

USING FUNCTIONAL BOXPLOTS TO VISUALIZE REFLECTANCE DATA AND DISTINGUISH  
BETWEEN AREAS OF NATIVE GRASSES AND INVASIVE OLD WORLD BLUESTEMS IN A  
KANSAS TALL GRASS PRAIRIE

by

GARTH HIGHLAND

B.S., Kansas State University, 2003

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics  
College of arts and Sciences

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

2012

Approved by:

Major Professor  
Dr. Leigh Murray

# **Copyright**

GARTH HIGHLAND

2012

## Abstract

Using remotely sensed reflectance data is an appealing tool for controlling invasive species of grasses by rangeland managers. Recent developments in functional data analysis include the functional boxplot (FBP) which is shown here to be a useful tool in the visualization of reflectance data. Functional boxplots are a novel method of visually inspecting functional data and determining the presence of outliers in the data. Implementation and interpretation of FBPs are both straightforward and intuitive. The goal of this study is to examine the use of FBPs for visualizing reflectance data, and to determine the efficacy of using the FBP to distinguish between native tall grasses and invasive Old World Bluestem (OWB, *Bothriochloa spp.*) monocultures in a Kansas prairie.

Validation trials were conducted in order to determine the stability of the FBP when used to analyze spectral data. FBPs were shown to be highly stable for use with both native and OWB grasses at all times and subsets of wavelengths tested.

Identification trials were conducted by introducing a single OWB observation to a test set of native tall grass observations and constructing a FBP. Results indicate that using observations recorded early in the growing season, the functional boxplot is able to successfully identify the OWB observation as an outlier in a test set of native tall grass observations with an estimated probability 100% and 95.45% when considering the visible and cellular spectrums, respectively. A 95% lower bound for the probability of successfully identifying the OWB observation using the cellular spectrum in May is found to be 89.67%.

# Table of Contents

List of Figures .....	v
List of Tables.....	vi
Chapter 1 - Introduction.....	1
Chapter 2 - Materials and Methods.....	4
2.1 Introduction to the Functional Boxplot.....	4
2.2 Description of Reflectance Data .....	7
2.3 Application of the Functional Boxplot to Reflectance Data.....	8
2.3.1 Validation Trials .....	8
2.3.2 Identification Trials.....	9
Chapter 3 - Results.....	11
3.1 Validation Trials .....	11
3.2 Identification Trials.....	12
Chapter 4 - Conclusions and Discussion .....	14
References.....	16
Appendix A - Validation Trial Code .....	18
Appendix B - Identification Trial Code .....	23

## List of Figures

Figure 2.1 Functional boxplot of all May (2009) native grass observations with outliers .....	5
Figure 2.2 Functional boxplot of all May (2009) OWB grass observations with outliers. ....	6
Figure 2.3 Native and OWB quadrats.....	7
Figure 2.4 Example run of identification trial.....	9
Figure 3.1 Histograms of proportion data generated in validation trials .....	11

## List of Tables

Table 3.1 Results from identification trials.....	12
---	----

# Chapter 1 - Introduction

Invasive species of grasses pose unique problems for those involved in native prairie conservation and rangeland management. Often, it is necessary to monitor vast expanses of land in order to effectively quell an invasion. Of particular interest for the Konza Prairie or to ranchers in the Great Plains is the management of the Old World Bluestem (OWB; *Bothriochloa spp.*) family of grasses. OWBs are a group of non-native grasses that were introduced to this country in the 1920's (Celarier and Harlen, 1955). A significant and documented introduction of a member of the OWB group is the Caucasian bluestem, (*Bothriochloa bladhii* (Retz.) S.T. Blake) that was widely distributed by the Manhattan, KS Soil Conservation Nursery after receipt by the US Department of Agriculture from a Georgian botanical garden in 1929 (Celarier and Harlen, 1955). OWB was widely used for soil stabilization, in road rights-of-way, and seeded for grazing or hay production (Harmony and Hickman, 2004).

Invasion by the OWB family of grasses has had negative implications for both ranchers and wildlife preservationists. For example, it has been shown that the average gains per acre for beef production were significantly lower for pastures planted to Caucasian Bluestem relative to pastures consisting of a native mixture (Launchbaugh 1971). OWB monocultures have also been shown to significantly reduce arthropod biomass as well as bird abundance and species diversity in a Kansas prairie habitat (Hickman et al. 2006).

OWB grasses are particularly aggressive when compared to the native blend of range grasses found in the Great Plains. In Harmony and Hickman's study (2004), it was observed that the rapid maturation of the OWB results in a poorer forage quality which

may result in higher utilization of native grasses. A higher rate of maturation combined with lower utilization by grazing animals results in an appreciable reproductive advantage of OWB over native grasses in settings such as the Konza prairie. Experiments have also confirmed that OWB in the presence of native species directly causes reduced growth in native species of grasses (Schmidt et al. 2008).

Using spectral signatures to distinguish between species of plants is of growing scientific interest. Land managers and conservation biologists who have a vested interest in documenting the spread of invasive species may have knowledge of plausible areas of invasion, but usually lack detailed maps (He et al. 2011). There are many methods available to distinguish between species of plants using spectral reflectance data. Current methods for discriminating between plant species include but are not limited to the minimum noise fraction (MNF) which consists of two principle component analysis rotations, band ratio indices, and continuum removal (Underwood et al 2003).

Problems can be encountered in classification of spectral signatures for a number of reasons. Variables encountered during real-world application of spectral reflectance data collection such as atmospheric conditions, illumination angle, viewing angle, and many others lead to difficulties in the unique identification of materials (Cochrane 2000). These variables are difficult to hold constant, especially if attempting to match spectral readings to those found in a spectral library where these conditions are not known. Furthermore, spectral signatures may not significantly differ between certain species due to variation within a species (Price 1994).

It is the goal of this study to examine using functional boxplots as a method of exploring the shape and variability of spectral signatures. Furthermore, this study aims to



assess the use of outlier detection as a method to identify an OWB observation when compared to a native species test data.

## Chapter 2 - Materials and Methods

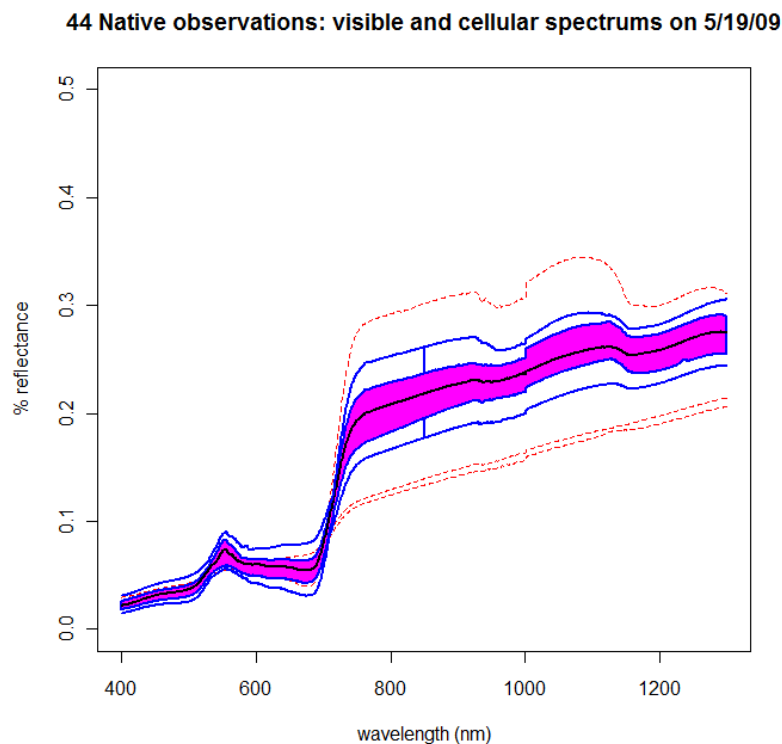
### 2.1 Introduction to the Functional Boxplot

Traditional boxplots provide a graphical means for visualizing univariate data (Tukey 1977). Boxplots give an assessment of symmetry, the location of non-outlying minimum and maximum observations as well as the first, second, and third quartiles for sample univariate data. An observation is flagged as an outlier if it falls outside of 1.5 times the inner quartile range (IQR) (Tukey 1977, Anderson et al 2005). The traditional boxplot is a robust tool for visualizing data, assessing symmetry of the data, and finding outliers. However, this method of assessing univariate data is unable to account for the complexity of functional data where multiple measurements are taken per observation.

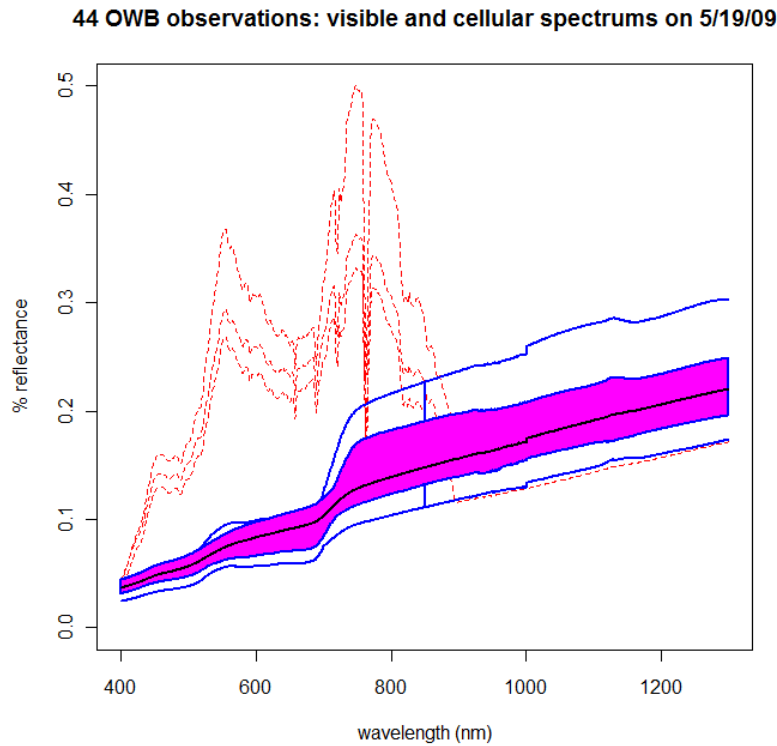
In the case of reflectance data, thousands of wavelengths are measured simultaneously. Because of the high correlation between wavelength measurements, the entire spectral signature is considered to be one functional observation. All observations used in this study are considered to be functional over their respective wavelength ranges. Due to the complexity of spectral signatures, it is appreciably more difficult to establish an order to functional observations as one naturally would with univariate data by sorting. When constructing a traditional boxplot, the observations are first sorted by numeric value so that the median observation may be found. For this reason various methods have been proposed by numerous authors to associate a depth score to each functional observation by which the observations can then be ranked (Sun and Genton 2011). López-Pintado and Romo (2009) pioneered the concept of the modified band depth score (MBD) as a robust way of ranking functional data.

The MBD measures the proportion of an observation that lies within a band bounded by two observations. In a data set with  $N$  observations, there are  $\binom{N}{2} = \frac{(N)!}{(2)!(N-2)!}$  number of bands for which a depth will be found for a single functional observation. The sum of these depth scores divided by the total number of bands is the MBD. Sun and Genton's work (2011, 2012) has used the MBD to create the functional boxplot (FBP). Figures 2.1 and 2.2 show FBPs that were created using 44 spectral observations from native and OWB grasses respectively (see section 2.2). Components include the median observation (black), 50% central region (magenta region), maximum envelope (blue curve), and plausible outliers (red dashed curves). The median curve is the most central curve with the highest MBD score (Sun and Genton 2011).

**Figure 2.1 Functional boxplot of all May (2009) native grass observations with outliers**



**Figure 2.2 Functional boxplot of all May (2009) OWB grass observations with outliers.**



The 50% central envelope of the FBP is analogous to the inner quartile range found in a univariate box plot (Sun and Genton 2011). Both the median and 50% central envelope are robust to outliers and thus provide an unbiased visualization of centrality and spread (Sun and Genton 2011). Functional observations that fall outside 1.5 times the IQR at any point are tagged as outliers. In practice, a spectral signature is considered to be an outlier if it is observed from a different biological process than the majority of the other observations (Hyndman and Shang 2010). This notion motivates an assessment of the FBP's efficacy when used to analyze reflectance data.

## 2.2 Description of Reflectance Data

Spectroradiometer measurement observations used in this study were collected between 10:00 AM and 2:00 PM CST on May 19, June 17, and July 19 in 2009 (Grabow 2010). The selected study consisted of areas dominated by either native tallgrass prairie species or OWB monocultures at a site located near Olsburg, Kansas.

**Figure 2.3 Native and OWB quadrats.**



Pictures are of actual quadrats where observations were taken in this study. On the left is a native grass quadrat, on the right is an OWB quadrat.

At this site, two locations were selected within each monoculture. Within each location two 200m transects were permanently identified. From each transect, eleven 1.0 m x 1.0 m quadrats were used as the observational unit. Readings were taken using an ASD Field Spec® 3 Portable Spectroradiometer with 2,150 discrete bands ranging from 350 to 2,500 nm wavelengths. Spectrometer readings were calibrated with a Spectralon (ASD) reference panel on each transect which was then used to convert the spectrometer readings on target materials to percent reflectance. For more information regarding the collection of this data, refer to Grabow (2010).

For the purpose of this study, the data from each of the two locations with the same monoculture were combined, resulting in 44 total functional observations for each monoculture at each date. Analysis was conducted using the visible spectrum (400 to 730 nm), cellular spectrum (731 to 1300 nm), or the combined visible and cellular spectrums.

## **2.3 Application of the Functional Boxplot to Reflectance Data**

All data manipulation and analysis in this study was conducted using R software (R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>). The *fbplot* command used to create the functional boxplots is a part of the *fda* package and is fully documented at: <http://cran.r-project.org/web/packages/fda/fda.pdf>. Applicable R-code and subsequent outputs used in this analysis are available in Appendices A and B.

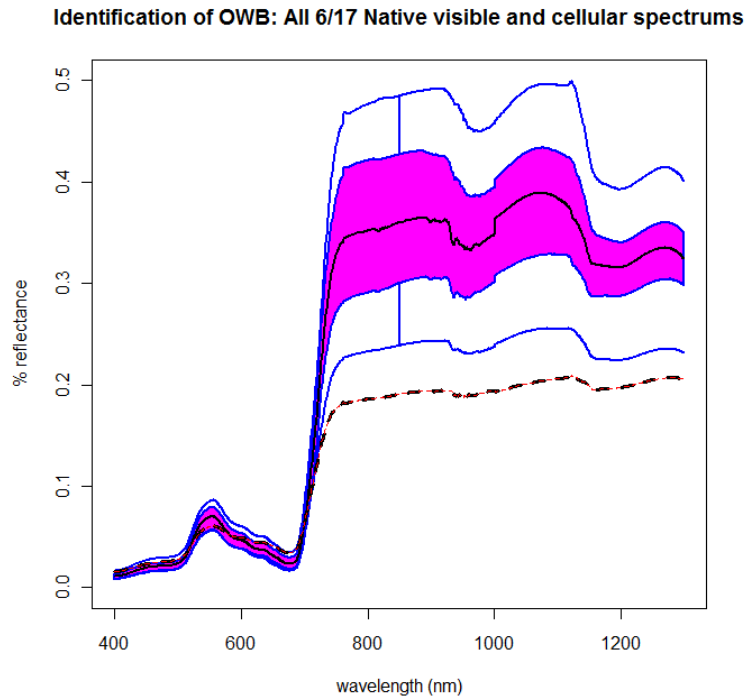
### **2.3.1 Validation Trials**

A validation trial was conducted to determine the stability of the FBP when used on spectral data. For each species, wavelength range, and time combination, 44 validation FBPs were created. Each validation FBP was created with a different observation omitted and subsequently plotted over the FBP created with the remaining observations. Note that the MBD is not able to be calculated for the omitted observation from the FBP because it has been omitted from the MBD algorithm. The proportion of the total wavelength range for the deleted observation that fell outside the resulting maximal envelope was recorded.

### 2.3.2 Identification Trials

Identification trials were conducted to estimate the ability of the FBP to correctly identify a single OWB observation as an outlier when considered with a test set comprised of only native grass observations. Each identification trial was conducted for a particular time and wavelength combination yielding a total of 9 trials. First, for each trial, a full FBP was constructed using all 44 native observations. All observations within the full native grass data set identified as outliers by the overall FBP were removed to obtain a clean test set, void of outliers. Second, using a looping scheme, identification FBPs were created after adding an individual OWB observation from the corresponding time and wavelength range to the test set. This process continued for all 44 of the OWB observations within each time and wavelength combination.

**Figure 2.4 Example run of identification trial.**



In this iteration of the identification trial, the FBP technique correctly identifies an OWB observation (black dashed line) as an outlier.

For each OWB observation, the MBD of the OWB observation was recorded as well as if the OWB observation was considered an outlier in accordance with the 1.5 IQR rule. A 95% lower confidence bound for  $\pi$  = the probability of successfully identifying the OWB observation as an outlier was calculated for each of the nine identification trials. This lower bound is the lowest value of  $\pi^*$  such that the probability of correctly identifying more than or equal to  $x_j$  out of 44 OWB observations is approximately 95%. The lower bound estimate satisfies the equation

$$0.95 \approx 1 - \sum_{i=0}^{x_j} \binom{44}{i} (\pi^*)^i (1 - \pi^*)^{44-i}$$

where  $x_j$  is the total number of successful identifications of the OWB observation in the  $j^{\text{th}}$  trial. Finding the lower bound was done by evaluating this equation for each  $\pi_i^*$  between 0 and 1 by increments of 0.0001, then minimizing the loss function which is defined as

$$L(\pi_i^*) = |0.095 - P(X \geq x_j | \pi_i^*, N=44)| .$$

The lower bound is not possible to calculate where  $\hat{\pi} = 1.0$  when all 44 of the OWB observations were correctly identified.

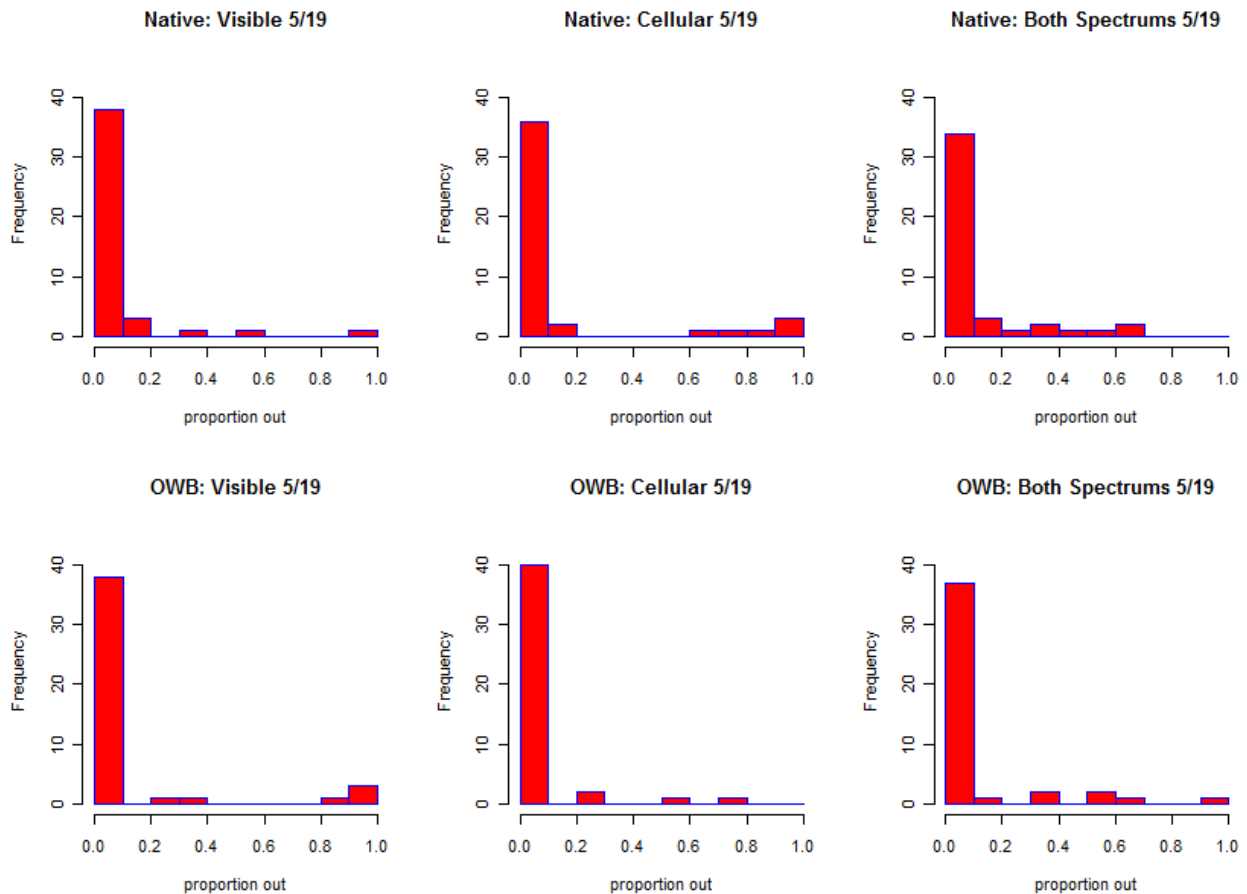


# Chapter 3 - Results

## 3.1 Validation Trials

Validation trials demonstrated that for this data, the FBP is extremely stable when subjected to removal of observations. Forty four validation FBPs were created for each species, time, and wavelength combination and each was compared to the full FBP constructed using all of the observations from the respective data. The proportion of a deleted observation that fell outside the validation FBP's maximal envelope was recorded. A set of histograms showing frequency of these proportions from the validation trial using May data are presented in Figure 3.1.

**Figure 3.1 Histograms of proportion data generated in validation trials**



The other two dates have similar results. Shapes of the resulting histograms indicate that the majority of the deleted observations clearly tended to be included, or mostly included in the resulting validation FBP's maximal envelope. On average, 88.63% of omitted observations had proportion readings less than or equal to 0.1.

### 3.2 Identification Trials

Table 3.1 displays the resulting estimated probabilities and 95% lower bound for identifying OWB quadrats as outliers.

**Table 3.1 Results from identification trials.**

May 19, 2009	Visible Spectrum (400-730nm)	Cellular Spectrum (731-1300nm)	Both Spectrums (400-1300nm)
Number of native observations in test set	43	39	41
Number (and proportion) of OWB identified as outliers	44 (1.0000)	42 (0.9545)	42 (0.9545)
95% lower bound of estimated probability	.	0.8967	0.8967

June 17, 2009	Visible Spectrum (400-730nm)	Cellular Spectrum (731-1300nm)	Both Spectrums (400-1300nm)
Number of native observations in test set	43	44	44
Number (and proportion) of OWB identified as outliers	30 (0.6818)	11 (0.2500)	30 (0.6818)
95% lower bound of estimated probability	0.5718	0.1655	0.5718

July 19, 2009	Visible Spectrum (400-730nm)	Cellular Spectrum (731-1300nm)	Both Spectrums (400-1300nm)
Number of native observations in test set	44	43	44
Number (and proportion) of OWB identified as outliers	25 (0.5682)	0 (0.0000)	16 (0.3636)
95% lower bound of estimated probability	0.4559	0.0012	0.2632

Results indicate that spectral differences between native and OWB grasses are easier to identify using the FBP earlier in the growing season, compared to later in the season. Estimated probabilities of successfully identifying an OWB observation as an outlier are the highest for May and are substantially lower for all spectrum ranges in June and July. In general, the identification FBP using the visible light spectrum performed better than that of the cellular spectrum. All 44 of the OWB observations were successfully identified as outliers when using the visible spectrum in May. When considering either the cellular or both spectrums, 42 of the 44 OWB observations were found to be outliers in May.

## Chapter 4 - Conclusions and Discussion

This research has illustrated that FBPs are a powerful tool for researchers working with reflectance data. Not only are they useful for visualizing and cleaning data sets, but also show great promise in being able to distinguish between spectral signatures observed from different species. Early in the growing season, identification trials were shown to be highly capable of differentiating between two monocultures of grasses. Using only the visible spectrum in May, the identification trial achieved a 100% rate of successfully identifying the OWB when compared to a native species test set, while the 95% lower bound for the probability of success was estimated to be greater than 0.89 for both the cellular and combined visible and cellular wavelengths.

Identification of materials using reflectance data poses challenges due to difficulty in managing variables such as angle of observation, atmospheric conditions, moisture content, and angle of illumination (Cochrane 1999). Inability to control these variables at different sample collection times can introduce error into the identification process when using a bank of observations to compare future observations to. Prior to collection of data over a large tract of land, an area representative of the native species should be identified. By using observations from this representative area to build the test set for analysis, the error introduced by these variables is likely to be minimized and could potentially provide an advantage over the bank method.

This work was conducted with a small data set of only 44 quadrats per vegetation type that was painstakingly collected by hand. Future work in this area should be focused on more realistic applications in the field, where data is collected remotely over large tracts

of land. This method could be used to generate maps of plausible invasion areas with which researchers could enter the field and establish the ground truth.

Remote sensing of spectral data is becoming more obtainable and useable as cost continues to decrease and resolution increases (Nebiker 2008). The promise of affordable, high quality reflectance data recorded over large tracts of land will eventually bring reflectance data into the workplace for rangeland managers and researchers. The simplicity and power of the FBP is likely to become an asset when working with these functional data sets in the future.

## References

- Anderson, D. R., Sweeney, D. J., & Williams T. A. (2005). Descriptive Statistics: Numerical Measures. *Statistics for Business and Economics 9e*, 102.
- Celarier, R. P. & Harlan, J. R. (1955). Studies on Old World Bluestems. Technical bulletin, Agricultural Experiment Station, Division of Agriculture, Oklahoma A.&M. College, 58, 5-31.
- Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*, 21(10), 2075-2087.
- Grabow, B. S. P. (2010). Comparing hyperspectral reflectance characteristics of Caucasian Bluestem and Native Tallgrass Prairie over a growing season. M.S. Thesis, Kansas State University, Manhattan KS, 1- 61.
- Harmoney, K. R. & Hickman, K. R. (2004). Comparative morphology of Caucasion Old World Bluestem and Native Grasses. *Agronomy Journal*, 96, 1540-1544.
- He, K. S., Rocchini, D., Neteler, M., & Nagendra, H. (2011). Benefits of hyperspectral remote sensing for tracking plant invasions. *Diversity and Distributions*, 17, 381-392.
- Hickman, K. R., Farley, G. H., Channell, R., & Steier, J. E. (2006). Effects of Old World Bluestem (*Bothriochloa Ischaemum*) on food availability and avian community composition within the mixed-grass prairie. *The Southwestern Naturalist*, 51(4), 524-530.
- Hyndman, R. J. & Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1), 29-45.
- Launchbaugh, J. L. (1971). Upland seeded pastures compared for grazing steers at Hays, Kansas. Technical bulletin for Kansas Agricultural Experiment Station, Kansas State University, 548, 1-29.
- López-Pintado, S. & Romo, J. (2009). On the concept of depth of functional data. *Journal of the American Statistical Association*, 104(486), 718-734.

Nebiker, S., Annen, A., Scherrer, M., & Oesch, D. (2008). A light-weight multispectral sensor for micro UAV- opportunities for very high resolution airborne remote sensing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 1193-1200.

Price, J. C. (1994). How unique are spectral signatures? *Remote Sensing of Environment*, 49, 181-186.

Schmidt, C. D., Hickman, K. R., Channell, R., Harmony, K., & Stark, W. (2008). Competitive abilities of native grasses and non-native (*Bothriochloa* spp.) grass. *Plant Ecology*, 197, 69-80.

Sun, Y., Genton, M. G. (2011). Functional Boxplots. *Journal of Computational and Graphical Statistics*, 20, 1-19.

Sun, Y., Genton, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, 23, 54-64.

Tukey, John W. (1977). Box-and-whisker plots. *Exploratory Data Analysis*, 39-47.

Underwood, E., Ustin, S., & DiPietro, D. (2003). Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment*, 86, 150-161.

## **Appendix A - Validation Trial Code**



```

#####
#reading in May data and wavelength information#
#####

wav=read.table("C:\\Users\\Garth Highland\\Desktop\\fbplot\\wavelength.txt", header=T)
spect=read.table("C:\\Users\\Garth Highland\\Desktop\\fbplot\\05.19.spectral.readings.txt",row.names=1, header=T)
is.data.frame(spect)
spect=data.matrix(spect)
wav=data.matrix(wav)
is.matrix(spect)
is.matrix(wav)
dim(spect)
spect=spect*100
library(fda)

#####
#creating functional boxplots of 5/19/2009 data#
#####

pdf(file="C:\\Users\\Garth Highland\\Desktop\\appendixA1.1.pdf")

fbplot(spect[50:380,1:44], x=wav[50:380,], ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="5/19 Native visible spectrum")

fbplot(spect[50:380,45:88], x=wav[50:380,], ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="5/19 OWB visible spectrum")
a
fbplot(spect[381:950,1:44], x=wav[381:950,], ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="5/19 Native cellular spectrum")

fbplot(spect[381:950,45:88], x=wav[381:950,], ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="5/19 OWB cellular spectrum")

fbplot(spect[50:950,1:44], x=wav[50:950,], ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="5/19 Native visible and cellular spectrums")

fbplot(spect[50:950,45:88], x=wav[50:950,], ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="5/19 OWB visible and cellular spectrums")

dev.off()

#####
#manipulation of fbplot() command to enable graphical overlay#
#####

fbboxplot=
function (fit, r=NULL , x = NULL, method = "MBD", depth = NULL, plot = TRUE,
        prob = 0.5, color = 6, outliercol = 2, barcol = 4, fullout = FALSE,
        factor = 1.5, ...)
{
  if (is.fdsSmooth(fit) | is.fdsPar(fit)) {
    fit = fit$fd
  }
  if (is.fds(fit)) {
    if (length(x) == 0) {
      x = seq(fit$basis$rangeval[1], fit$basis$rangeval[2],
              len = 101)
    }
    fit = eval.fds(x, fit)
  }
  tp = dim(fit)[1]
  n = dim(fit)[2]
  if (length(x) == 0) {
    x = 1:tp
  }
}

```

```

}
if (length(depth) == 0) {
  if (method == "BD2") {
    depth = BD2(t(fit))
  }
  else if (method == "BD3") {
    depth = BD3(t(fit))
  }
  else if (method == "MBD") {
    depth = MBD(t(fit))
  }
  else if (method == "Both") {
    depth = round(BD2(t(fit)), 4) * 10000 + MBD(t(fit))
  }
}
dp_s = sort(depth, decreasing = T)
index = order(depth, decreasing = T)
if (plot) {
  plot(x, fit[, index[1]], lty = 1, lwd = 2, col = 1, type = "l",
    ...)
}
for (pp in 1:length(prob)) {
  m = ceiling(n * prob[pp])
  center = fit[, index[1:m]]
  out = fit[, index[(m + 1):n]]
  inf = apply(center, 1, min)
  sup = apply(center, 1, max)
  if (prob[pp] == 0.5) {
    dist = factor * (sup - inf)
    upper = sup + dist
    lower = inf - dist
    outly = (fit <= lower) + (fit >= upper)
    outcol = colSums(outly)
    remove = (outcol > 0)
    colum = 1:n
    outpoint = colum[remove == 1]
    out = fit[, remove]
    woout = fit
    good = woout[, (remove == 0), drop = FALSE]
    maxcurve = apply(good, 1, max)
    mincurve = apply(good, 1, min)
    if (sum(outly) > 0) {
      if (plot) {
        matlines(x, out, lty = 2, col = outliercol,
          type = "l", ...)
      }
    }
    barval = (x[1] + x[tp])/2
    bar = which(sort(c(x, barval)) == barval)[1]
    if (plot) {
      lines(c(x[bar], x[bar]), c(maxcurve[bar], sup[bar]),
        col = barcol, lwd = 2)
      lines(c(x[bar], x[bar]), c(mincurve[bar], inf[bar]),
        col = barcol, lwd = 2)
    }
  }
  xx = c(x, x[order(x, decreasing = T)])
  supinv = sup[order(x, decreasing = T)]
  yy = c(inf, supinv)
  if (plot) {
    if (prob[pp] == 0.5) {
      polygon(xx, yy, col = color[pp], border = barcol,
        lwd = 2)
    }
    else {
      polygon(xx, yy, col = color[pp], border = NA)
    }
  }
}
if (plot) {
  lines(x, fit[, index[1]], lty = 1, lwd = 2, col = 1,

```

```

        type = "l")
lines(x, maxcurve, col = barcol, lwd = 2)
lines(x, mincurve, col = barcol, lwd = 2)

lines(x, r, col="black", lwd=3, lty=2 )

if (fullout) {
  if (sum(outly) > 0) {
    if (plot) {
      matlines(x, out, lty = 2, col = outliercol,
              type = "l", ...)
    }
  }
}
return(list(depth = depth, outpoint = outpoint))
}

#####
#validation trials for 5/19/2009: visible spectrum, both species#
#####

X=0
pdf(file="C:\\Users\\Garth Highland\\Desktop\\Native.5.19.visible.pdf")
for(i in 1:44){
data=spect[50:380,1:44]
use=data[,-i]
X=fboxplot(use,
fullout=T,x=wav[50:380,],r=data[,i],ylim=c(0,25),
ylab="% reflectance",xlab="wavelength (nm)",
main="Delete an observation stability test: All 5/19 Native visible spectrum")
write(as.vector(X$depth), file="C:\\Users\\Garth
Highland\\Desktop\\Native.5.19.visible.banddepth.xls",sep = "\t",ncolumns=43,append=TRUE)
write(as.vector(X$outpoint), file="C:\\Users\\Garth
Highland\\Desktop\\Native.5.19.visible.outliers.xls",sep = "\t",ncolumns=43,append=TRUE)
use=0
X=0
}
dev.off()

X=0
pdf(file="C:\\Users\\Garth Highland\\Desktop\\OWB.5.19.visible.pdf")
data=spect[50:380,45:88]
for(i in 1:44){
use=data[,-i]
X=fboxplot(use,
fullout=T,x=wav[50:380,],r=data[,i],ylim=c(0,25),
ylab="% reflectance",xlab="wavelength (nm)",
main="Delete an observation stability test: All 5/19 OWB visible spectrum")
write(as.vector(X$depth), file="C:\\Users\\Garth
Highland\\Desktop\\OWB.5.19.visible.banddepth.xls",sep = "\t",ncolumns=43,append=TRUE)
write(as.vector(X$outpoint), file="C:\\Users\\Garth
Highland\\Desktop\\OWB.5.19.visible.outliers.xls",sep = "\t",ncolumns=43,append=TRUE)
use=0
X=0
}
dev.off()

#####
#validation trials for 5/19/2009: cellular spectrum, both species#
#####

X=0
pdf(file="C:\\Users\\Garth Highland\\Desktop\\Native.5.19.cellular.pdf")
for(i in 1:44){
data=spect[381:950,1:44]
use=data[,-i]

```

```

X=fboxplot(use,
fullout=T,x=wav[381:950,],r=data[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="Delete an observation stability test: All 5/19 Native cellular spectrum")
write(as.vector(X$depth), file="C:\\Users\\Garth
Highland\\Desktop\\Native.5.19.cellular.banddepth.xls",sep = "\t",ncolumns=43,append=TRUE)
write(as.vector(X$outpoint), file="C:\\Users\\Garth
Highland\\Desktop\\Native.5.19.cellular.outliers.xls",sep = "\t",ncolumns=43,append=TRUE)
use=0
X=0
}
dev.off()

X=0
pdf(file="C:\\Users\\Garth Highland\\Desktop\\OWB.5.19.cellular.pdf")
data=spect[381:950,45:88]
for(i in 1:44){
use=data[, -i]
X=fboxplot(use,
fullout=T,x=wav[381:950,],r=data[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="Delete an observation stability test: All 5/19 OWB cellular spectrum")
write(as.vector(X$depth), file="C:\\Users\\Garth
Highland\\Desktop\\OWB.5.19.cellular.banddepth.xls",sep = "\t",ncolumns=43,append=TRUE)
write(as.vector(X$outpoint), file="C:\\Users\\Garth
Highland\\Desktop\\OWB.5.19.cellular.outliers.xls",sep = "\t",ncolumns=43,append=TRUE)
use=0
X=0
}
dev.off()

#####
#validation trials for 5/19/2009: combined visible and cellular spectrums, both species#
#####

X=0
pdf(file="C:\\Users\\Garth Highland\\Desktop\\Native.5.19.both.pdf")
for(i in 1:44){
data=spect[50:950,1:44]
use=data[, -i]
X=fboxplot(use,
fullout=T,x=wav[50:950,],r=data[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="Delete an observation stability test: All 5/19 Native full spectrum")
write(as.vector(X$depth), file="C:\\Users\\Garth
Highland\\Desktop\\Native.5.19.both.banddepth.xls",sep = "\t",ncolumns=43,append=TRUE)
write(as.vector(X$outpoint), file="C:\\Users\\Garth
Highland\\Desktop\\Native.5.19.both.outliers.xls",sep = "\t",ncolumns=43,append=TRUE)
use=0
X=0
}
dev.off()

X=0
pdf(file="C:\\Users\\Garth Highland\\Desktop\\OWB.5.19.both.pdf")
for(i in 1:44){
data=spect[50:950,45:88]
use=data[, -i]
X=fboxplot(use,
fullout=T,x=wav[50:950,],r=data[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main="Delete an observation stability test: All 5/19 OWB full spectrum")
write(as.vector(X$depth), file="C:\\Users\\Garth
Highland\\Desktop\\OWB.5.19.both.banddepth.xls",sep = "\t",ncolumns=43,append=TRUE)
write(as.vector(X$outpoint), file="C:\\Users\\Garth
Highland\\Desktop\\OWB.5.19.both.outliers.xls",sep = "\t",ncolumns=43,append=TRUE)
use=0
X=0
}
dev.off()

```

## **Appendix B - Identification Trial Code**

```

#####
#identification of outliers#
#####

fbplot(spect[50:380,1:44],fullout=T,x=wav[50:380,],
ylim=c(0,50),ylab="% reflectance",xlab="wavelength (nm)",
main="All Native Observations on May 19: visible spectrum")
#outlier:30#

fbplot(spect[381:950,1:44],fullout=T,x=wav[381:950],
ylim=c(0,50),ylab="% reflectance",xlab="wavelength (nm)",
main="All Native Observations on May 19: cellular spectrum")
#outlier:5,16,17,44#
#secondary outlier:2#

fbplot(spect[50:950,1:44],fullout=T,x=wav[50:950],
ylim=c(0,50),ylab="% reflectance",xlab="wavelength (nm)",
main="All Native Observations on May 19: visible and cellular spectrum")
#outlier:5,16,17#

#####
#removal of outliers#
#####

pdf(file="C:\\Users\\Garth Highland\\Desktop\\Appendix.B.1.pdf")

vis=spect[50:380,1:44]
fbplot(vis[,-30],fullout=T,x=wav[50:380,],
ylim=c(0,50),ylab="% reflectance",xlab="wavelength (nm)",
main="Clean Native test set on May 19: visible spectrum")

cell=spect[381:950,1:44]
fbplot(cell[,-c(2,5,16,17,44)],fullout=T,x=wav[381:950],
ylim=c(0,50),ylab="% reflectance",xlab="wavelength (nm)",
main="Clean Native test set on May 19: cellular spectrum")

full=spect[50:950,1:44]
fbplot(full[,-c(5,16,17)],fullout=T,x=wav[50:950],
ylim=c(0,50),ylab="% reflectance",xlab="wavelength (nm)",
main="Clean Native test set on May 19: visible and cellular spectrums")

dev.off()

#####
#definitions#
#####

title=c(
"C:\\Users\\Garth Highland\\Desktop\\5.19.visible.identification.pdf",
"C:\\Users\\Garth Highland\\Desktop\\5.19.cellular.identification.pdf",
"C:\\Users\\Garth Highland\\Desktop\\5.19.both.identification.pdf")

mains=c(
"Identification of OWB test: All 5/19 Native visible spectrum",
"Identification of OWB test: All 5/19 Native cellular spectrum",
"Identification of OWB test: All 5/19 Native full spectrum")

owb1=spect[50:380,45:88]
owb2=spect[381:950,45:88]
owb3=spect[50:950,45:88]

native1=(vis[,-30])
native2=cell[,-c(2,5,16,17,44)]
native3=full[,-c(5,16,17)]

xx1=wav[50:380]
xx2=wav[381:950]
xx3=wav[50:950]

```

```

#####
#identification trials for 5/19/2009 visible spectrum#
#####

out=0
depth=0
bdout=0
bdin=0
pdf(file=title[1])
for(i in 1:44){
use=as.matrix(cbind(native1,owb1[,i]))
X=fboxplot(use,
fullout=T,x=xx1,r=owb1[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main=mains[1])
bdepth=signif(X$depth[length(X$depth)],digits=4)
if(length(X$outpoint>0)){
for(k in 1:(length(X$outpoint))){
if (X$outpoint[k]==ncol(use))
{out[i]=1
bdout=append(bdout,bdepth,after=length(bdout))}
else
{out[i]=0}
}}else{out[i]=0}
bdin=append(bdin,bdepth,after=length(bdin))}
legend(400,50,c("OWB band depth",bdepth))
depth[i]=bdepth
bdepth=0
owbobs=0
X=0
}
dev.off()

#####
#identification trials for 5/19/2009 cellular spectrum#
#####

out=0
depth=0
bdout=0
bdin=0
pdf(file=title[2])
for(i in 1:44){
use=as.matrix(cbind(native2,owb2[,i]))
X=fboxplot(use,
fullout=T,x=xx2,r=owb2[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main=mains[2])
bdepth=signif(X$depth[length(X$depth)],digits=4)
if(length(X$outpoint>0)){
for(k in 1:(length(X$outpoint))){
if (X$outpoint[k]==ncol(use))
{out[i]=1
bdout=append(bdout,bdepth,after=length(bdout))}
else
{out[i]=0}
}}else{out[i]=0}
bdin=append(bdin,bdepth,after=length(bdin))}
legend(731,50,c("OWB band depth",bdepth))
depth[i]=bdepth
bdepth=0
owbobs=0
X=0
}
dev.off()

```

```

#####
#identification trials for 5/19/2009 visible and cellular spectrums#
#####

out=0
depth=0
bdout=0
bdin=0
pdf(file=title[3])
for(i in 1:44){
use=as.matrix(cbind(native3,owb3[,i]))
X=fboxplot(use,
fullout=T,x=xx3,r=owb3[,i],ylim=c(0,50),
ylab="% reflectance",xlab="wavelength (nm)",
main=mains[3])
bdepth=signif(X$depth[length(X$depth)],digits=4)
if(length(X$outpoint>0)){
for(k in 1:(length(X$outpoint))){
if (X$outpoint[k]==ncol(use))
{out[i]=1
bdout=append(bdout,bdepth,after=length(bdout))}
else
{out[i]=0}
}}else{out[i]=0
bdin=append(bdin,bdepth,after=length(bdin))}
legend(400,50,c("OWB band depth",bdepth))
depth[i]=bdepth
bdepth=0
owbobs=0
X=0
}
dev.off()

```