Analyzing and Modeling Canvasback Populations

Anusha Kothapalli and Fei Wang

December 20, 2015

1. Background

Canvasback (*Aythya valisineria*) ducks are the largest North American diving ducks, but also one of the least abundant species. Canvasbacks breed in prairie potholes, and during the winter, on ocean bays. Its sloping profile distinguishes it from other ducks. Males have rust-colored heads, white bodies, and black chests and rears. Female canvasbacks are not very distinctively colored, and juvenile ducks have similar coloring as adult females. Aquatic plants are a large portion of the diet of canvasbacks, which is unusual for diving ducks. Therefore, this is a large interest in being able to hunt these ducks for their meat.

Male

Female

1.1. Motivation

However, there is also an interest to maintain a large enough, breeding-age population so that this species is not depleted to the point of endangerment and/or extinction by hunting. The US Fish and Wildlife Service began a project to predictively model canvasback populations to balance these two interests around 1975, and until about 1991, they did not sanction any hunting. However, in an effort to balance the two interests properly, they began to significantly improve this model and set the hunting regulations accordingly, starting at around 1992. Better prediction of the canvasback populations for the next year(s) would enable better hunting regulations, thereby allowing for the most hunting while also protecting a significant (goal: 500,000) breeding-age populations of canvasback ducks.

1.2. Goals

In the long-term, this project has the potential to significantly improve current efforts by the US Fish and Wildlife Service to conserve canvasback populations while allowing for maximal hunting. The project in the short-term focuses on: - Through data visualization, we aim to make the modeling of Canvasback duck populations more accessible and interpretable - Through model assessment, we hope to suggest improvements to increase the prediction accuracy of the model in the future

2. Alternative Strategy: Multistage Sampling

Since we do not have a sampling frame for canvasback population, the traditional method of simple random sampling is not a good option for estimating the duck population. Therefore, the survey conducted by the US Fish and Wildlife Service employed an alternative strategy called *Multistage Sampling*.

This process of monitoring canvasback populations and determining the appropriate hunting regulations has multiple steps: Firstly, the population is partitioned into primary sampling units (PSUs); each unit is subsequently partitioned, and subsampling occurs. This is repeated over multiple stages. Each stage of sampling is independent of the other sampling stages, and subsampling within one PSU is independent of subsampling within all other PSUs.

In the context of this project, the primary sampling unit is a stratum, a geographical region determined by the US Fish and Wildlife Service (portions of the US & Canada lands where canvasbacks are usually observed). Subsampling occurs within each stratum, by sampling with segments (i.e., subunits of a given stratum). Within each segment, there are many survey ponds (i.e., subunits of a given segment).

Certain ponds of selected segments are surveyed by two observers on a particular date. They approximate the number of canvasbakc by observing the number of single males, pairs of male and female, and groups of canvasback ducks. Single females are rare as they are usually the couples of sigle males and hiden in the nearby reeds. Groups large than 45 are excluded in the estimation as they are assumed to be non-breeding birds. These estimates are all combined to get an estimate of the total population of canvasback ducks for that year.

3. The Predictive Model

After obtaining an estimate of the current population, the future population is estimated using predictive modeling. To estimate the population of canvasback ducks expected in the coming year(s), this prediction model utilizes observed data including an estimate of the current population, estimates of the number of juvenile ducks as opposed to adult ducks present in the population (age ratio), winter and summer survival rates, and crippling-loss rate (an estimate of the number of unreported, accidental duck killings by humans). After obtaining this estimate of the population for the next spring, the hunting policy is set accordingly.

The current model is as follows:

$N_{t+1} = (N_t S + N_t S A_t - (H_t/(1 - CLR))) * S_w$

- N_t : abundance/population size in spring for year t
- N_{t+1} : model output - estimated population size in spring for year t+1

- S : constant summer (May to August) survival rate for adults = 0.936
- A_t : year t estimated production rate = juveniles/adults
- H_t : year t harvest
- CLR : constant crippling-loss rate = 0.30
- S_w : constant winter (February to May) survival rate for adults = 0.926

In the model, N_t represents the estimated total population size in spring in $year_t$. The term N_tS and N_tSA_t give adult and juvenile canvasback population in fall respectively, where S is the summer survival rate. Harvest is adjusted by crippling-loss rate to compensate the uncounted killing. The remaining population is further adjusted by winter survival rate. Hence, it gives the predicted population, N_{t+1}, in spring in $year_{t+1}$.

The possible hunting policies are:

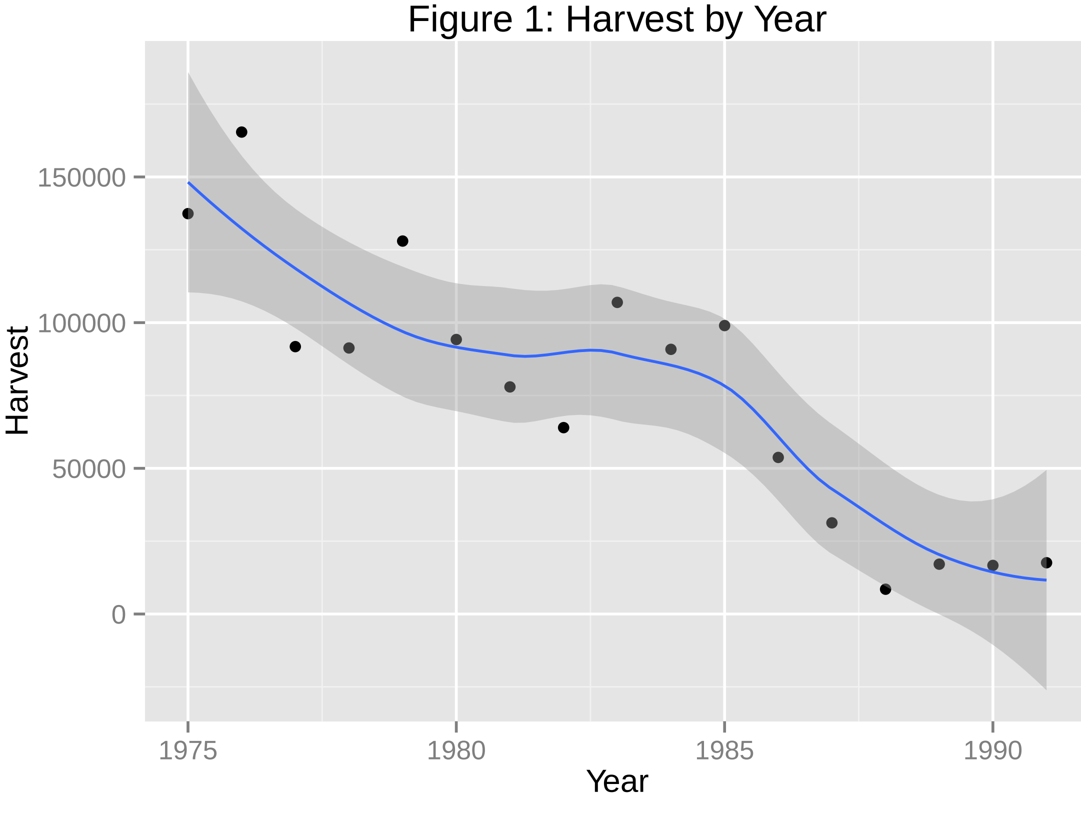
- Closed ("c"), where no hunting is legal (though some occurs, illegally)
- Restricted ("r"), in which some hunting is allowed but quite limited
- Liberal with 1-bird bag ("l"), in which hunters are allowed to kill 1 duck
- Liberal with 2-bird bag ("l2"), in which hunters may kill 2 ducks.

4. Model Assessment through Multivariate Plots

4.1. Harvest

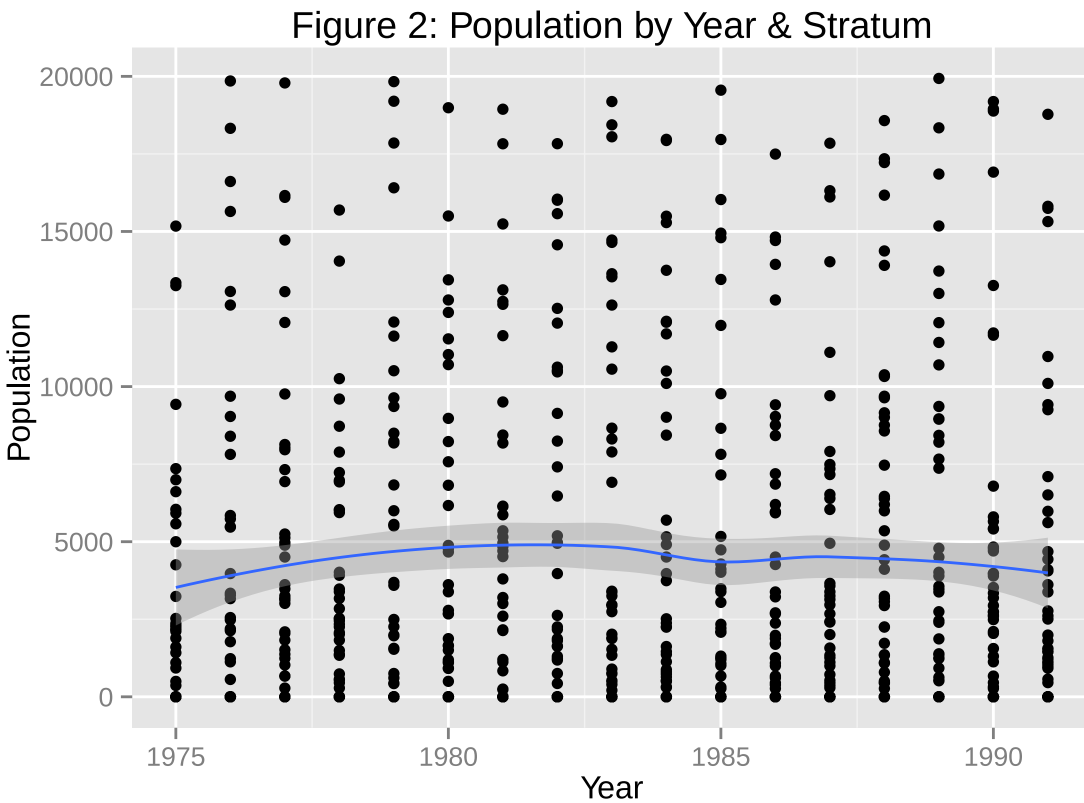
Take a look at canvasback harvest before 1992. As shown in Figure 1, in the early years of the US Fish and Wildlife project, the hunting regulation was always designated closed. As is evident, harvest began to decline.

Figure 1: Harvest by Year



4.2. Population

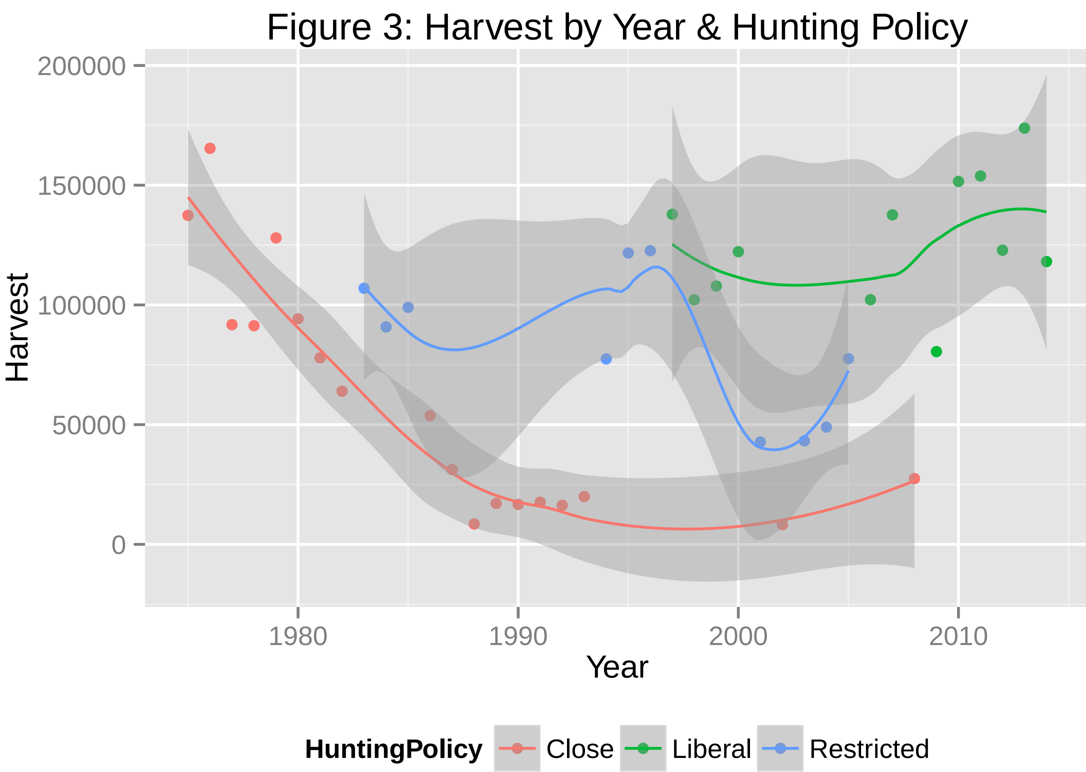
Correspondingly, Figure 2 shows the canvasback populations before 1992. The dots represent the strata in a particular year that had population less than 20,000. The blue line is the trend line for population over the years. It is evident that the canvasback population did not drop substaintially, despite the drastically declining harvest. Therefore, there was no need for such a restrictive ‘closed’ policy to be in place constantly.



The model shown before (see “The Predictive Model”) was the result of rigorous studies, documentation, and policy re-evaluation. The model was instituted as a way to estimate the canvasback population for the coming year’s spring season, based on observed population estimates from the current year. Therefore, a more appropriate hunting policy could be instituted to balance harvest and the need to conserve a canvasback population of (ideally) 500,000 breeding-age ducks.

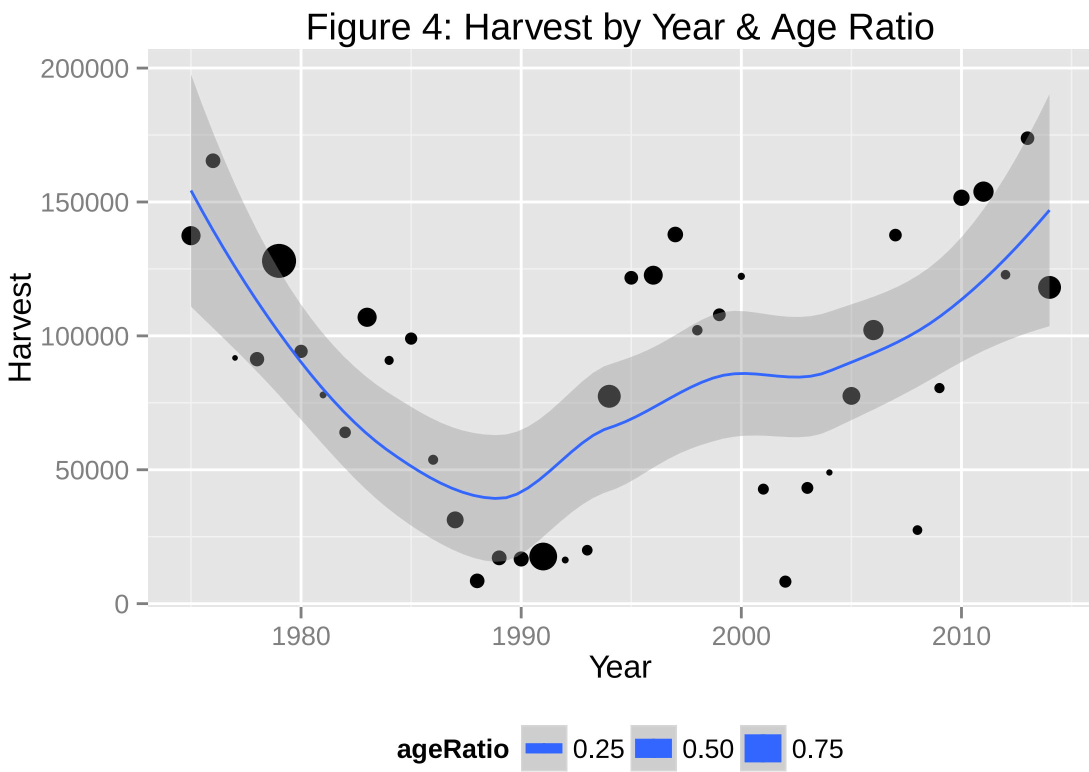
4.3. Hunting Policy

Figure 3 shows us the hunting policy over the years. As is clear in plot, a closed season policy results in declining harvests overall. Recently, as a result of the model improvements, the model seems to be effective, allowing for more years with a liberal or restrictive policy (as opposed to closed). This allows for higher harvests overall.



4.4. Age Ratio

Figure 4 depicts the curve of harvests in relation to age ratio. The size of the points correspond to the value of age ratio. It is important to account for age ratio because juvenile ducks, once shot, cannot be returned to their environment, and therefore do not mature and breed. Juvenile ducks become the breeding population, so it is important that there is more harvesting of older ducks relative to juvenile ducks.



5. Sample Modelling Application

In this section, we provide sample code for generating canvasback population estimates from the model provided by the US Fish and Wildlife Service.

Use 2013 data to estimate 2014 population using the model

```
harvest2013 <- subset(canvasbackHarvest, year == 2013)
estimates2013 <- subset(canvasbackEstimates, year == 2013)

# Actual population data for 2014, observed in 2014
harvest2014 <- subset(canvasbackHarvest, year == 2014)
estimates2014 <- subset(canvasbackEstimates, year == 2014)
```

An R function to generate population estimates

Input observed data from year t, outputs numerical population estimates for year t+1 (using the model that was described before)

```
pop <- function(Nt, At, Ht) {
  est = 0
  Sw = 0.926
  C = 0.30
  Ss = 0.936
  for (i in 1:length(Nt)) {
```

```
    est = est + Nt[i]*Ss + Nt[i]*Ss*At
  }
  popEst = Sw*(est-(Ht/(1-C)))
  return (popEst)
}
```

Generate vectors

This includes population estimates(from model), observed population estimates (from observed data), and harvest (from observed data).

```
# Population estimates (from model)
pop(estimates2013$popEst, harvest2013$ageRatio, harvest2013$harvest)
```

```
## [1] 685286.7
```

```
est <- numeric(40)
est[1] = 0
for (i in 2:length(canvasbackHarvest$year-1)){
est[i] <- pop(subset(canvasbackEstimates,
                     year == canvasbackHarvest$year[i])$popEst,
              subset(canvasbackHarvest,
                     year == canvasbackHarvest$year[i])$ageRatio,
              subset(canvasbackHarvest, year == canvasbackHarvest$year[i])$harvest)
}

# Population estimates (from observed data)
Population <- numeric(40)
for (i in 1:length(canvasbackHarvest$year)){
Population[i] <- sum(subset(canvasbackEstimates,
                            year == canvasbackHarvest$year[i])$popEst)
}

# Harvest (from observed data)
Year <- canvasbackHarvest$year
Harvest <- canvasbackHarvest$harvest
```

Combine dataframe

Combine year, estimated population (from model), harvest, and observed population estimates into dataset

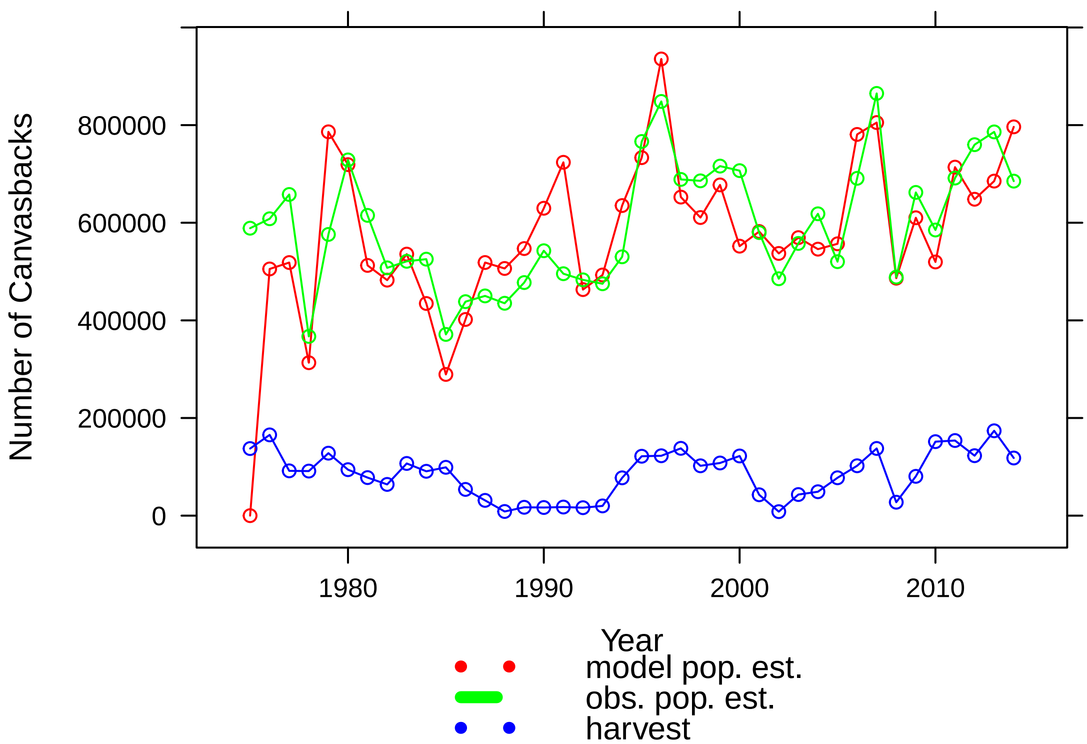
```
canvasbacks <- as.data.frame(cbind(Year, est, Harvest, Population))
```

In the code above, it takes in data observed from the current year, and outputs an estimate of number of canvasbacks (using the model) expected for the next year. This estimate is what is used to decide on a hunting policy. The decision is made a year in advance, which is why a predictive model is used.

The final output is a data frame of the years of this project, 1975-2014, the observed number of harvests (hunting) of canvasbacks, the observed population estimates for that year, and what the model predicted for that year.

Estimated and Predicted Populations

Figure 5: Canvasbacks Estimated and Predicted Populations Over Time



In Figure 5, the model estimate of the population (red line), the observed estimates of the population (green line), and the harvests (blue line) for each given year, from 1975 to 2014, is depicted. The trends over time can be seen. Note that the model and observed population estimates get relatively closer after 1991 or so, indicating that the model improvements that occurred around then were effective. The harvest also stopped declining, and began to rise around then, also indicative of the effectiveness of the new process instituted around 1991-1992.

However, it can also be noted that the model estimates slightly lag behind the observed population estimates. For this reason, we believe a logical next step (and a limitation of our current work) to improve the model is to incorporate error terms, perhaps the standard errors for the observed population estimates.

6. Exploration through Leaflet and Shiny

Our Shiny application truly visualizes this model, through a comprehensive map of the relevant areas of the US and Canada, divided into the designated stratum. The interactive map enables users to slide to the year they wish to explore. They can zoom in and out, and get more information on a particular stratum by clicking on it, which then displays a pop-up with information on that stratum's model/predicted population estimate, observed population estimate, and standard errors for the observed estimate. Users can also view a map of harvests by year, by using the tabs to navigate to the map of the year they wish to see. The link to our Shiny application is: **insert link here!**. Happy exploring!

on boundaries of stratum, which are the US Fish and Wildlife geographical designations of region, which we used in our interactive maps of canvasback populations that can be found in our Shiny application.

7. Limitations and Future Directions

The limitations of our work is that all the values are estimates, so inherently it is any limitations associated with sampling, such as sampling error, and specifically multistage sampling, that limit our work. The model itself is limited in its prediction accuracy, likely in part due to the fact that error terms are not incorporated. Our data only applies to the canvasback populations in North America - specifically, regions of the US and Canada as designated by the US Fish and Wildlife Service. The visualizations in Shiny are still somewhat choppy, and maps are organized by tab. The Shiny app could be made more dynamic if maps were fluid, and changed by a slider, rather than still maps that are individually tabbed by year. The fluid visualizations, that are controlled by a slider, are still choppy. Using packages that allow for higher quality, smoother visuals would greatly improve the Shiny application.

Our project could be improved by looking into improving the estimation of the observer's error, looking into improving the overall sampling procedure, and by looking into incorporating error terms into the model.

8. Conclusion

In conclusion, the model can be made more accessible to the general public by investigating computations of various estimates and creating interactive visualizations. Doing so is the first step to improving this model - and ultimately, efforts to conserve canvasback populations while maximizing hunting opportunity.

9. Codebook

9.1 canvasbackEstimates.csv

- stratum: geographic location
- year: year in which the data was collected
- popEst: estimated canvasback population in strata
- sePopEst: standard error of the canvasback population estimate
- VCF: visibility correction factor to account for observer error
- seVCF: standard error of visibility correction factor

9.2 pondEstimates.csv

- stratum: geographic location
- year: year in which the data was collected
- pondEst: estimated number of ponds in strata
- sePondEst: standard error of pond estimate
- VCF: visibility correction factor to account for observer error
- seVCF: standard error of visibility correction factor

9.3 canvasHarvest.csv

- year: year in which the data was collected
- harvest: estimated number of canvasbacks hunted
- ageRatio: estimated production rate (immature/adults)
- season: hunting designation - open, closed, restricted with 1-bird bag, restricted with 2-bird bags

9.4 segmentCounts-1975-99.csv

- Month,Day,Year: 1975-1999
- Str: stratum
- Seg: segment
- AOU: 1470=canvasback, 9991=pond
- Single: number of single male canvasback
- Pair: number of pairs of male and female
- Grp: 3 or more birds that cannot be divided into pairs and single males, lone females are not counted

9.5 segmentCounts-2000-13.csv

- Year: 2000-2013
- Str: stratum
- Seg: segment
- Species: canv/pond (refering to canvasback)
- Single: number of single male canvasback
- Pair: number of pairs of male and female
- Grp: 3 or more birds that cannot be divided into pairs and single males, lone females are not counted
- GrpLG: large group: flocks greater than 45 are not included in the estimates because these are assumed to be non-breeding birds

10. References

The data used in this project come from the US Fish and Wildlife Service studies of canvasback populations, as part of a larger project that models and predicts waterfowl populations in order to set effective hunting regulations. Dr. Emily Silverman of the US Fish and Wildlife Service generously provided us with the data and resources.

We have the following data files:

- canvasbackEstimates.csv: Estimation of Canvasback population for stratum 1-77 from year 1975 to 2015
- canvasbackHarvest.csv: Estimation of Canvasback harvest and age-ratio in general from year 1975 to 2014, including the information for hunting season scenario
- pondEstimates.csv: Estimation of Canadian ponds for stratum 26-49 from year 1975 to 2015
- segmentCounts_1975-99.csv: Segment surveyed on a particular day, indicating the total number of canvas (single males, pairs or open groups) on that day from year 1975 to 1999
- segmentCounts_2000-13.csv: Segment surveyed on a particular day, indicating the number of canvas (single males, pairs, groups, large groups) on that day from year 2000 to 2013 surveydesign

We also have the following documents, which are reports about the canvasback hunting strategy and model evaluation:

- assessmentReportRevised.4.22.03.pdf: provides the model for age-ratio estimation and population estimation
- Canvasback review document 1993.pdf: provides information/background on development on strategy for determining guidelines
- Smith 1995 Critical Review of Aerial and Ground Surveys of Breeding Waterfowl in NA Reduced Size.pdf: formulas for calculating stratum estimates and variances using raw data & VCFs
- strategy2004.pdf and description.docx: background on canvasbacks and the model and relevant variables, overview of the strategy development process, overall goals and strategies

The following files, especially spatial data in shapefiles, are needed to incorporate geographical locations of the ponds in our interactive application:

- WBPHS2010StrataMap.pdf: general demarcation of stratum boundaries
- shape files: geographical locations for strata boundaries and segment locations

These can all be found on GitHub repository, [AmherstCollege/CanvasBacks](https://github.com/AmherstCollege/CanvasBacks).

Please refer to Appendix B for Codebook

Other References:

- For resources on shapefiles, and how to use them in R, see Shapefiles in R.
- For resources on Leaflet, see Leaflet for R.
- For resources on multistage sampling, see ABC and Model Assisted Survey Sampling.

11. Acknowledgement

We would like to thank:

- Dr. Silverman of the US Fish and Wildlife Service, for all of her guidance and for providing us with the reports and data we used for this project
- Professor Nicholas Horton of Amherst College, for all of his help and guidance throughout this project

Appendix A: R Markdown

Load Data

```
#created a R variables to hold the canvasback harvest & canvasback estimates data
load("canvasbackHarvest.Rda")
load("canvasbackEstimates.Rda")

#load required packages
require(ggplot2)
require(mosaic)
require(readr)
require(tidyr)
require(xtable)
```

Figure 1

```
#Note: season is a variable code for hunting policy
canvasbackHarvest1 <- subset(canvasbackHarvest, year < 1992)

ggplot(data = canvasbackHarvest1, mapping = aes(x = year, y = harvest)) +
  layer(geom = "point") +
  geom_point(alpha=0.001) +
  ggtitle("Figure 1: Harvest by Year") +
  labs(x="Year", y="Harvest") +
  stat_smooth()
```

Figure 2

```
#take a look at canvasback populations before 1992
canvasbackEstimates1 <- canvasbackEstimates %>%
  filter(year < 1992, popEst < 20000)

ggplot(data = canvasbackEstimates1, mapping = aes(x = year, y = popEst)) +
  layer(geom = "point") +
  geom_point(alpha=0.001) +
  ggtitle("Figure 2: Population by Year & Stratum") +
  labs(x="Year", y="Population") +
  stat_smooth()
```

Figure 3

```
# Renmae hunting policy
canvasbackHarvest2 <- rename(canvasbackHarvest, HuntingPolicy = season)
canvasbackHarvest2$HuntingPolicy <- as.character(canvasbackHarvest2$HuntingPolicy)
canvasbackHarvest2$HuntingPolicy[canvasbackHarvest2$HuntingPolicy == "c"] <- "Close"
canvasbackHarvest2$HuntingPolicy[canvasbackHarvest2$HuntingPolicy == "l"] <- "Liberal"
canvasbackHarvest2$HuntingPolicy[canvasbackHarvest2$HuntingPolicy == "r"] <- "Restricted"

#plot for harvest by year & hunting policy
ggplot(data = canvasbackHarvest2, mapping = aes(x = year, y = harvest, color = HuntingPolicy)) +
  layer(geom = "point") +
  geom_point(alpha=0.001) +
```

```
  ggtitle("Figure 3: Harvest by Year & Hunting Policy") +
  labs(x="Year", y="Harvest") +
  stat_smooth() +
  theme(legend.position="bottom")
```

Figure 4

```
#plot for harvest by year & age ratio
ggplot(data = canvasbackHarvest, mapping = aes(x = year, y = harvest, size = ageRatio)) +
  layer(geom = "point") + geom_point(alpha=0.001) +
  ggtitle("Figure 4: Harvest by Year & Age Ratio") +
  labs(x="Year", y="Harvest") +
  stat_smooth() +
  theme(legend.position="bottom")
```

Figure 5

```
# load required package
require(latticeExtra)

# generates individual plots:
# estimated population (from model) over time,
# observed population estimates (from data) over time, harvest (from data) over time

plot1 <- xyplot(est~Year, data=canvasbacks, ylab="Number of Canvasbacks",
                main = "Figure 5: Canvasbacks Estimated and Predicted Populations Over Time",
                par.settings=list(par.main.text=list(cex=1)),
                type=c('l','p'), col="red", key=list(space="bottom",
         lines=list(col=c("red", "green", "blue"), lty=c(3,2), lwd=6),
         text=list(c("model pop. est.", "obs. pop. est.", "harvest"))
))
plot2 <- xyplot(Population~Year, data=canvasbacks,
                type=c('l','p'), col = "green", key=list(space="bottom",
         lines=list(col=c("red", "green", "blue"), lty=c(3,2), lwd=6),
         text=list(c("model pop. est.", "obs. pop. est.", "harvest"))
))
plot3 <- xyplot(Harvest~Year, data=canvasbacks, type=c('l','p'), col="blue",
                key=list(space="bottom", lines=list(col=c("red", "green", "blue"),
                                                    lty=c(3,2), lwd=6),
         text=list(c("model pop. est.", "obs. pop. est.", "harvest"))
))

#combined plot of observed estimates of population, model estimates of population, and harvest over tim
plot1+plot2+plot3
```