# MT5767 Project 1

Lachlan MacLean, Misha Tseitlin, Liam Gerrity

23/10/2023

## Question 1

### Part a)

For this BAS model, we first look at the potential for stochastic components modelling the sub-process abundances. For survival, we use a binomial distribution to model this sub-process abundance for the first three age classes, with parameters $n_{i,t-1}$ and $\phi_i$ for $i = 1, 2, 3$, respectively. The last age class abundance for this sub-process is deterministically 0, as we are told that: "Fourth year individuals always die". Next we look at the ageing sub-process. As all animals must necessarily age between time periods (years), we know this must be a deterministic process where all individuals from the previous sub-process for each age class move into to age class above, with the final fourth year individuals staying in this absorbing state. However, this absorbing state is trivial as none of the fourth year individuals have survived into the time period. Finally for reproduction, we choose a Poisson distribution for the first sub-process age class, whilst keeping all other age classes deterministically the same as in the previous sub process. A Poisson distribution is chosen as we know one of the reproduction rates is greater than 1 ($\rho_3 = 1.9$) so a binomial distribution is not appropriate.

The only other stochastic part of this model are the observations. Given we are told the probability of detection is a static $p = 0.5$ and that no double counting can occur, a binomial distribution for the observations of each age class seems appropriate. Here for time $t$, we use the abundance for class $i$ at time $t$, $n_{i,t}$, for the number of trials and $p = 0.5$ for the probability of success.

### Part b)

In this part we functionalise the theory from a) in order to simulate the dynamics of this model.

```
# b)

# Write an R script to simulate dynamics from this model.
# we do this in a function.

# INPUT :
# n0 = initial populations
# phi = survival probabilities
# rho = reproduction rates
# p = probability of detection
# nyears = number of years over which population is projected
# PROCESS :
# calculate the stochastic population dynamics and observations for each year
# OUTPUT :
# y = array of observations for each year, across all age classes
# n = array of abundances for each year, across all age classes

BAS_stoch <- function(n0, phi, rho, p, nyears) {
```

```r
# initialise

# matrix of all abundances, each column is a year, row is age class
n <- matrix(data = NA, nrow = length(n0), ncol = (nyears+1))

# matrix of all observations
y <- n #replicate n

# initial abundance
# let the first set of abundances (at t = 0) be the initial population size
n[,1] <- n0

# initial observation
# ASSUMPTION: binomial distribution as constant probability of detection (see Newman)
y[1,1] <- rbinom(n = 1, size = n[1,1], prob = p)
y[2,1] <- rbinom(n = 1, size = n[2,1], prob = p)
y[3,1] <- rbinom(n = 1, size = n[3,1], prob = p)
y[4,1] <- rbinom(n = 1, size = n[4,1], prob = p)

# loop over all (other) years
for (i in 2:(nyears+1)) {

  # calculate stochastic sub-processes for BAS model

  # Survival process
  u_1s1t <- rbinom(n = 1, size = n[1,i-1], prob = phi[1])
  u_1s2t <- rbinom(n = 1, size = n[2,i-1], prob = phi[2])
  u_1s3t <- rbinom(n = 1, size = n[3,i-1], prob = phi[3])
  u_1s4t <- 0

  # Ageing process
  u_2a1t <- 0
  u_2a2t <- u_1s1t
  u_2a3t <- u_1s2t
  u_2a4t <- u_1s3t + u_1s4t

  # Reproduction/Birth process
  u_3b1t <- rpois(n = 1, lambda = rho[1] * u_2a2t) +
            rpois(n = 1, lambda = rho[2] * u_2a3t)
  u_3b2t <- u_2a2t
  u_3b3t <- u_2a3t
  u_3b4t <- u_2a4t

  # new abundances
  n[,i] <- c(u_3b1t, u_3b2t, u_3b3t, u_3b4t)

  # new observations
  y[1,i] <- rbinom(n = 1, size = n[1,i-1], prob = p)
  y[2,i] <- rbinom(n = 1, size = n[2,i-1], prob = p)
  y[3,i] <- rbinom(n = 1, size = n[3,i-1], prob = p)
  y[4,i] <- rbinom(n = 1, size = n[4,i-1], prob = p)
}
```

```
  # return
  return(list(n=n,y=y))


}

# parameters from specification
phi <- c(0.45, 0.7, 0.7)
rho <- c(0.9, 1.9)
p <- 0.5
n0 <- c(150, 70, 50, 30)
```

## Part c)

Finally, we run the function over 25 years. Storing the values for the population dynamics and observations, we then convert this into an appropriate format. This data can then be used to produce clear visualisations in the form of line plots.

```
# c)

# Simulate 25 years of age-specific population dynamics and observations and
# produce an informative visualisation of the data.

# specify how many years to project over
nyears <- 25

# run function for 25 years
BAS_proj <- BAS_stoch(n0 = n0, phi = phi, rho = rho, p = p, nyears = nyears)

# visualise the data using a faceted ggplot.

# convert both outputs to dataframes
proj_n_df <- as.data.frame( cbind(0:nyears, t(BAS_proj$n),
                    rep(2,(nyears+1))) )
proj_y_df <- as.data.frame( cbind(0:nyears, t(BAS_proj$y),
                    rep(1,(nyears+1))) )
# rename
colnames(proj_n_df) <- c("Year", "First Year", "Second Year", "Third Year",
                        "Fourth Year", "State")
colnames(proj_y_df) <- c("Year", "First Year", "Second Year", "Third Year",
                        "Fourth Year", "State")
# pivot longer
proj_n_df_long <- pivot_longer(data = proj_n_df,
                            cols = c("First Year", "Second Year",
                                    "Third Year", "Fourth Year"))
proj_y_df_long <- pivot_longer(data = proj_y_df,
                            cols = c("First Year",  "Second Year",
                                    "Third Year", "Fourth Year"))
proj_df_long <- rbind(proj_n_df_long, proj_y_df_long)
colnames(proj_df_long) <- c("Year", "State", "Age Class", "Abundance")

# change variable types
proj_df_long$State <- as.factor(proj_df_long$State)
proj_df_long$`Age Class` <- as.factor(proj_df_long$`Age Class`)
```

```
# Can now easily use faceted ggplot
plot_BAS <- ggplot(data = proj_df_long, aes(x = Year, y = Abundance)) +
  geom_line(aes(colour = State), size = 1) +
  scale_colour_manual("State", breaks = c(2, 1),
                      values = c("blue", "grey"),
                      labels = c("Dynamics","Observations")) +
  facet_wrap(~factor(`Age Class`, levels = c("First Year", "Second Year",
                                             "Third Year", "Fourth Year")),
             scales = "free_y") +
  ylab("Abundance") +
  ggtitle("Population Dynamics and Observations Projected over Time")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
plot_BAS
```
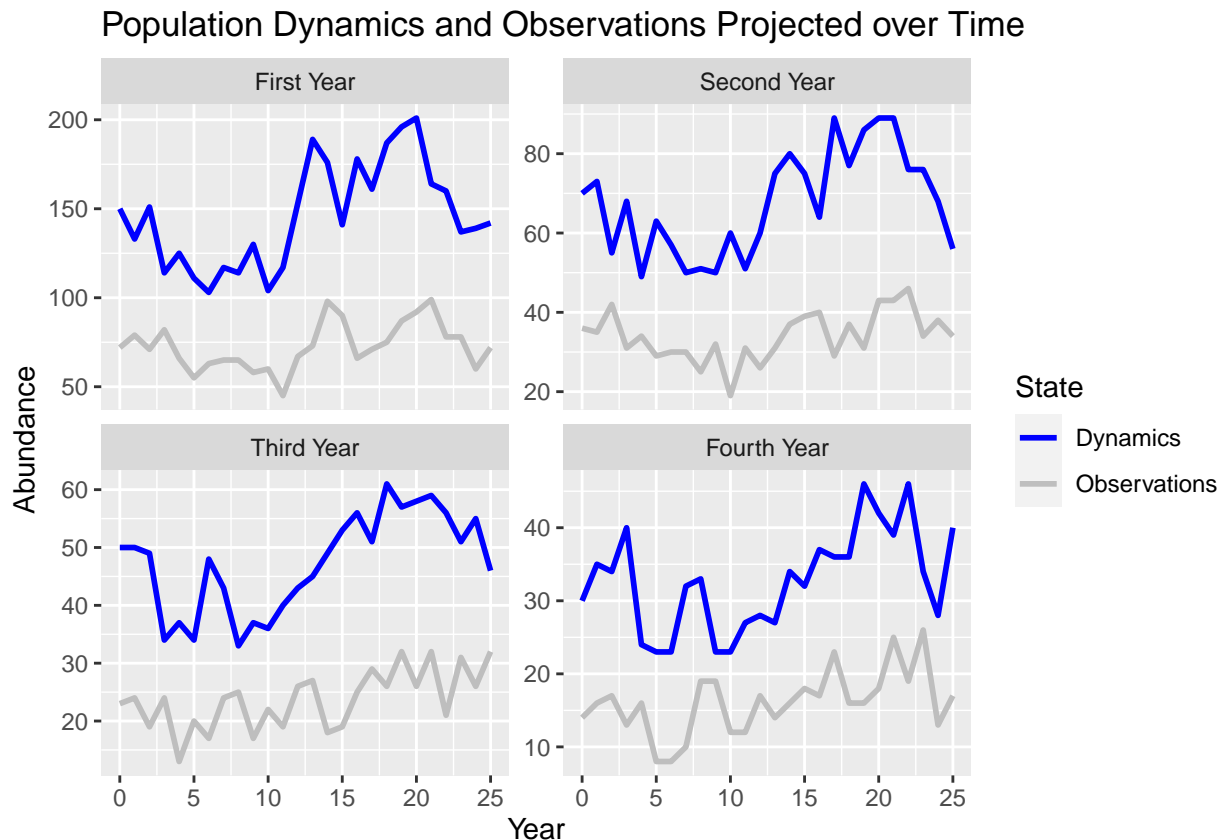


## Question 3

We will fit models to the data using values for growth rate, $r$, and carrying capacity, $K$, dependent on the recorded rainfall, $R$, according to the data set. We will then look to test the impact of time lag when fitting models to the observed data by finding the coefficients of $r$ and $K$ with respect to $R_t$ and $R_{t-1}$.
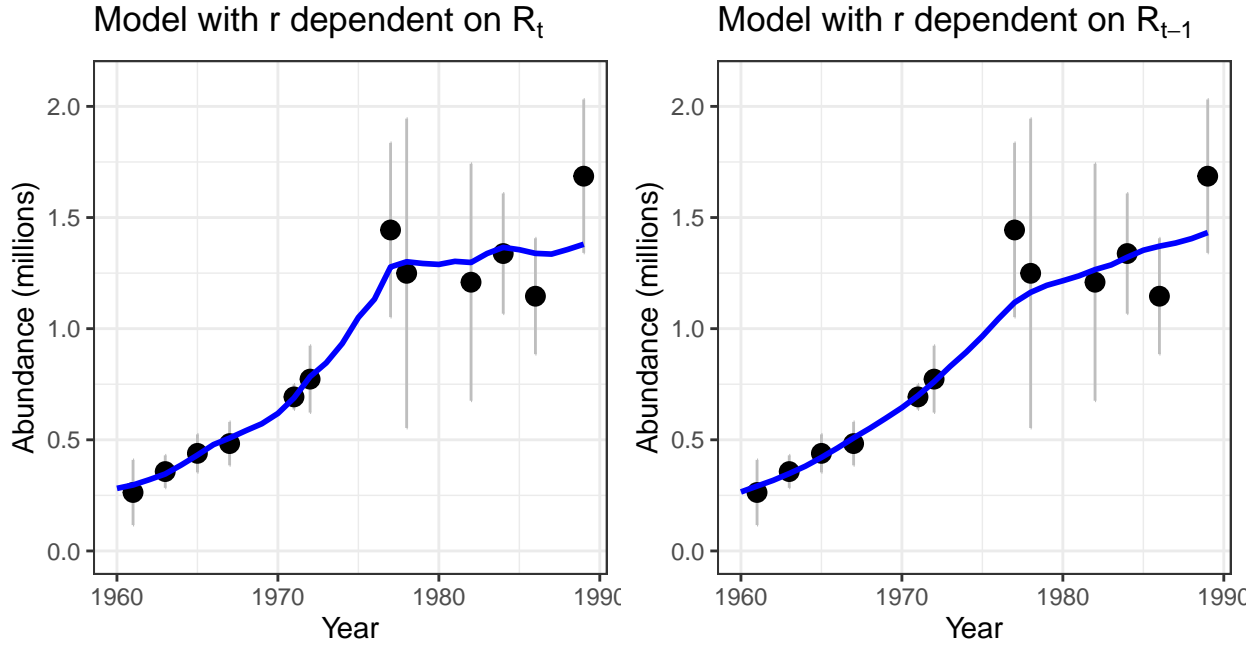
**modelling $r$ dependent on $R_t$ and $R_{t-1}$**

The previously used model has now been extended to include temporal variation in $r$:

$r = exp(\alpha_0 + \alpha_1 R_t)$

$r = exp(\alpha_0 + \alpha_1 R_{t-1})$

By finding the likelihood of the model and using an optimisation function, parameter values of $\alpha_0$ and $\alpha_1$ can be found to calculate $r$ for each time period. Each model can then be used to estimate N as shown by Figure ???. The AIC of each model is displayed in Table ??? and used for model comparison.



| | ALPHA 0 | ALPHA 1 | AIC |
|---|---|---|---|
| r ~ Rt | -2.984800 | 0.6323085 | -18.96679 |
| r ~ Rt-1 | -2.344128 | 0.1565288 | -17.14583 |

By comparing AIC values, the first model $r = exp(\alpha_0 + \alpha_1 R_t)$ is preferabe as it produces the lowest AIC value.

**Modelling $K$ dependent on $R_t$ and $R_{t-1}$**

Using the same optimization methods used previously when modelling the temporal variation of $r$, we can now extend the model to include temporal variation of $k$. The two formulations of K modeled are:

$K = exp(\beta_0 + \beta_1 R_t)$

$K = exp(\beta_0 + \beta_1 R_{t-1})$

Again through the use of optimisation, coefficient value for $\beta_0$ and $\beta_1$ were used to model estimates for $N$, plotted in Figure ???, alongside the respective AIC values displayed in table ???.
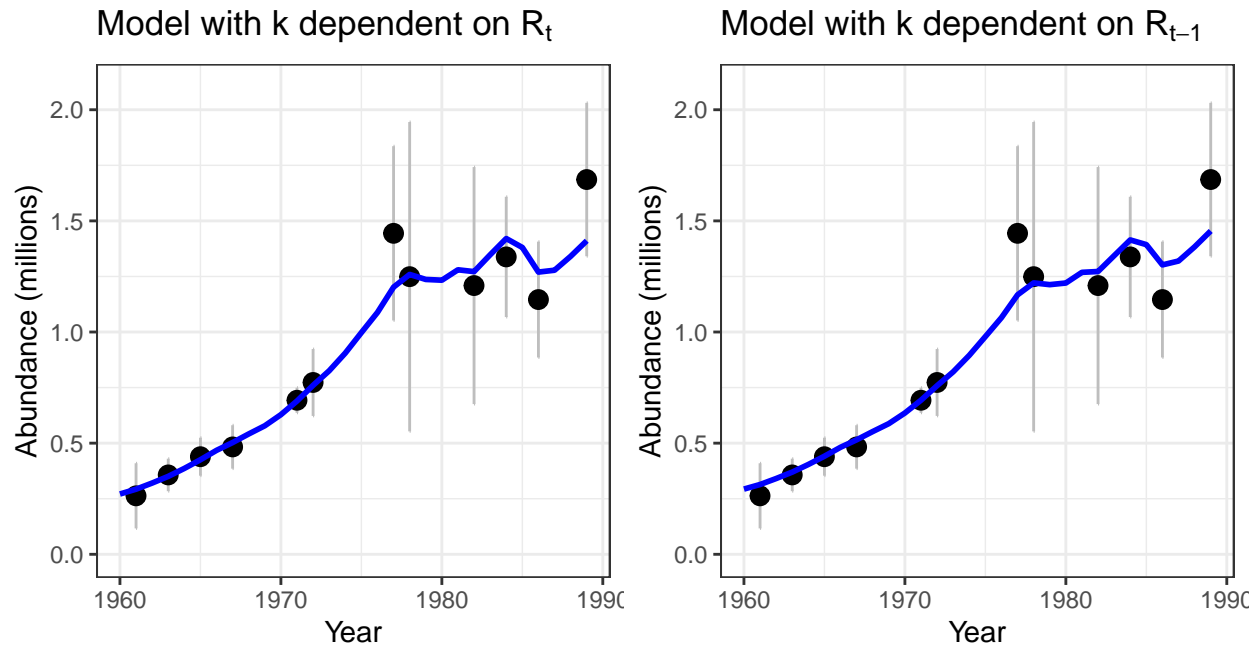
|          | BETA 0     | BETA 1    | AIC       |
|----------|------------|-----------|-----------|
| k ~ Rt   | -0.4969040 | 1.020015  | -19.89674 |
| k ~ Rt-1 | -0.5030827 | 1.194308  | -14.42609 |

From comparing the AIC values of the competing models, the optimal choice appears to use $K$ as a function of $R_t$.

**The Optimal Model?**

By comparing the AIC scores of all 4 models, it would appear that the optimal model according to the lowest score is the one which uses K dependent on $R_t$.

**modelling both temporal $r$ and $K$**