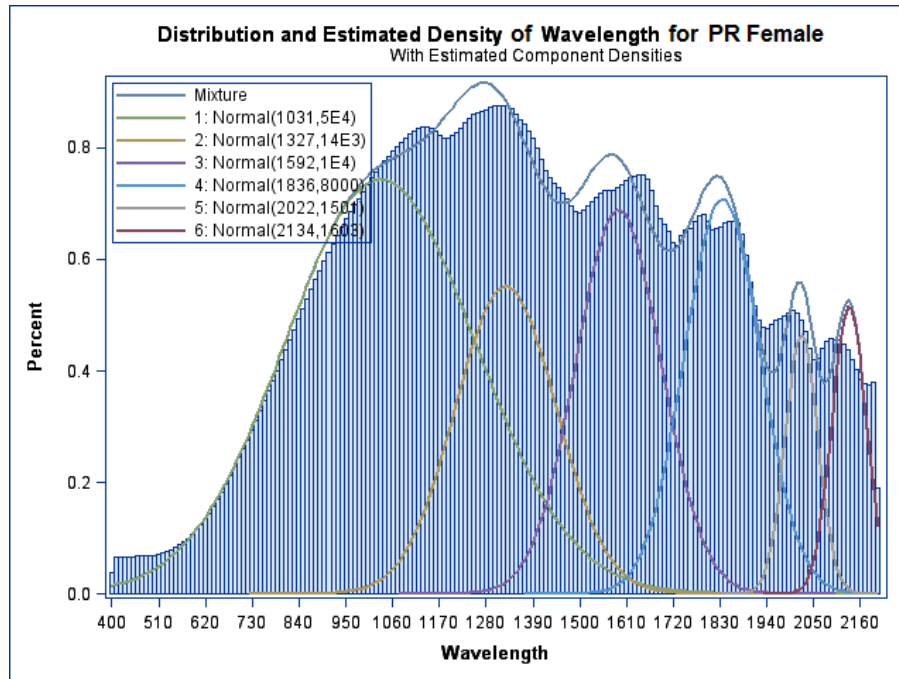


Figure D15. Finite mixture model for Cerambycidae: *Prionus californicus* (PR) female with six normal curves.

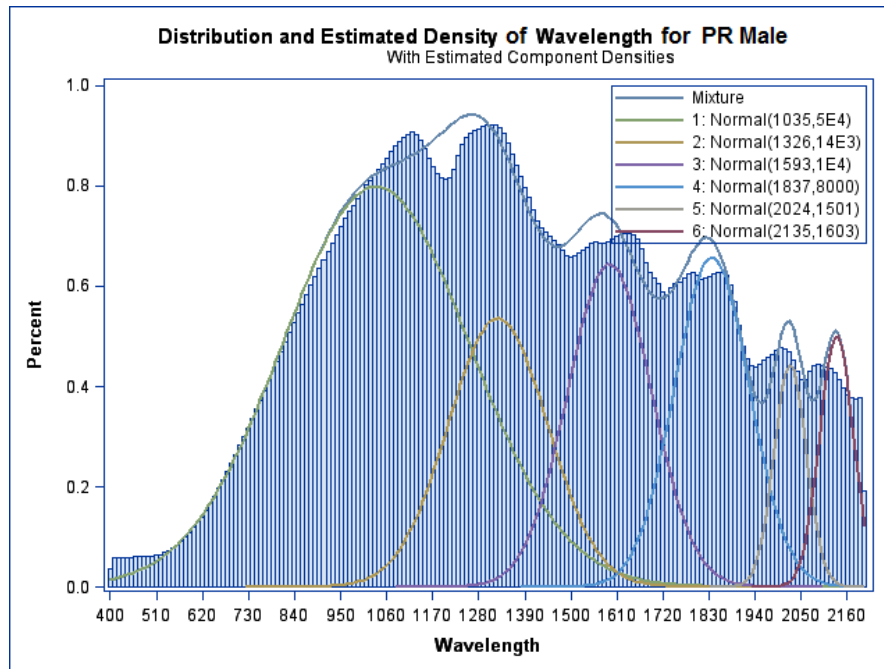


Figure D16. Finite mixture model for Cerambycidae: *Prionus californicus* (PR) male with six normal curves.

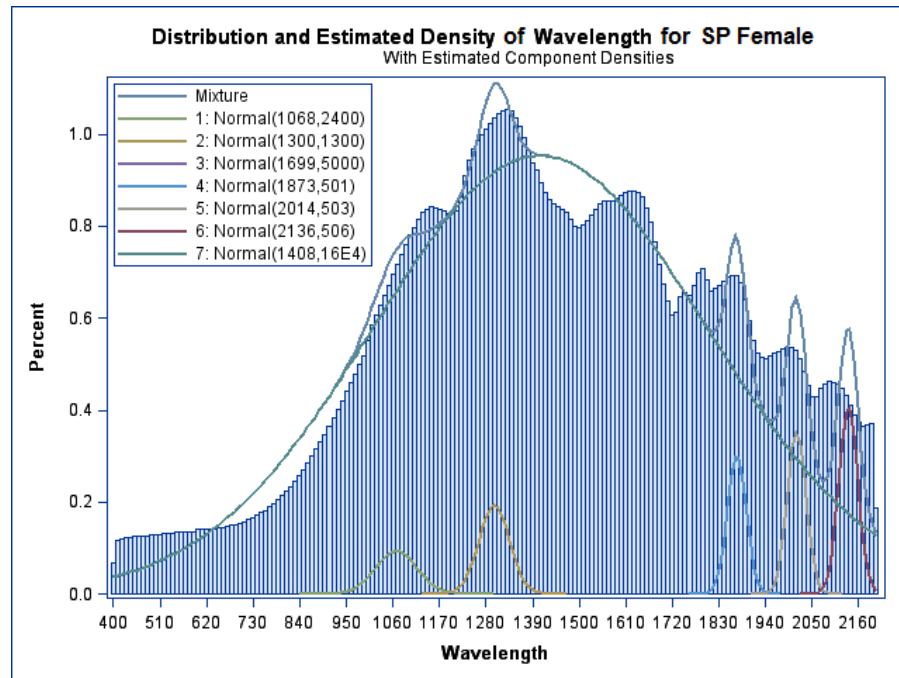


Figure D17. Finite mixture model for Cerambycidae *Spondylis upiformis* (SP) female with six normal curves.

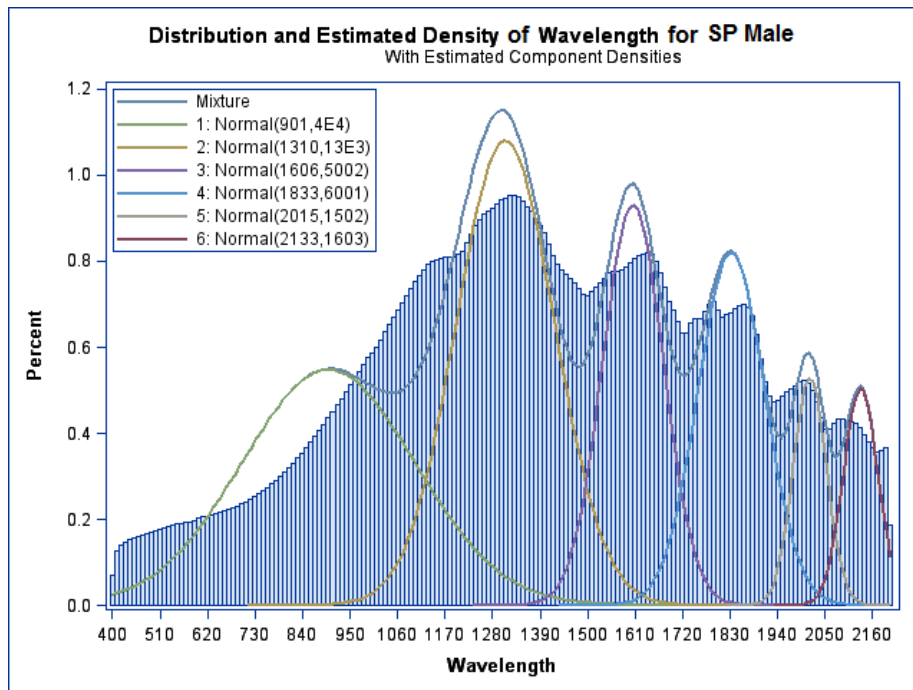


Figure D18. Finite mixture model for Cerambycidae *Spondylis upiformis* (SP) male with six normal curves.

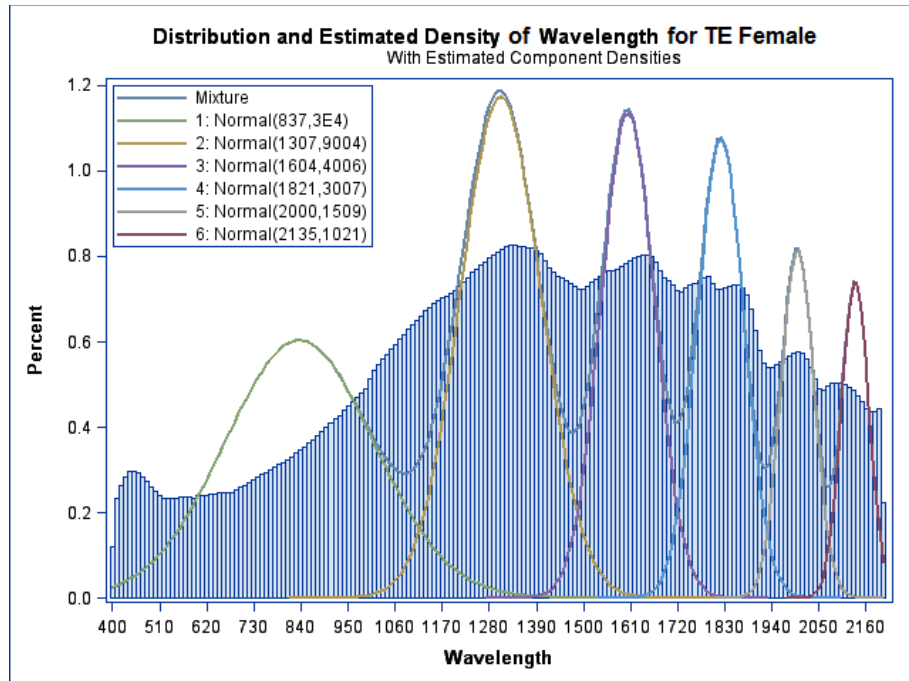


Figure D19. Finite mixture model for Trogossitidae: *Temnocheila chlorodia* (TE) female with six normal curves.

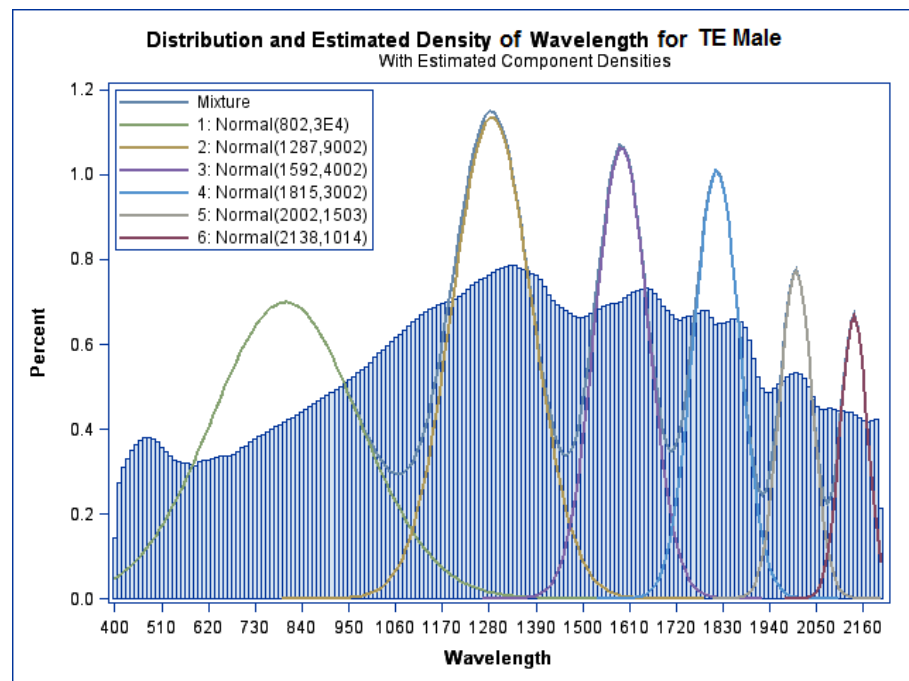


Figure D20. Finite mixture model for *Temnocheila chlorodia* (TE) chlorodia male with six normal curves.

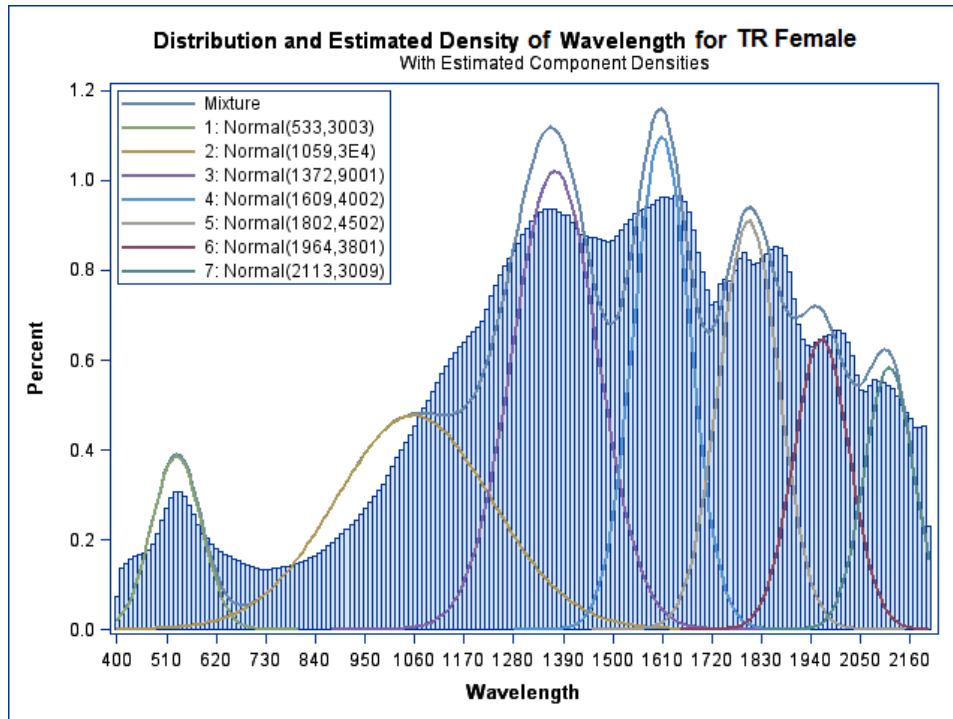


Figure D21. Finite mixture model Buprestidae: *Trachykele blondeli blondeli* (TR) female with seven normal curves.

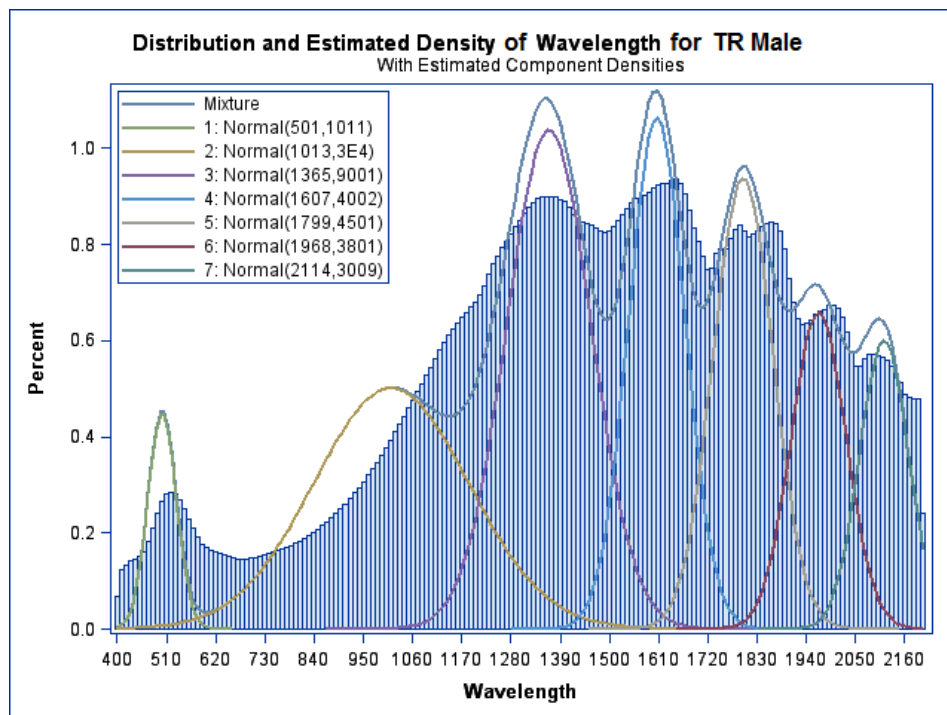


Figure D22. Finite mixture model for Buprestidae: *Trachykele blondeli blondeli* (TR) male with seven normal curves.

Appendix E

Correlation Matrix of R1 through R18 variables generated from the finite mixture models.

Table E1. Correlation coefficients among variables R1 through R18 generated using finite mixture models.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18
R1	1	0.95	0.84	0.22	0.06	0.04	0.01	0.10	0.05	0.13	0.26	0.36	0.37	0.36	0.40	0.45	0.57	0.55
R2	0.95	1	0.88	0.20	0.02	0.00	-0.05	0.05	0.01	0.10	0.24	0.34	0.36	0.35	0.40	0.44	0.58	0.54
R3	0.84	0.88	1	0.52	0.32	0.29	0.21	0.27	0.21	0.27	0.39	0.46	0.47	0.45	0.49	0.52	0.63	0.60
R4	0.22	0.20	0.52	1	0.97	0.95	0.89	0.83	0.75	0.69	0.64	0.57	0.53	0.51	0.51	0.47	0.39	0.41
R5	0.06	0.02	0.32	0.97	1	0.99	0.97	0.90	0.84	0.76	0.66	0.57	0.52	0.50	0.50	0.43	0.30	0.33
R6	0.04	0.00	0.29	0.95	0.99	1	0.98	0.92	0.87	0.79	0.70	0.61	0.56	0.54	0.53	0.45	0.32	0.35
R7	0.01	-0.05	0.21	0.89	0.97	0.98	1	0.96	0.93	0.85	0.75	0.65	0.60	0.58	0.57	0.48	0.32	0.35
R8	0.10	0.05	0.27	0.83	0.90	0.92	0.96	1	0.99	0.96	0.90	0.81	0.77	0.76	0.74	0.65	0.49	0.49
R9	0.05	0.01	0.21	0.75	0.84	0.87	0.93	0.99	1	0.98	0.91	0.82	0.79	0.78	0.76	0.66	0.48	0.48
R10	0.13	0.10	0.27	0.69	0.76	0.79	0.85	0.96	0.98	1	0.97	0.91	0.88	0.88	0.86	0.77	0.60	0.60
R11	0.26	0.24	0.39	0.64	0.66	0.70	0.75	0.90	0.91	0.97	1	0.98	0.97	0.96	0.95	0.88	0.76	0.75
R12	0.36	0.34	0.46	0.57	0.57	0.61	0.65	0.81	0.82	0.91	0.98	1	0.99	0.99	0.98	0.93	0.86	0.85
R13	0.37	0.36	0.47	0.53	0.52	0.56	0.60	0.77	0.79	0.88	0.97	0.99	1	1	0.99	0.95	0.89	0.87
R14	0.36	0.35	0.45	0.51	0.50	0.54	0.58	0.76	0.78	0.88	0.96	0.99	1	1	1.00	0.96	0.88	0.87
R15	0.40	0.40	0.49	0.51	0.50	0.53	0.57	0.74	0.76	0.86	0.95	0.98	0.99	1	1	0.97	0.90	0.89
R16	0.45	0.44	0.52	0.47	0.43	0.45	0.48	0.65	0.66	0.77	0.88	0.93	0.95	0.96	0.97	1	0.95	0.93
R17	0.57	0.58	0.63	0.39	0.30	0.32	0.32	0.49	0.48	0.60	0.76	0.86	0.89	0.88	0.90	0.95	1	0.99
R18	0.55	0.54	0.60	0.41	0.33	0.35	0.35	0.49	0.48	0.60	0.75	0.85	0.87	0.87	0.89	0.93	0.99	1