Classification - Part 2

$$Tradeoff_{t=i} = LRRD_{trt1,t=i}/LRRD_{trt2,t=i}$$

$$= (LRR_{trt1} - LRR_{control})/(LRR_{trt2} - LRR_{control})$$

$$= \frac{log(X_{trt1,t=i}/X_{trt1,t=0}) - log(X_{control,t=i}/X_{control,t=0})}{log(X_{trt2,t=i}/X_{trt2,t=0}) - log(X_{control,t=i}/X_{control,t=0})}$$
$$= \frac{(log(X_{control,t=i} + TRT1)/X_{control,t=0}) - log(X_{control,t=i}/X_{control,t=0})}{(log(X_{control,t=i} + TRT2)/X_{control,t=0}) - log(X_{control,t=i}/X_{control,t=0})}$$
$$= \frac{A - C}{B - C}$$

Fitting Multi-state (Markov) models with MSM

Multi-state Markov models might be a relevant way to analyze this data, where repeated observations of a single subject are quantified as a set of discrete states with temporal transitions in state assignment described by a matrix of transition properties. Also possible with these models is the addition of covariates that may change transition probabilities, rather than operating with only a fixed probability for each given state change.

Raw matrix of transition probabilities

The simplest way to assess differences in transition probability is to just plot the total number of transition events observed within the dataset. This matrix, shown below, describes the general pattern visualized earlier in the plotting of states over time:

- Both the "natives" and "WAPS" states are relatively stable over the time period we observed.
- All states were quite likely to shift to the "dry annuals" state (likely the result of the influx of Avena / Bromus diandrus during the peak of the drought)
- "Priority Annuals", which did very well once initially planted, were quick to be lost from the dataset.

```
##          to
## from      Natives WetA WAPS DryA
##   Natives      96    8    7   29
##   WetA         10   50   30   29
##   WAPS         25   11  115   22
##   DryA         19   21    7   76
```

Fitting state-transition models:

Fitting these models is relatively simply using the MSM package – the function attempts to find the best values of the transition matrix that predict the observed frequencies of each state over time.

MSMs also allow for the use of likelihood ratio tests – quantitative measures for how additional covariates in the model improve fit. However, one important caveat is that each element of the transition matrix is fit by a unique parameter (12 total for a 4x4 matrix, as diagonal entries are equal to 1 - sum of all transition probabilities), so the addition of a covariate must change a number of unique parameters, greatly reducing the residual degrees of freedom.

To assess whether priority effects impact transition probabilities, we can include planting mixture effects on transition probabilities. To reduce the total number of degrees of freedom associated with each covariate, I've restricted their impact to only those transitions from one state to a group found in the initial planting composition. E.g. asking whether the addition of natives to the initial planting group influences the likelihood that a given community will later transition to a native state. This would only affect 3 elements of the transition matrix governing the transitions to state 1 (native state) from states 2, 3, or 4 (wet annual, WAPS, or dry annual states, respectively).

The resulting likelihood ratio test suggests that the addition of these parameters significantly improves model fit (p > .001).

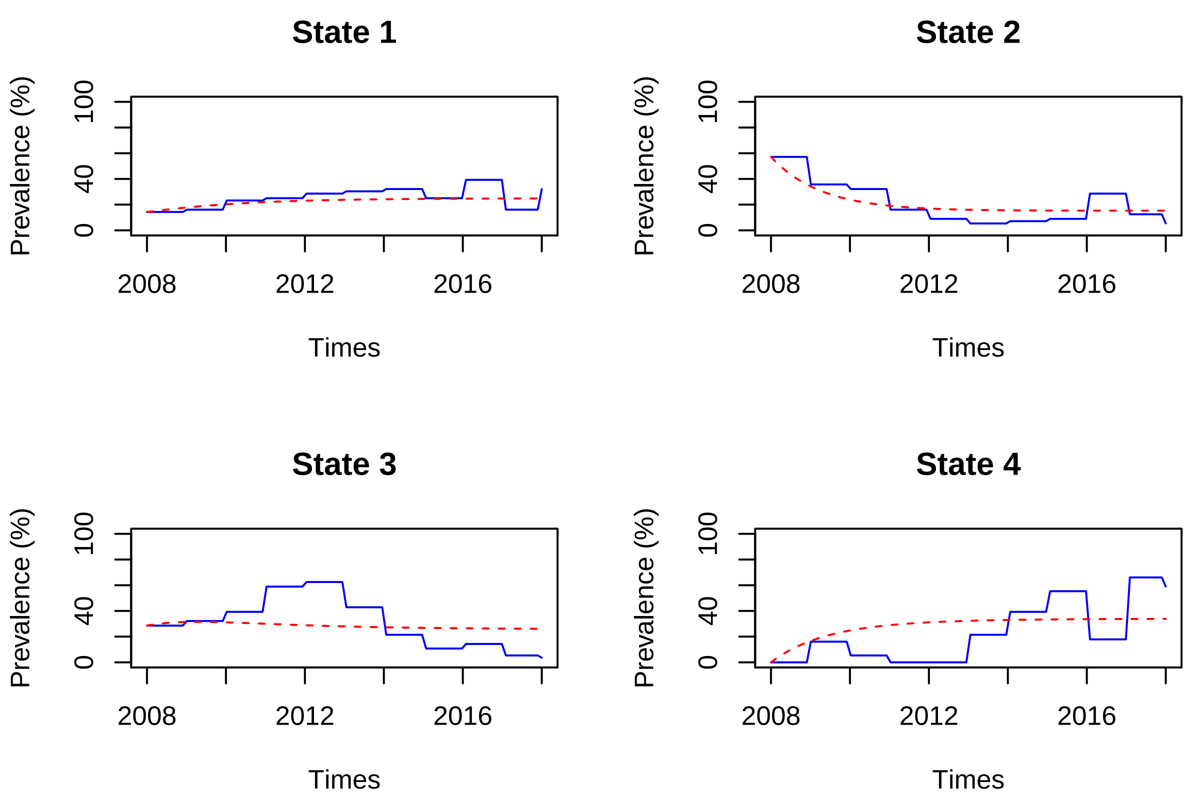
```
## Likelihood ratio test:
```

```
##              -2 log LR df            p
## msm.priority  53.15036 12 3.873242e-07
```

We can also visualize how well our model explains the observed patterns using a simple prevalence plot, which shows expected frequencies of a state in red with estimated state frequencies in blue.

I think that this model does quite a good job for two of the different states – Natives (state 1), and Priority Annuals (state 2), which are either quite stable over time, or decrease rapidly.

However, we seem to do poorly with WAPS and Dry Annuals states, whose transition probabilities seem to correlate strongly with climatic variation.

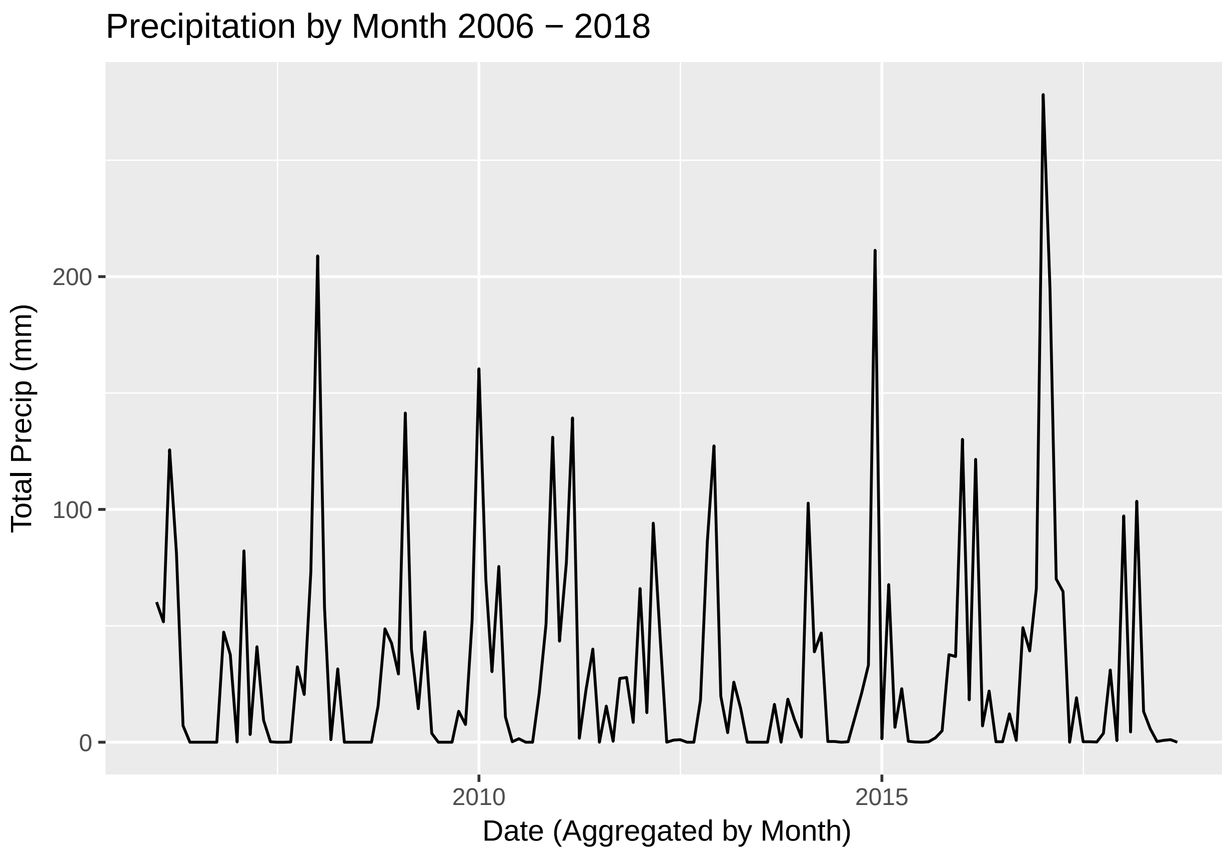


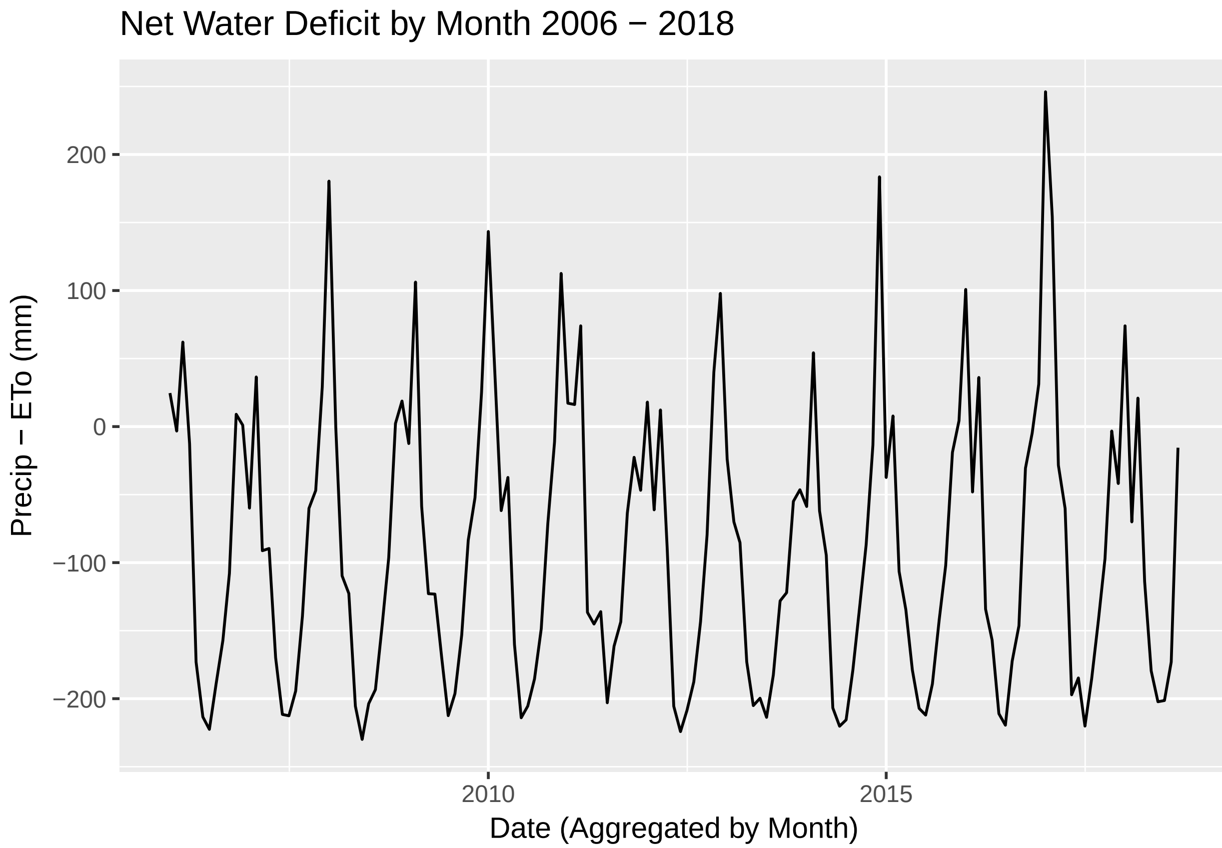

```
## 1 - Native, 2 - Priority Annuals, 3 - WAPS, 4 - Dry Annuals
```

Including climate data into analyses:

To improve the model fit above, we could try to incorporate environmental covariates into our data as well. We know that precipitation patterns can have major effects on community assembly, but it is unclear how to best examine the influence of this driver on state change rates.

To start, figures below show data on monthly precipitation and evapotranspiration measures taken from the Davis CIMIS (irrigation management) system. These fluctuations show a somewhat expected pattern - years 2012 - 2015 show a profound drought, while the periods between 2008 - 2011 and 2015-2017 are quite wet.

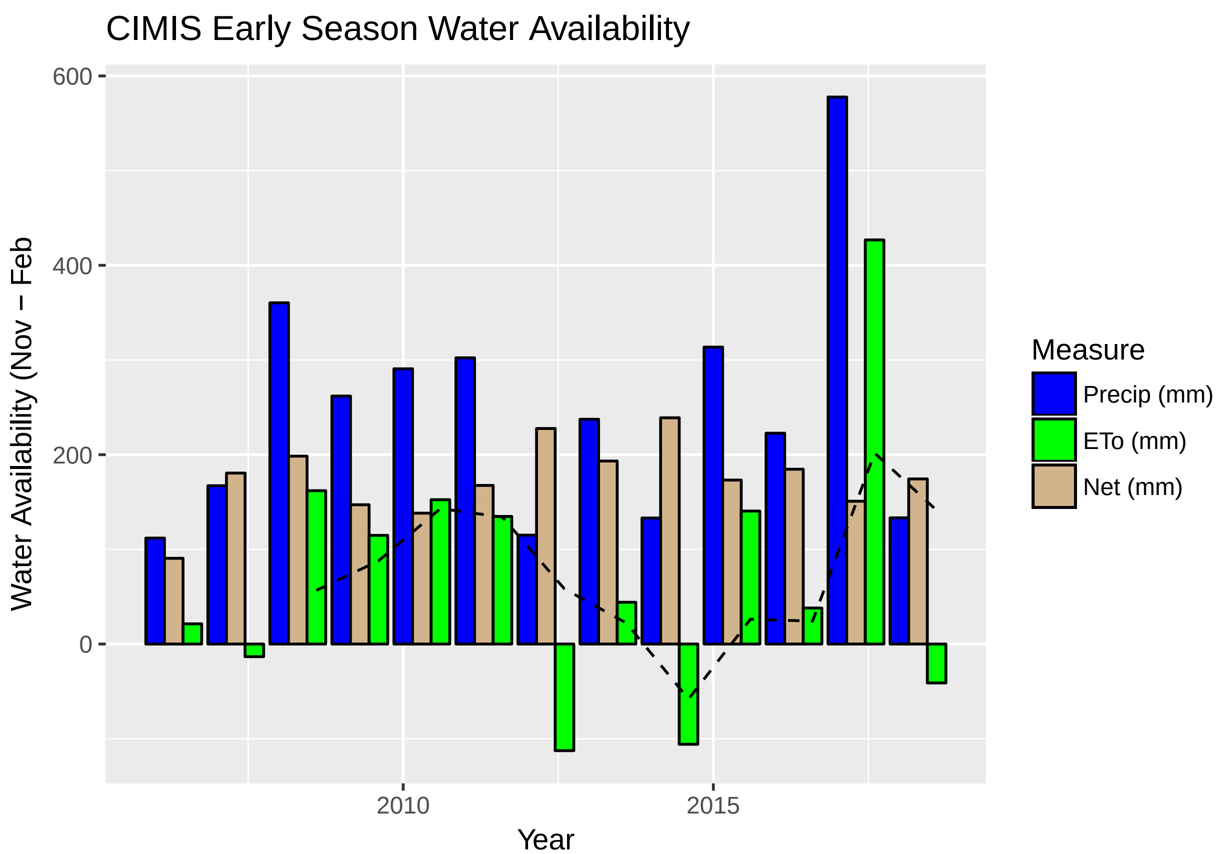




Many studies focusing on climate-driven changes in species abundance have focused on early-season precipitation as a main determinant of community composition. As a start, I will focus here on precipitation and ETo values between December and February.

There are a number of different ways to use this data to evluate changes in community transition rates – total precipitation, net water availability (precip - ETo), and lagged precipitation, to name a few. Dudney et al (2017), which attempts to evaluate current year and lagged precipitation effects, may be of some use.

The figure below shows this early precipitation change, aggregated into monthly intervals, with a three-year moving average indicated by the dashed line.



Because I don't seem to see a consistent trend here where lower values of net precipitation minus ETo produce within year-effects, I think a running average approach might be best. Rather than accounting for just a single year of data, taking a running average of drought stress can account for successive effects of multiple years.

```
## Likelihood ratio test (with env. covariates vs. null):
```

```
##         -2 log LR df            p
## msm.env  34.79283 12 0.0005052327
```

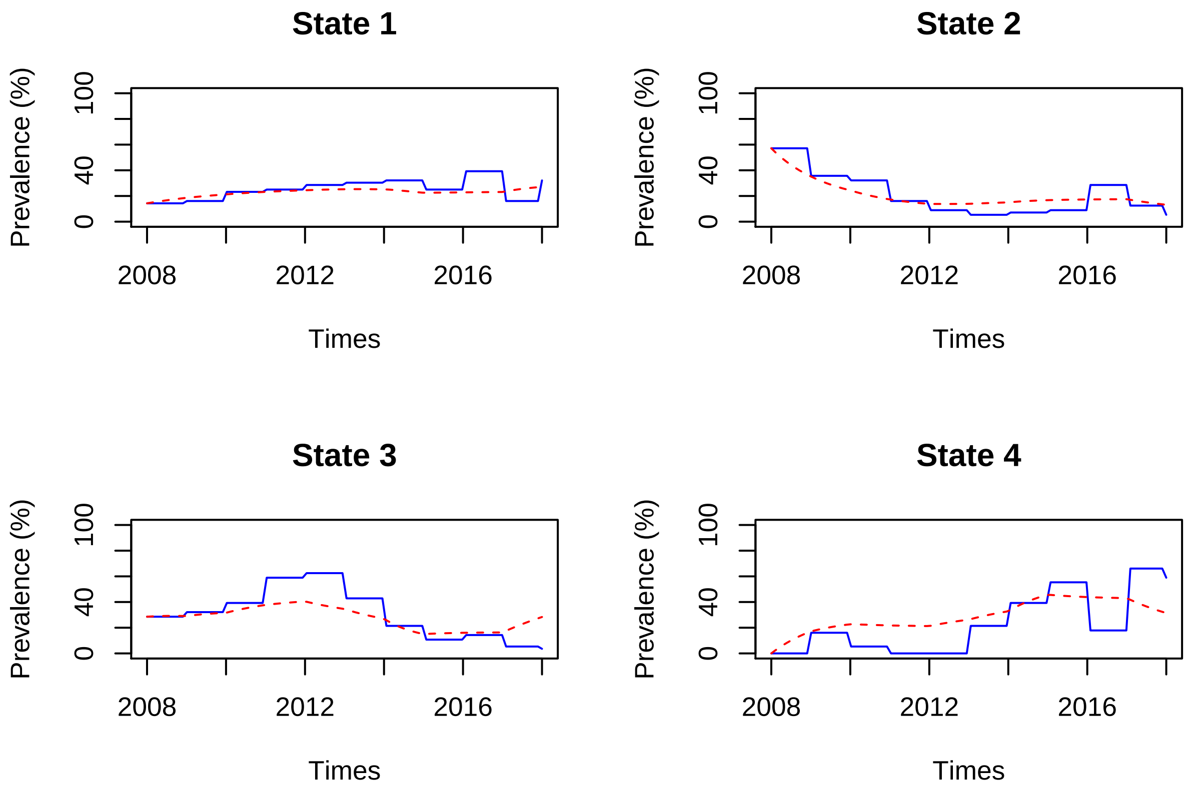
```
## Likelihood ratio test (with env. covariates + priority vs. priority:
```

```
##         -2 log LR df p
## msm.env -18.35753  0 1
```

```
## AIC scores of all models (lowest score is best):
```

```
##            model      AIC
## 1           Null 1289.975
## 2       Priority 1260.825
## 3    Environment 1279.182
## 4 Env + Priority 1248.810
```

The subsequent prevalence plot shows a substantially better fit to the models with this environmental data included - though it's not perfect.




```
## 1 - Native, 2 - Priority Annuals, 3 - WAPS, 4 - Dry Annuals
```

Next steps:

Going forward, I think the most important ways to refine this analysis are:

1. Make clustering approaches more concrete - what are the best steps in subsetting data and classifying community types?
2. Define questions clearly - what exactly is this approach answering?
3. Determine the best way to analyze climatic data - how to include variables in transition models?
4. Follow-up with clearer visualizations of how climate impacts persistence / state change - visualize how probabilities of transitions change with priority effects across a gradient of climatic values.
5. Identify other tests and analyses, as needed.

Visualizing response to climate:

Look at raw transition probabilities (add covariates list)

See MSM documentation on how to estimate probabilities with covariates, stability, etc.

Formulation of transition probabilities from Qmatrix:

When all transitions are possible, the probability of transitioning to either states 2, 3, or 4 from state 1 can be expressed as:

$$P_{11} = e^{-(q_{12}+q_{13}+q_{14})}$$

and

$$P_{12} = 1 - e^{-q_{12}}$$
$$P_{13} = 1 - e^{-q_{13}}$$
$$P_{14} = 1 - e^{-q_{14}}$$

```r
qmatrix.msm(msm.env_priority, covariates=list(net_cum_3 = 100))
```

```
##         State 1                      State 2
## State 1 -0.24171 (-0.378758,-0.15425)  0.02581 ( 0.006454, 0.10325)
## State 2  0.04767 ( 0.014337, 0.15850) -0.62031 (-1.058698,-0.36345)
## State 3  0.03337 ( 0.010762, 0.10345)  0.10022 ( 0.047650, 0.21077)
## State 4  0.07601 ( 0.033915, 0.17034)  0.20193 ( 0.099914, 0.40810)
##         State 3                      State 4
## State 1  0.02605 ( 0.007363, 0.09219)  0.18984 ( 0.113848, 0.31656)
## State 2  0.12239 ( 0.059920, 0.24999)  0.45025 ( 0.223819, 0.90575)
## State 3 -0.20556 (-0.342060,-0.12354)  0.07198 ( 0.029960, 0.17295)
## State 4  0.03794 ( 0.009291, 0.15496) -0.31588 (-0.530386,-0.18813)
```

```r
p12 = 1 - exp(-.02581)
p13 = 1 - exp(-.02605)
p14 = 1 - exp(-.1894)
tot = sum(p12,p13,p14)
p12/tot
```

```
## [1] 0.1138822
```

```r
p13/tot
```

```
## [1] 0.1149274
```

```r
p14/tot
```

```
## [1] 0.7711904
```

Below is code to pull out transition probabilities for each state. However, there is likely a better way:

- Create matrix of Q values
- Create function that extracts probabilities based on the function exp(q) or 1 - exp(-q).
- Assign diagonals to the probability of maintaining a state
- Assign non-diagonals to the probability of transitioning
- Repeat for confidence intervals (both low and high)

```r
Nprob = list()
counter = 1

for(pri in c(0:1)){
  for(netprecip in seq(-100, 500, by = 100)){
    qm = as.matrix(qmatrix.msm(msm.env_priority, covariates=list(net_cum_3 = netprecip,
                                                               natives = pri, annuals = pri, waps = pri
    ql = as.matrix(qmatrix.msm(msm.env_priority, covariates=list(net_cum_3 = netprecip,
```

```
                                                              natives = pri, annuals = pri, waps = pr
    qu = as.matrix(qmatrix.msm(msm.env_priority, covariates=list(net_cum_3 = netprecip,
                                                              natives = pri, annuals = pri, waps = pr

    qdf = data.frame(qm) %>% mutate(From = rownames(qm)) %>%
      gather("key" = "To", "value" = "Trans", -From) %>%
      mutate(To = gsub("\\.", " ", To))

    qlow = data.frame(ql) %>% mutate(From = rownames(ql)) %>%
      gather("key" = "To", "value" = "Lower", -From) %>%
      mutate(To = gsub("\\.", " ", To))

    qupp = data.frame(qu) %>% mutate(From = rownames(qu)) %>%
      gather("key" = "To", "value" = "Upper", -From) %>%
      mutate(To = gsub("\\.", " ", To))

    qdf$Trans = if_else(qdf$From == qdf$To, exp(qdf$Trans), 1 - exp(-qdf$Trans))
    qlow$Lower = if_else(qlow$From == qlow$To, exp(qlow$Lower), 1 - exp(-qlow$Lower))
    qupp$Upper = if_else(qupp$From == qupp$To, exp(qupp$Upper), 1 - exp(-qupp$Upper))

    output <- left_join(qdf, qlow) %>% left_join(qupp)
    output$precip = rep(netprecip, nrow(output))
    output$priority = rep(pri, nrow(output))

    Nprob[[counter]] <- output
    counter = counter + 1
  }
}
```

```
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
```

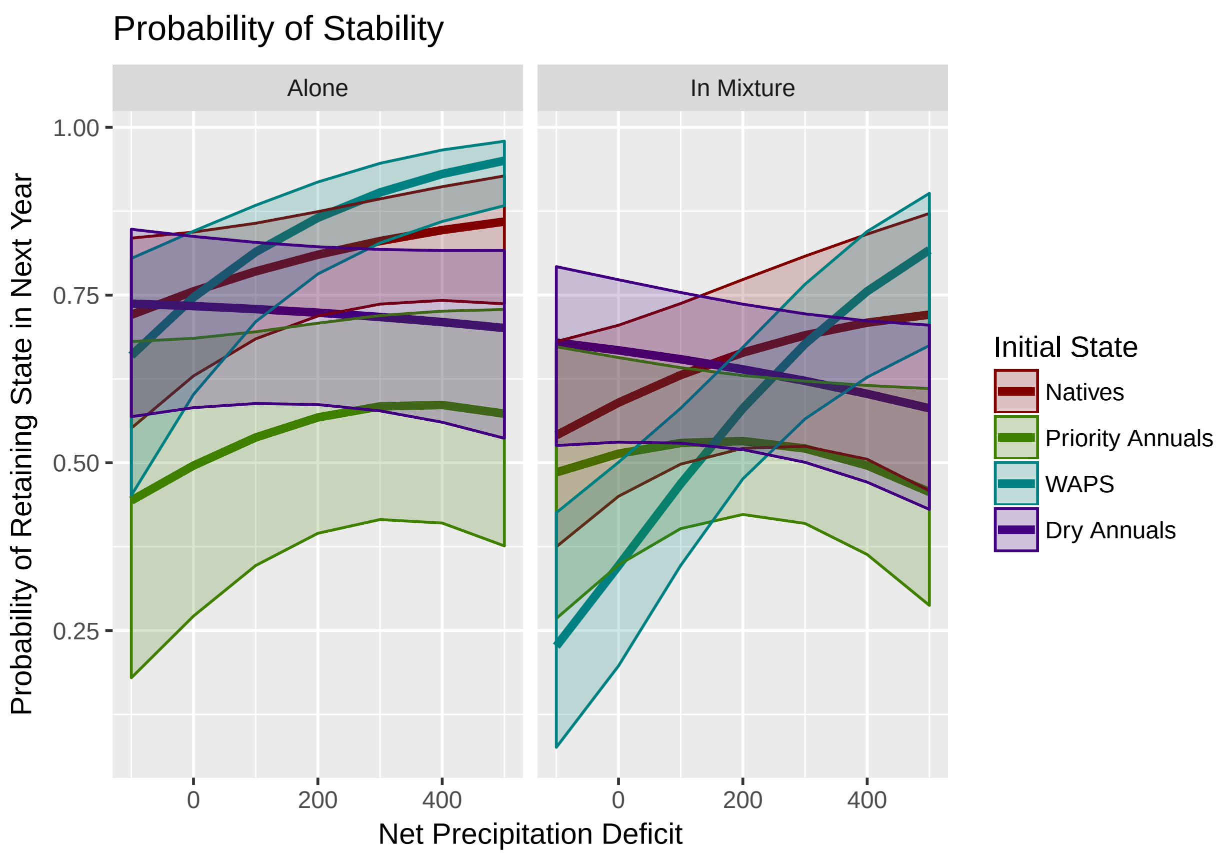
```
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
## Joining, by = c("From", "To")
```

```r
library(viridis)
```

```
## Loading required package: viridisLite
```

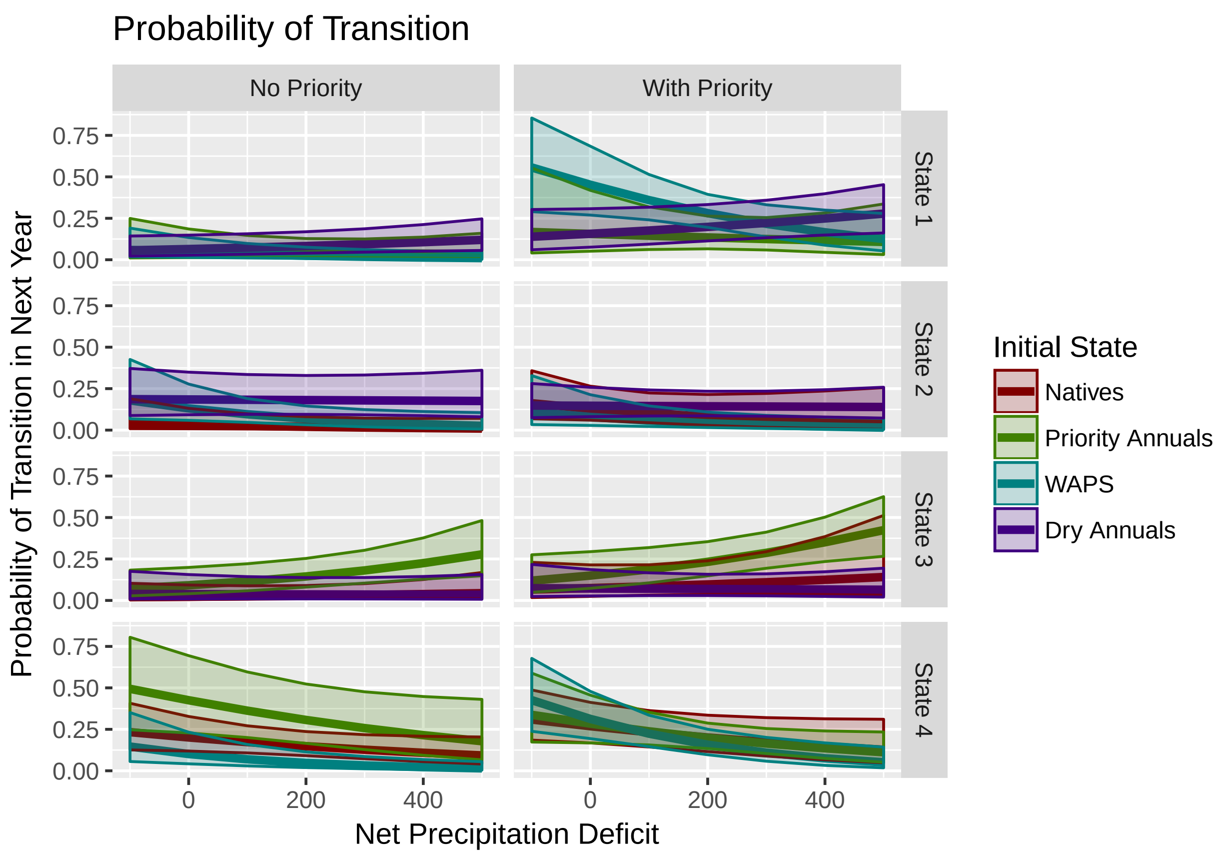
```r
cols = rainbow(4, v = .5)

bind_rows(Nprob) %>%
  filter(From == To) %>%
  mutate(priority = if_else(priority == 1, "In Mixture", "Alone")) %>%
  ggplot(aes(x = precip,
             color = From,
             fill = From)) +
  geom_line(aes(y = Trans), size = 1.5) +
  geom_ribbon(aes(ymax = Upper, ymin = Lower), alpha = .2) +
  facet_wrap(~priority) +
  ggtitle("Probability of Stability") +
  scale_color_manual(name = 'Initial State',
         values =c("State 1" = cols[1], "State 2" = cols[2],
                   "State 3" = cols[3], "State 4" = cols[4]),
         labels = c('Natives','Priority Annuals', 'WAPS','Dry Annuals')) +
  scale_fill_manual(name = 'Initial State',
         values =c("State 1" = cols[1], "State 2" = cols[2],
                   "State 3" = cols[3], "State 4" = cols[4]),
         labels = c('Natives','Priority Annuals', 'WAPS','Dry Annuals')) +
  xlab("Net Precipitation Deficit") +
  ylab("Probability of Retaining State in Next Year")
```




```
ggsave("../figures/StabilityProbabilities.pdf", height = 6, width = 8)
```

```
trtnames = c("to Natives", "to Priority Annuals", "to WAPS", "to Dry Annuals")
priorities = c("No Priority", "Priority")
bind_rows(Nprob) %>%
  filter(From != To) %>%
  mutate(priority = if_else(priority == 1, "With Priority", "No Priority")) %>%
  ggplot(aes(x = precip,
             color = From,
             fill = From)) +
  geom_line(aes(y = Trans), size = 1.5) +
  geom_ribbon(aes(ymax = Upper, ymin = Lower), alpha = .2) +
  facet_grid(To~priority) +
  ggtitle("Probability of Transition") +
  scale_color_manual(name = 'Initial State',
         values =c("State 1" = cols[1], "State 2" = cols[2],
                   "State 3" = cols[3], "State 4" = cols[4]),
         labels = c('Natives','Priority Annuals', 'WAPS','Dry Annuals')) +
  scale_fill_manual(name = 'Initial State',
         values =c("State 1" = cols[1], "State 2" = cols[2],
                   "State 3" = cols[3], "State 4" = cols[4]),
         labels = c('Natives','Priority Annuals', 'WAPS','Dry Annuals')) +
  xlab("Net Precipitation Deficit") +
  ylab("Probability of Transition in Next Year")
```




```
ggsave("../figures/TransitionProbabilities.pdf", height = 14, width = 8)
```