VTADS preliminary report

Reed Scott

# Key points

I have done the preliminary analysis you requested on the samples provided by Jake. I realized that we never specified what type of analysis I should do. After some thought, I believe that the single season multi-scale occupancy model is a good fit for this preliminary analysis. In this case I worked under the assumption that = occupancy at a pond, = whether BD/RV was detected *during a survey* and p = whether BD/RV was detected on an individual (prevalence). One thought I have had is that we *could* change the analysis to a dynamic single species model, which may be more fitting to compare between surveys. Also please note that I did not do any type of co-occurrence analysis. I’m happy to do so if you think that would be helpful, I just haven’t yet because it wasn’t on our list of questions.

Below, I provide detail on how I conducted this analysis, however, I wanted to provided a brief summary of my interpretation of the results. For this analysis, we asked three key questions:

1. Did time, which I interpreted as survey number, have an effect on the detection of BD/RV at each site?
2. Did probability of occupancy of BD/RV vary by pond?
3. Did probability of detection of BD/RV vary by species?

In my opinion, preliminary results do not provide strong evidence for any affects of time, pond, or species. So I think we could consider running a subset of the data, such as only looking at one of the survey periods.

This is the short version of what I found, for a more in depth description of my methods, results, and relevant graphs, please read on.

# Prepping the data

I’m going to try to keep this section short, as I think I have a tendency to explain data preparation in more detail than is necessary. However, there are a few key points I wanted to highlight. I think it’s worth mentioning the raw data. First, lets discuss Jake’s data:

dim(Disease\_Data)

[1] 359 3

This data contains a Sample.Name, whether the sample was positive or negative for BD, and whether it was positive or negative for ranavirus. We have 359 samples from this data.

We also have our field data. This has a large number of variables so I’m not going to show them all. However, I would like to highlight the variables we will use for this analysis: PondID, Survey Number, and Species ID. In this preliminary data we had 24 sites, 3 surveys, and 4 species.

So the first thing I had to do was merge the data. Because we have sample ID’s in both dataframes this is a fairly straightforward process. However, because there is some confusion with labeling, only 201 of the 359 samples from Jake had matching SampleIDs.

#remove spaces from disease\_data  
Disease\_Data$Sample.Name <- str\_remove(Disease\_Data$Sample.Name," ")  
field\_data$SampleID <- str\_remove\_all(field\_data$SampleID,"-")  
  
colnames(Disease\_Data)[1] <- 'SampleID'  
  
Disease\_Data$BD.Status[Disease\_Data$BD.Status == 'Negative'] <- '0'  
Disease\_Data$RV.Status[Disease\_Data$RV.Status == 'Negative'] <- '0'  
  
Disease\_Data$BD.Status <- as.numeric(Disease\_Data$BD.Status)  
Disease\_Data$RV.Status <- as.numeric(Disease\_Data$RV.Status)  
  
  
prelim\_data <- Disease\_Data %>%   
 left\_join(field\_data, by = "SampleID")  
  
prelim\_data2 <- prelim\_data %>% filter(!is.na(ComplexID)) #has 201 obs. 359-201 = 158. Which is the amount of samples  
dim(prelim\_data2)

[1] 201 26

From here there’s some data cleaning / organization that I just want to briefly go over. I limited the number of columns to just the factors we’re interested in. I changed the status of BD / RV from zoospres/PFU equivalents to a binary 0/1. I created dummy data (similar to what we did in the initial stages of the NEARMI project) to create all possible combinations of number of visits (max = 3) and number of frogs (max = 14). This way it created an “even” dataframe that could be read in to unmarked. Then I widened the dataframe, and finally broke it in to 3 dataframes (one for site co-variates, one for “observation” co-variates, and one for the detection data) which unmarked requires. Lastly, unmarked does not have a built in variation by so I had to code it in. All of this code can be seen below, and I’m happy to discuss it in more detail. But with this we’re ready to use unmarked.

### Limit number of columns ###  
prelim\_data\_RV <- prelim\_data2 %>% select(c(RV.Status, ComplexID, PondID, SiteID, Survey.Number, CatchOfDay, Species))  
prelim\_data\_BD <- prelim\_data2 %>% select(c(BD.Status, ComplexID, PondID, SiteID, Survey.Number, CatchOfDay, Species))  
  
#Change BD/RV status to 0/1  
prelim\_data\_RV$RV.Status <- ifelse(prelim\_data\_RV$RV.Status > 0, 1,0)  
prelim\_data\_BD$BD.Status <- ifelse(prelim\_data\_BD$BD.Status > 0, 1,0)  
  
# Let's create some dummy data  
Num\_visits <- max(prelim\_data2$Survey.Number)  
Num\_frogs <- max(prelim\_data2$CatchOfDay)  
dummy\_RV <- data.frame(RV.Status = rep(NA, time = Num\_visits\*Num\_frogs),  
 ComplexID = rep("Dummy", time = Num\_visits\*Num\_frogs),  
 PondID = rep("Dummy", time = Num\_visits\*Num\_frogs),  
 SiteID = rep("Dummy", time = Num\_visits\*Num\_frogs),  
 Survey.Number = rep(1:Num\_visits, times = Num\_frogs),  
 CatchOfDay = rep(1:Num\_frogs, times = Num\_visits),  
 Species = rep(NA, time = Num\_visits\*Num\_frogs))  
dummy\_BD <- data.frame(BD.Status = rep(NA, time = Num\_visits\*Num\_frogs),  
 ComplexID = rep("Dummy", time = Num\_visits\*Num\_frogs),  
 PondID = rep("Dummy", time = Num\_visits\*Num\_frogs),  
 SiteID = rep("Dummy", time = Num\_visits\*Num\_frogs),  
 Survey.Number = rep(1:Num\_visits, times = Num\_frogs),  
 CatchOfDay = rep(1:Num\_frogs, times = Num\_visits),  
 Species = rep(NA, time = Num\_visits\*Num\_frogs))  
  
  
#then we can rbind to the rv and bd data  
prelim\_data\_RV <- rbind(prelim\_data\_RV, dummy\_RV)  
prelim\_data\_BD <- rbind(prelim\_data\_BD, dummy\_BD)  
  
#Pivot wider so this can be read in to unmarked  
RV\_wider <- prelim\_data\_RV %>% pivot\_wider(names\_from = c("Survey.Number","CatchOfDay"), values\_from = c(RV.Status, Species))  
BD\_wider <- prelim\_data\_BD %>% pivot\_wider(names\_from = c("Survey.Number","CatchOfDay"), values\_from = c(BD.Status, Species))  
  
#Need to get rid of dummy rows  
RV\_wider <- filter(.data = RV\_wider, !SiteID %in% "Dummy")  
BD\_wider <- filter(.data = BD\_wider, !SiteID %in% "Dummy")  
  
#Okay now we need to order the columns  
#separate in to different dataframes  
RV\_wider\_sitecovs <- RV\_wider[,1:3]  
RV\_wider\_obsdet <- RV\_wider[,4:ncol(RV\_wider)]  
RV\_wider\_obsdet <- RV\_wider\_obsdet[,sort(colnames(RV\_wider\_obsdet))]  
RV\_wider\_dets <- RV\_wider\_obsdet[,1:42]  
RV\_wider\_ObsCovs <- RV\_wider\_obsdet[,43:ncol(RV\_wider\_obsdet)]  
  
BD\_wider\_sitecovs <- BD\_wider[,1:3]  
BD\_wider\_obsdet <- BD\_wider[,4:ncol(BD\_wider)]   
BD\_wider\_obsdet <- BD\_wider\_obsdet[,sort(colnames(BD\_wider\_obsdet))]  
BD\_wider\_dets <- BD\_wider\_obsdet[,1:42]  
BD\_wider\_ObsCovs <- BD\_wider\_obsdet[,43:ncol(BD\_wider\_obsdet)]  
  
# #reordering  
# RV\_wider\_dets <- RV\_wider\_dets[,sort(colnames(RV\_wider\_dets))]  
# RV\_wider\_ObsCovs <- RV\_wider\_ObsCovs[,sort(colnames(RV\_wider\_ObsCovs))]  
#   
# BD\_wider\_dets <- BD\_wider\_dets[,sort(colnames(BD\_wider\_dets))]  
# BD\_wider\_ObsCovs <- BD\_wider\_ObsCovs[,sort(colnames(BD\_wider\_ObsCovs))]  
  
  
#Need to code in variable for survey #  
S <- 3  
SurveyNum\_RV <- data.frame(Survey = matrix(rep(1:S, times = nrow(RV\_wider)),  
 nrow = nrow(RV\_wider),   
 ncol = S,  
 byrow = TRUE))  
SurveyNum\_BD <- data.frame(Survey = matrix(rep(1:S, times = nrow(BD\_wider)),  
 nrow = nrow(BD\_wider),   
 ncol = S,  
 byrow = TRUE))

# Creating models + comparing

I’m not going to explain unmarked in to much detail. The first step is to create an “Unmarked frame” for each disease.

#okayyyy lets get read to read these in to unmarked!  
RV\_frame <- unmarkedMultFrame(y = RV\_wider\_dets,   
 siteCovs = RV\_wider\_sitecovs,   
 yearlySiteCovs = list(SurveyNum = SurveyNum\_RV),  
 obsCovs = list(Species = RV\_wider\_ObsCovs),  
 numPrimary = 3)

Warning: siteCovs contains characters. Converting them to factors.

Warning: obsCovs contains characters. Converting them to factors.

BD\_frame <- unmarkedMultFrame(y = BD\_wider\_dets,   
 siteCovs = BD\_wider\_sitecovs,   
 yearlySiteCovs = list(SurveyNum = SurveyNum\_BD),  
 obsCovs = list(Species = BD\_wider\_ObsCovs),  
 numPrimary = 3)

Warning: siteCovs contains characters. Converting them to factors.  
Warning: obsCovs contains characters. Converting them to factors.

Then, you can start creating models. It’s really quite simple! First I created the null model for each disease

Null\_RV <- goccu(psiformula = ~1, phiformula = ~1, pformula= ~1, RV\_frame)  
Null\_BD <- goccu(psiformula = ~1, phiformula = ~1, pformula= ~1, BD\_frame)  
  
plogis(coef(Null\_RV))

psi(Int) phi(Int) p(Int)   
0.9998651 0.9367100 0.3710238

plogis(coef(Null\_BD))

psi(Int) phi(Int) p(Int)   
0.9999933 0.8389172 0.3598025

I then assessed availability of each disease by survey. Here, I will show the AIC tables so you can assess for yourself.

#"availability" by survey  
surveyMod\_RV <- goccu(psiformula = ~1, phiformula = ~SurveyNum, pformula= ~1, RV\_frame)  
surveyMod\_BD <- goccu(psiformula = ~1, phiformula = ~SurveyNum, pformula= ~1, BD\_frame)  
  
#fitting models on survey  
surveyfit\_RV <- fitList('null' = Null\_RV, "SurveyEffect" = surveyMod\_RV)  
modSel(surveyfit\_RV)

nPars AIC delta AICwt cumltvWt  
null 3 261.31 0.00 0.68 0.68  
SurveyEffect 4 262.83 1.52 0.32 1.00

surveyfit\_BD <- fitList('null' = Null\_BD, "SurveyEffect" = surveyMod\_BD)  
modSel(surveyfit\_BD)

nPars AIC delta AICwt cumltvWt  
null 3 233.43 0.00 0.71 0.71  
SurveyEffect 4 235.25 1.82 0.29 1.00

Then I fit occupancy by pond.

# Occupancy by pond  
PondMod\_RV <- goccu(psiformula = ~SiteID, phiformula = ~1, pformula = ~1, RV\_frame)  
PondMod\_BD <- goccu(psiformula = ~SiteID, phiformula = ~1, pformula = ~1, BD\_frame)  
  
# plogis(coef(PondMod\_RV))  
# plogis(coef(PondMod\_BD))  
  
#fitting models on pond  
occufit\_RV <- fitList('null' = Null\_RV, "PondEffect" = PondMod\_RV)  
modSel(occufit\_RV)

nPars AIC delta AICwt cumltvWt  
null 3 261.31 0.00 1.0e+00 1.00  
PondEffect 24 303.30 41.99 7.6e-10 1.00

occufit\_BD <- fitList('null' = Null\_BD, "PondEffect" = PondMod\_BD)  
modSel(occufit\_BD)

nPars AIC delta AICwt cumltvWt  
null 3 233.43 0.00 1.0e+00 1.00  
PondEffect 24 275.43 42.00 7.6e-10 1.00

And lastly, I fit detection by species

#detection by species  
SppMod\_RV <- goccu(psiformula = ~ 1, phiformula = ~1, pformula = ~ Species, RV\_frame)  
SppMod\_BD <- goccu(psiformula = ~ 1, phiformula = ~1, pformula = ~ Species, BD\_frame)  
  
#fitting models on species  
det\_fit\_RV <- fitList('null' = Null\_RV, "SpeciesEffect" = SppMod\_RV)  
modSel(det\_fit\_RV)

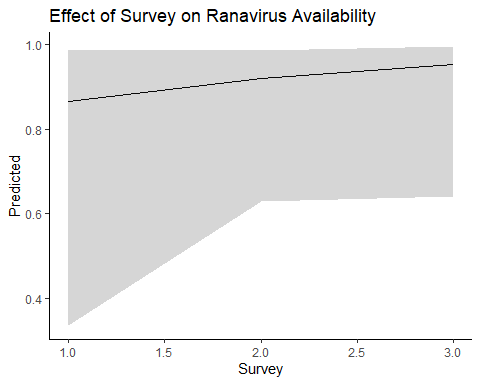
nPars AIC delta AICwt cumltvWt  
null 3 261.31 0.00 0.907 0.91  
SpeciesEffect 6 265.87 4.56 0.093 1.00

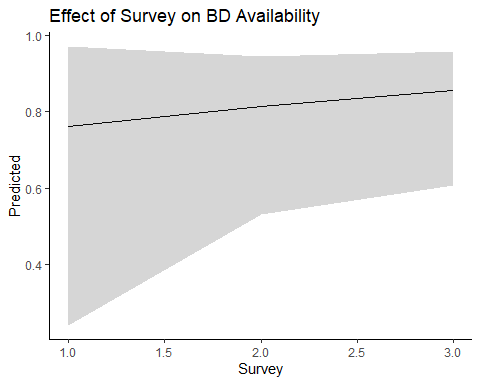
det\_fit\_BD <- fitList('null' = Null\_BD, "SpeciesEffect" = SppMod\_BD)   
modSel(det\_fit\_BD)

nPars AIC delta AICwt cumltvWt  
null 3 233.43 0.00 0.61 0.61  
SpeciesEffect 6 234.32 0.89 0.39 1.00

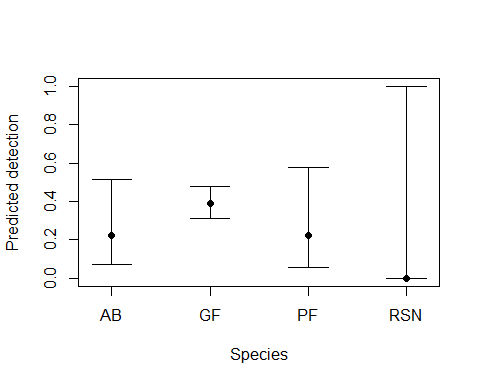
# Plots

There are just a few plots I would like to show. First, since the survey model was relatively close ( AIC scores within 2 points of each other) I decided to graph the effect of survey on availability of each disease





Additionally, the models for detection showed that, for BD, species may have had an effect. So I decided to plot that.



# Conclusions

Based on the data provided, it is worth considering looking at only one survey period. I think this warrants some discussion, and want to be sure I’m interpreting this analysis correctly. I also think it’s worth considering other factors, such as temperature, and how they may vary with survey period. I personally think that the potential species effect shown in BD means that we should at minimum continuing sampling all species.

This is what I have done for now. I would love to hear your thoughts!