DYNAMIC OCCUPANCY MODELING

One of the natural extensions of single-season single-species occupancy models formulated by Mackenzie et al. (2002) are *dynamic occupancy models*. These models rely on a similar set of assumptions as single-season models (e.g., independence of observations, no false detection, each site is 'closed' within the season), but relaxes the population closure assumption by explicitly estimating probabilities of colonization ad exinction between successive years. The basic idea is the same: survey a set of sites several times during a set *season*. However, one of the interesting questions from a metapopulation perspective is whether the sites are subject to dynamics characterized by local extinction and colonization from one year (or season) to another. *Dynamic occupancy models* address this question by explicitly estimating rates of extinction and colonization, as well as turnover rates, between successive seasons. This is particularly important for species that are characterized by spatial and temporal turnover in occupancy, such as amphibians or territorial species.

As such, *dynamic occupancy models* are used for estimating the change in probability of occupancy based on observed extinction and colonization rates, while accounting for imperfect detection. These quantities can be taken a step further and indirectly get at population growth rates and persistence. In simple terms, the persistence of a metapopulation (or fraction of occupied patches) is dependent on the ratio between colonization and extinction rates. When extinction is higher than colonization, the fraction of occupied patches decreases, leading to extinction of the population over some period of time. This can be also thought in terms of lambda (growth rate) for the entire (meta)population, so *dynamic occupancy models* can be used to infer population dynamics based on observed rates of exctintion and colonization while addressing imperfect detection (i.e., wherether a site with no detections in subsequent years is truly extinct).

To better understand *dynamic occupancy models* we will use data simulations and implementing simple models to test hypotheses (e.g., whether colonization and extinction vary by year)

##LET'S START CODING!

We will first set up paramaters needed to simulate a multi-season occupancy dataset.

```
nyear=5       # number of years to simulate
n_sites=50    # number of sites to simulate
n_vis=4       # number of visits per year
psi1=0.66     # initial occupancy (in first survey year)
p=0.8         # detection probability
ext=0.2       # extinction rate
col=0.4       # colonization rate
```

Similarly to single-season occupancy, we will set up an empty matrix of 'true' occupancy for each of the season we are simulating (**STATE PROCESS**), and an empty matrix of observations (**OBSERVATION PROCESS**)

```
# true occupancy matrix has #rows =  #sites and #columns = #years
tocc <- matrix(rep(0, nyear*n_sites), ncol=nyear)
tocc <- as.data.frame(tocc)
names(tocc) <- paste("t", 1:nyear, sep="")
tocc

# observations matrix has #rows = #sites and #columns = #years * #surveys per season
obsocc <- matrix(rep(0, nyear*n_sites*n_vis), ncol=(nyear*n_vis),
                 dimnames=list(paste("site", 1:n_sites, sep=""),
                 paste("Yr", rep(1:nyear, each=n_vis), "v", 1:n_vis, sep="")))
```

```
obsocc <- as.data.frame(obsocc)
obsocc
```

Next, we simulate TRUE occupancy and OBSERVATED data matrices. This is a nested FOR LOOP with the first loop iterating through sites, and second loop through seasons. In other words, this FOR LOOP fills in data for Site 1 for Year 1, than calculates the outcome for Year 2 based on the dynamics between occupancy, colonization and extinction: - if the Site is unoccupied, then it can be colonized with probability r - if the site is occupied, it can go extinct with probability phi

The same procedure is applied to year 3, 4, 5 and so forth. The FOR LOOP then moves to Site 2 and applies the same procedure, then Site 3, Site 4, and so forth.

```
for(i in 1:n_sites) {
  # First simulate TRUE and OBSERVED simulate data for Year 1
  # (same as for single-season occupancy)

  # Simulate TRUE site occupancy In Year 1 with prob psi1
  tocc[i,1] = rbinom(1, 1, psi1)

  # Simulate OBSERVED presence/absence data in Year 1 using detection prob p
  obsocc[i,1:n_vis] = rbinom(n_vis, 1, tocc[i,1]*p)

  # Then simulate TRUE occupancy data for Years 2 through nyear based on occupancy in
  # Year 1 and probabilities of extinction and colonization
  for(t in 2:nyear) {

    # if site is occupied at Year T, occupancy at T+1 depends on extinction only
    if (tocc[i,t-1] == 1) {
      tocc[i,t] <- rbinom(1,1,(1-ext))    # we need to use (1-extinction)

    # # if site is not occupied at T, it can be colonized with probability col at T+1
    } else {
      tocc[i,t] <- rbinom(1,1,col)
    }

    # Finally, for each year, the observed data incorporates detection probability
    for(j in 1:n_vis) {
      obsocc[i,((t-1)*n_vis)+j] <- rbinom(1,1,tocc[i,t]*p)
    }
  }
}
```

Let’s check out the OBSERVED data and examine the simulated naive occupancy

```
str(obsocc)
naive_occ <- sum(apply(obsocc[,1:4], 1, sum) > 0) / nrow(obsocc)
naive_occ
```

In order to run models, we want to create several variables that we could use to model *extinction* and *colonization* rates, we need to create a matrix of Site * Year covariates containing “Year” as a factor. For example, column 1 in the matrix with have the value “Yr1”, the second column will be “Yr2” and so forth. Note that *extinction* and *colonization* rates express **transitions** between years, so we will estimate 1 fewer parameters. As such, the last column in the newly created matrix is not used, but still has to be present. We will populate this last column with “Junk”, just for fun!

```r
years <- matrix(rep(seq(1:(nyear-1)), n_sites), ncol=nyear-1, byrow=T,
                dimnames=list(paste("site", 1:n_sites, sep=""),
                paste("Yr", rep(1:(nyear-1)), sep="")))

# we need to add a column that will never be used
yrs.Junk <- matrix(rep("Junk", n_sites*1), ncol=1)
years <- cbind(years,yrs.Junk)
colnames(years) <- paste("Yr", rep(1:nyear), sep="")
years <- as.data.frame(years)
years <- data.frame(lapply(years, as.factor))

### Combine the observed P/A data, the yearly covariates, treatments (Years)
final.dataset <- cbind(obsocc, years)
```

Next, we will analyze the simulated data in package *unmarked*, using the function that fits dynamic occupancy models, *colext*. But first, we need to build the object that can be recognized by *unmarked* as an occupancy dataset. It contans the observed data (y), site covariates for modeling occupancy in Year 1 (not in our case), survey (or observation) specific covariates, and *yearly site covariates* for modeling extinction and colonization rates (see *years* matrix above).

```r
# create the y matrix (OBSERVATIONS) columns 1 to 20 in our case
yMat = final.dataset[,1:(nyear*n_vis)]

# prepare the matrix for yearly covariates for modeling colonization and extinction
# columns 21-25 in our case
yearMat = final.dataset[,((nyear*n_vis)+1):((nyear*n_vis)+nyear)]
yearMat

# we have to transpose this matrix to be read in unmarked
yearly_covs = data.frame(YEARS = matrix(t(yearMat)))
yearly_covs
```

Prepare the unmarked multi-year colext data.frame

```r
oUMF <- unmarkedMultFrame(
  y = yMat,
  # site covariates for modeling occupancy in Year1 - NOT USED HERE
  # siteCovs = 1,
  # yearly site covariates for modeling extinction and colonization
  yearlySiteCovs = yearly_covs,
  numPrimary=nyear)

summary(oUMF)
plot(oUMF)
```

It is finally time to run the models on the simulated data. Similarly to the single-season occupancy simulations, will run a Null model, and an Alternative model with covariates for *extinction* and *colonization*, in which these two rates vary by year (i.e., the *year* matrix that we just put together)

The model contains 4 components, *IN THIS ORDER*: - psiformula for initial occupancy - gammaformula for colonization probability - epsilonformula for extinction probability - pformula for detection probability

```r
Null.fit <- colext(psiformula = ~1, gammaformula = ~1,
                   epsilonformula = ~1, pformula = ~1, oUMF)
summary(Null.fit)


Alt.fit <- colext(psiformula = ~1, gammaformula = ~YEARS-1,
                  epsilonformula = ~YEARS-1, pformula = ~1, oUMF)
summary(Alt.fit)
```

We can now predict extinction and colonization for N-1 years (note that we only have 4 transitions for 5 years of data), and detection probabilities for each year.

```r
# first create a new object that fits the predctions (in terms of number of years)
nd <- data.frame(YEARS=c('01','02','03','04'))

E.ext <- predict(Alt.fit, type='ext', newdata=nd)
E.ext

E.col <- predict(Alt.fit, type='col', newdata=nd)
E.col

# predict detection probabilities by year
nd <- data.frame(YEARS=c('01','02','03','04','05'))
E.det <- predict(Alt.fit, type='det', newdata=nd)
E.det
```

Lastly, we may want to know what the occupancy outcome for years 2 - 5 based on the interplay between extinction and colonization rates (under the assumption of stationarity). The occupancy probabilities for each year *psi* are not calculated directly in this parameterization (we only have a component for explaining occupancy in year 1 in this model). Instead, calculating *psi* draws on metapopulation theory, and it is expressed as: **psi = colonziation / (colonization + extinction)**

The nexst piece of code computes occupancy probabilities for years 2,3,4,5 given the estimated extinction and colonization rates using a direct or smoothed approach. We also want to know the uncertainity around the derived occupancy estimates, which we can calculate using non-paramtetric bootstrapping, as well as the number of sites occupied and unoccupied when accounting for imperfect detection.

```r
# predict occupancy probabilities for years 2,3,4,5
E.psi <- projected(Alt.fit)
E.psi

E.psi.smoothed <- smoothed(Alt.fit)
E.psi.smoothed

## Find bootstrap standard errors for smoothed trajectory
Alt.fit <- nonparboot(Alt.fit, B = 100)  # This takes a while!
Alt.fit@smoothed.mean.bsse

# Find the estimates of number of sites occupied in each year
round(E.psi.smoothed*n_sites)
```

END OF SIMULATIONS!

To fit a *dynamic occupancy model* to real data, you must follow the same steps as with the single-season model. First, evaluate and identify the best variables that explain **detection probability**. Detection is key, especially when dealing with dynamic occupancy: if species is not detected at a site at time 2, but was detected at time 1, then the observation could be an outcome of either be *extinction* or *non-detection*. If there is no unmodeled heterogeneity in detection, then these models can correcty predcit that an *extinction* event has occurred, which will influence the predicted dynamics in the system. If detection is not modeled appropriately, then we would incorrectly assume *extinction*, which will underestimate the likelihood of persistence n this (meta)population.

LA COMMEDIA E FINITA!