

Community Ecology With GLLVMs

Sam Perrin

Day Two

The data we're using for this session was collected largely by the ECOSPAT Group at the University of Lausanne. It looks at co-occurrence patterns in different plants in the Swiss Alps. The original dataset can be found at <https://doi.org/10.5061/dryad.8mv11> (<https://doi.org/10.5061/dryad.8mv11>), which also contains links to a fantastic paper by Manuela D'Amen, Heidi Mod, Nicholas Gotelli and Antoine Guisan which uses the data, entitled "Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence". The elevational data has been provided by SwissTopo.

As a very basic overview, the original data includes presence-absence data for 183 plants over 912 sites, and the following environmental covariates. The dataset included below is a subset, so that this all runs a bit faster during the workshop. Obviously subsetting like this isn't always recommended, but if you'd like to run it with the full dataset we're happy to provide it.

We start by loading all the requisite packages.

```
library(gllvm)
library(dplyr)
library(grDevices)
```

The data we're using for this session was collected largely by the ECOSPAT Group at the University of Lausanne. It looks at co-occurrence patterns in different plants along an elevational gradient in the Swiss Alps. The original dataset can be found at <https://doi.org/10.5061/dryad.8mv11> (<https://doi.org/10.5061/dryad.8mv11>), which also contains links to a fantastic paper by Manuela D'Amen, Heidi Mod, Nicholas Gotelli and Antoine Guisan which uses the data, entitled "Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence". The elevational data we use here comes from SwissTopo.

As a very basic overview, the original data includes presence-absence data for 183 plants over 912 sites, and the following environmental covariates. The dataset included below is a subset, so that this all runs a bit faster during the workshop. Obviously subsetting like this isn't always recommended, but if you'd like to run it with the full dataset that has also been included in the GitHub.

These are our environmental covariates.

DDEG0 - Days over zero degrees

SOLRAD - Summed annual solar radiation

SLOPE - Slope angle in degrees

MIND - Moisture index

TPI - Topographic position index

ELEVATION - Number of kangaroos present at site (joking it's just elevation)

Each of these environmental covariates has been standardised to a mean of 0 and SD of 1. This is just to help model convergence.

```
load("WorkshopData.RDA")

Y <- WorkshopData$Y
X <- WorkshopData$X
```

For starters, let's run a basic gllvm using no environmental variables.

```

time1 <- Sys.time()
fit_base <- gllvm(Y, num.lv = 2, family = binomial(link="probit"), sd.errors = FALSE)
Sys.time()-time1

```

```
## Time difference of 36.65673 secs
```

Before we look at our latent variables, let's have a look at collinearity. The following code shows any significant collinearity between our environmental covariates.

```

source("http://www.sthda.com/upload/rquery_cormat.r")
colin <- rquery.cormat(X, type = "flatten", graph = FALSE)

```

```
colin$r %>% filter(abs(cor) > 0.4)
```

```

##      row      column  cor      p
## 1 DDEG0      MIND -0.89 6.9e-141
## 2 SOLRAD      MIND -0.46 1.5e-22
## 3 SLOPE      MIND  0.46 9.1e-23
## 4 DDEG0 ELEVATION -0.99 0.0e+00
## 5 MIND ELEVATION  0.90 4.0e-148

```

The high cor values and low p values indicate some serious collinearity between positive degree days, slope, elevation and moisture index.

Let's see if our latent variables correspond to any of these covariates.

We define colours according to the values of covariates. The darker blue indicates a higher value of the relevant covariate.

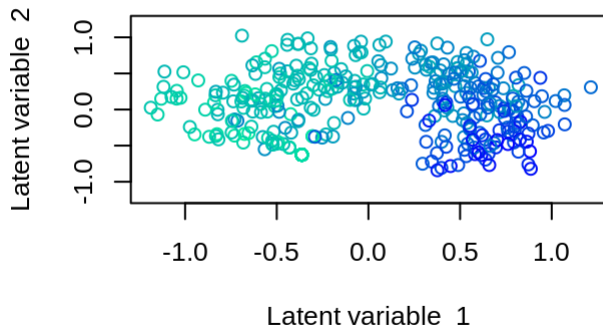
```

par(mfrow=c(2,2))
for (i in 1:length(colnames(X))) {
  covariate <- X[,i]
  rbPal <- colorRampPalette(c('mediumspringgreen', 'blue'))
  Colorsph <- rbPal(20)[as.numeric(cut(covariate, breaks = 20))]
  breaks <- seq(min(covariate), max(covariate), length.out = 30)

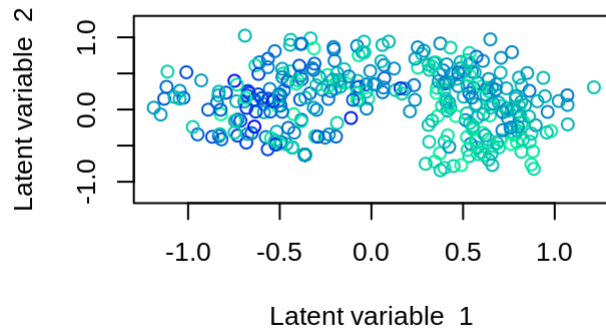
  ordiplot(fit_base, main = paste0("Ordination of sites, color: ",colnames(X)[i]),
           symbols = TRUE, s.colors = Colorsph, xlim = c(-1.2,1.2), ylim = (c(-1.2, 1.2
)))
}

```

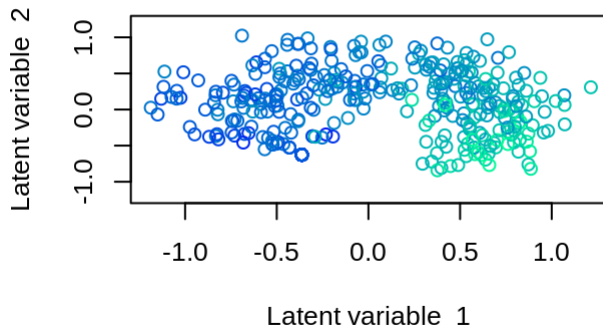
Ordination of sites, color: DDEG0



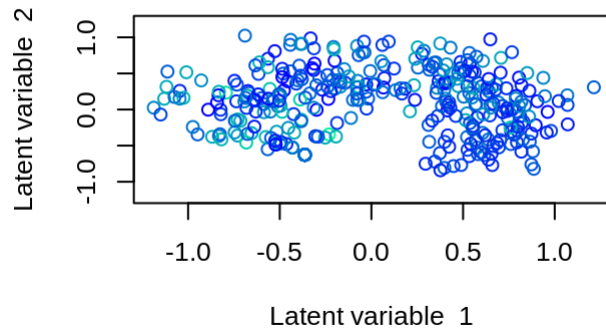
Ordination of sites, color: SLOPE



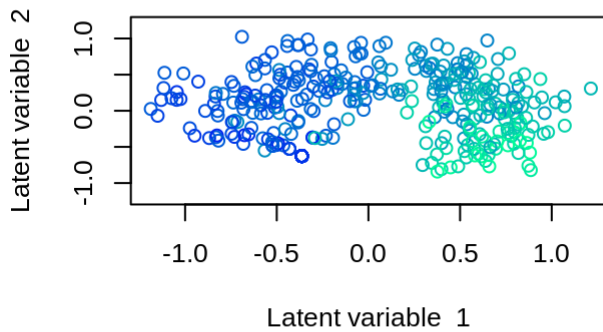
Ordination of sites, color: MIND



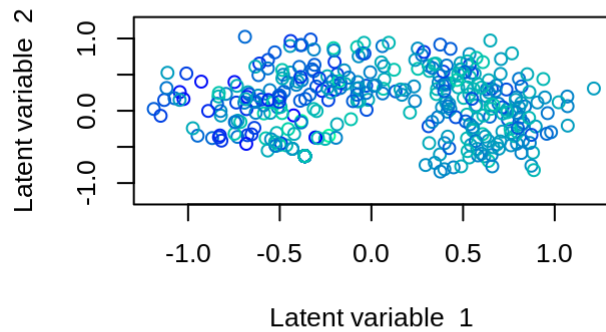
Ordination of sites, color: SOLRAD



Ordination of sites, color: ELEVATION



Ordination of sites, color: TPI



We can see some quite clear gradients related to the four collinear variables we mentioned above. At this point let's take the two climate variables which are likely to have a direct impact on the community, since ELEVATION is likely to be a proxy for other covariates (more on that later).

Let's use these two and use some code from last session to figure out how many latent variables would be appropriate. Basically we're running the same model again and again, but increasing the number of latent variables each time. I'm using the lowest AICc value to determine the model which fits best.

```
fit_list <- list()
for(i in 0:3){
  fit_sub <- gllvm(Y, X, family = binomial(link="probit"), num.lv = i, sd.errors = FALSE,
    formula = ~ DDEG0 + MIND, seed = 1234)
  fit_list[[i+1]] <- fit_sub
}
```

Let's have a look at our AICc values.

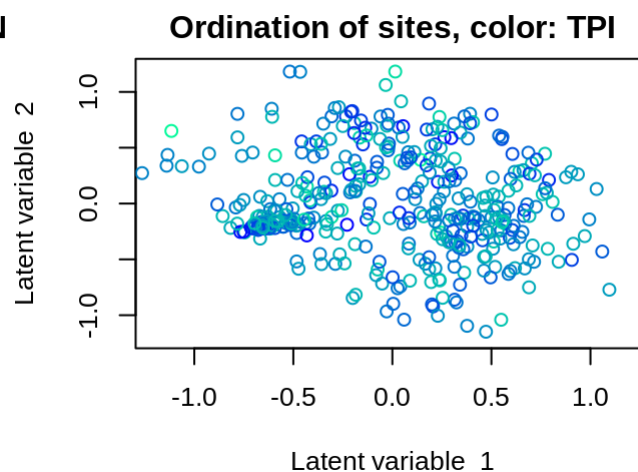
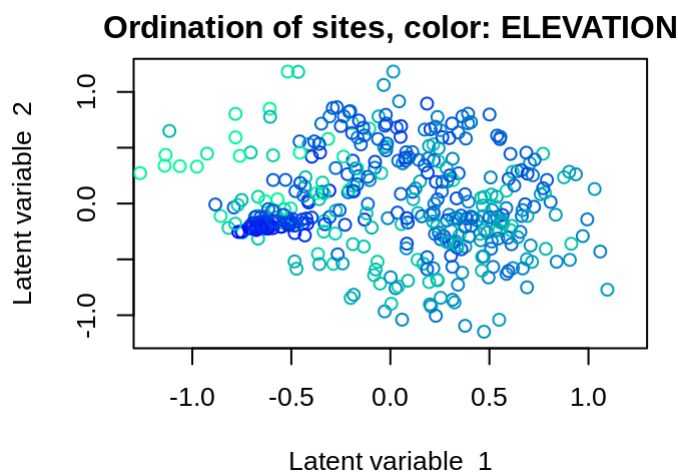
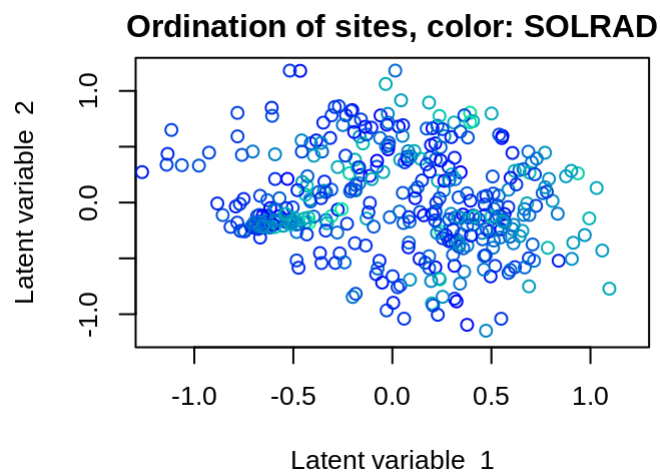
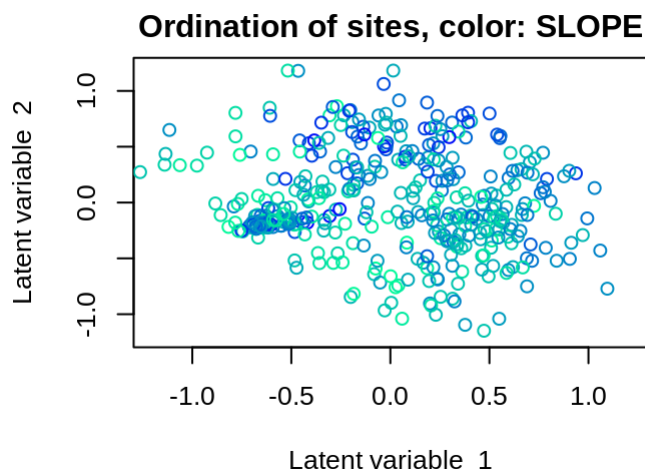
```
AICc <- sapply(fit_list, function(X) {summary(X)$AICc})
data.frame(AICc, model= paste0("LV-",0:3))
```

```
##      AICc model
## 1 14417.02 LV-0
## 2 12672.38 LV-1
## 3 12655.35 LV-2
## 4 12998.40 LV-3
```

We can see that the best model here uses 2 latent variables. Let's have a look at how these variables compares to our remaining environmental covariates.

```
par(mfrow = c(2, 2), mar = c(4, 4, 2, 1))
remaining_covariates <- c("SLOPE","SOLRAD","ELEVATION","TPI")

for(i in 1:length(remaining_covariates)) {
  covariate <- X[,remaining_covariates[i]]
  rbPal <- colorRampPalette(c('mediumspringgreen', 'blue'))
  Colorsph <- rbPal(20)[as.numeric(cut(covariate, breaks = 20))]
  breaks <- seq(min(covariate), max(covariate), length.out = 30)
  ordiplot(fit_list[[3]], main = paste0("Ordination of sites, color: ",remaining_covariates[i]),
    symbols = TRUE, s.colors = Colorsph, xlim = c(-1.2,1.2), ylim = (c(-1.2, 1.2)))
}
```



From this we can see that there is still a bit of variation explained by elevation (and potentially slope). So, is it worth including?

Let's have a look at our LVs if we include elevation.

```
fit_Elevation <- gllvm(Y, X, family = binomial(link="probit"), num.lv = 2, sd.errors = FALSE,
                      formula = ~ DDEG0 + MIND + ELEVATION, seed = 1234)

summary(fit_Elevation)$AICc
```

```
## [1] 12571.26
```

There's a slight drop in AICc values. But is it useful to include a proxy like elevation here?

► EXTENDED QUESTION

Breakout questions

What other sort of variables would serve as proxies to account for variation?

What's the difference between species associations and species interactions?

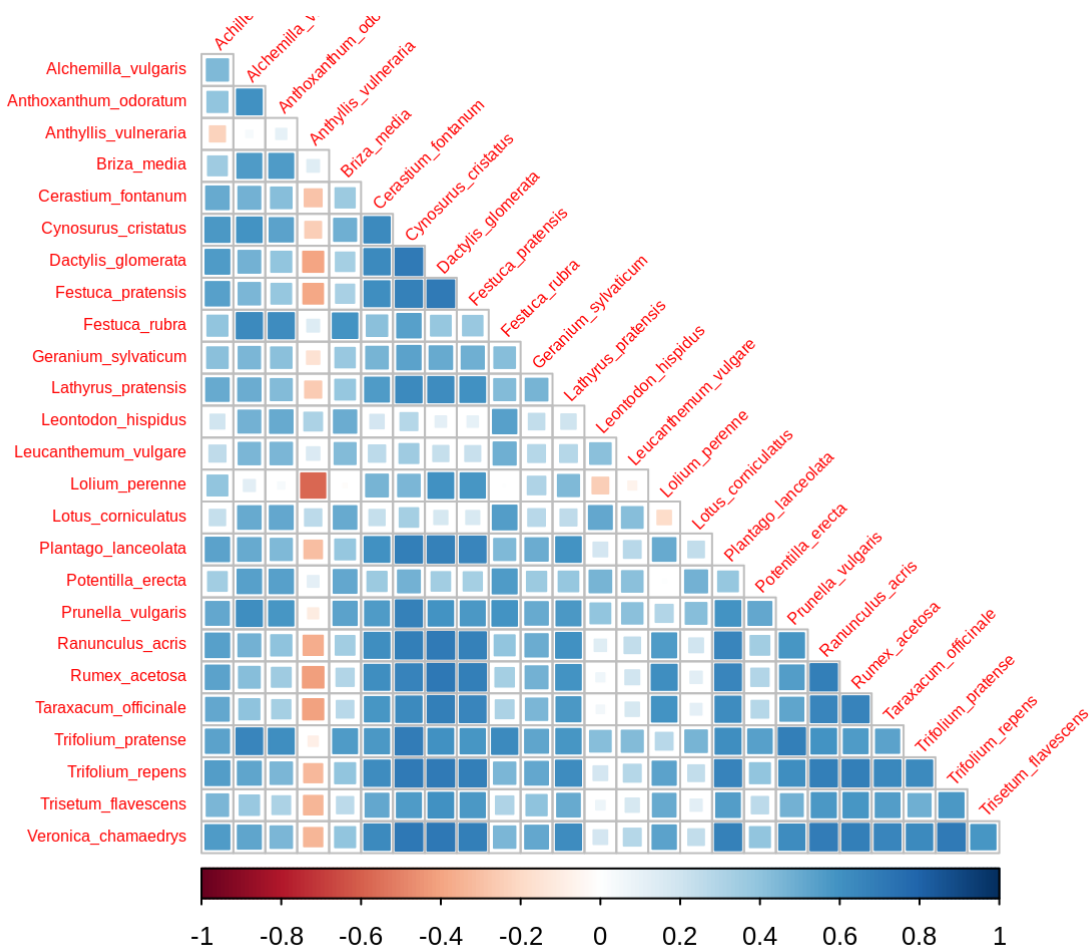
Species Associations

There are a lot of species down here, so let's do what the researchers did and look at a subset of elevational species, in this case only colline species, species which occur below 700m above sea level.

```
colline_species <- WorkshopData$colline_species
```

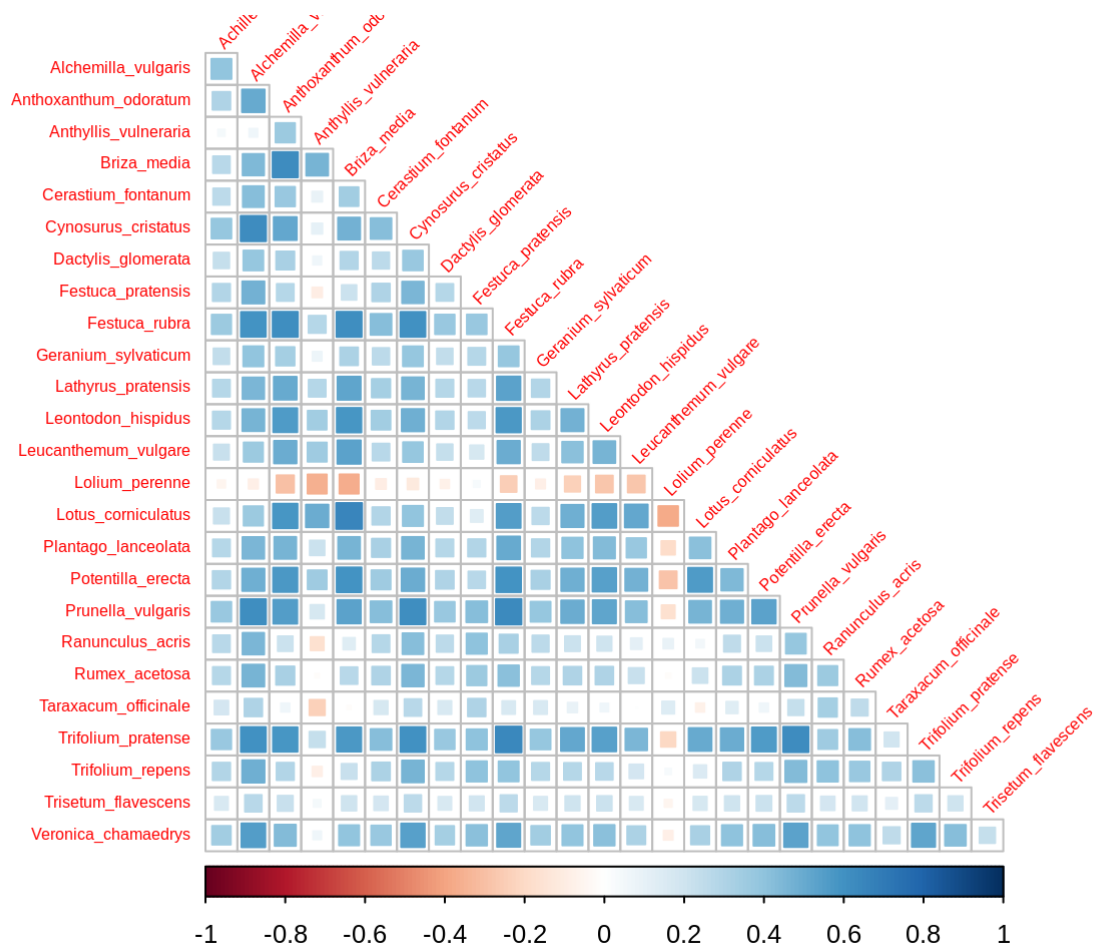
Let's start by having a look at the species relationships when we have no covariates.

```
cr0 <- getResidualCor(fit_base)
corrplot(cr0[colline_species,colline_species], diag = FALSE, type = "lower",
         method = "square", tl.cex = 0.5, tl.srt = 45, tl.col = "red")
```



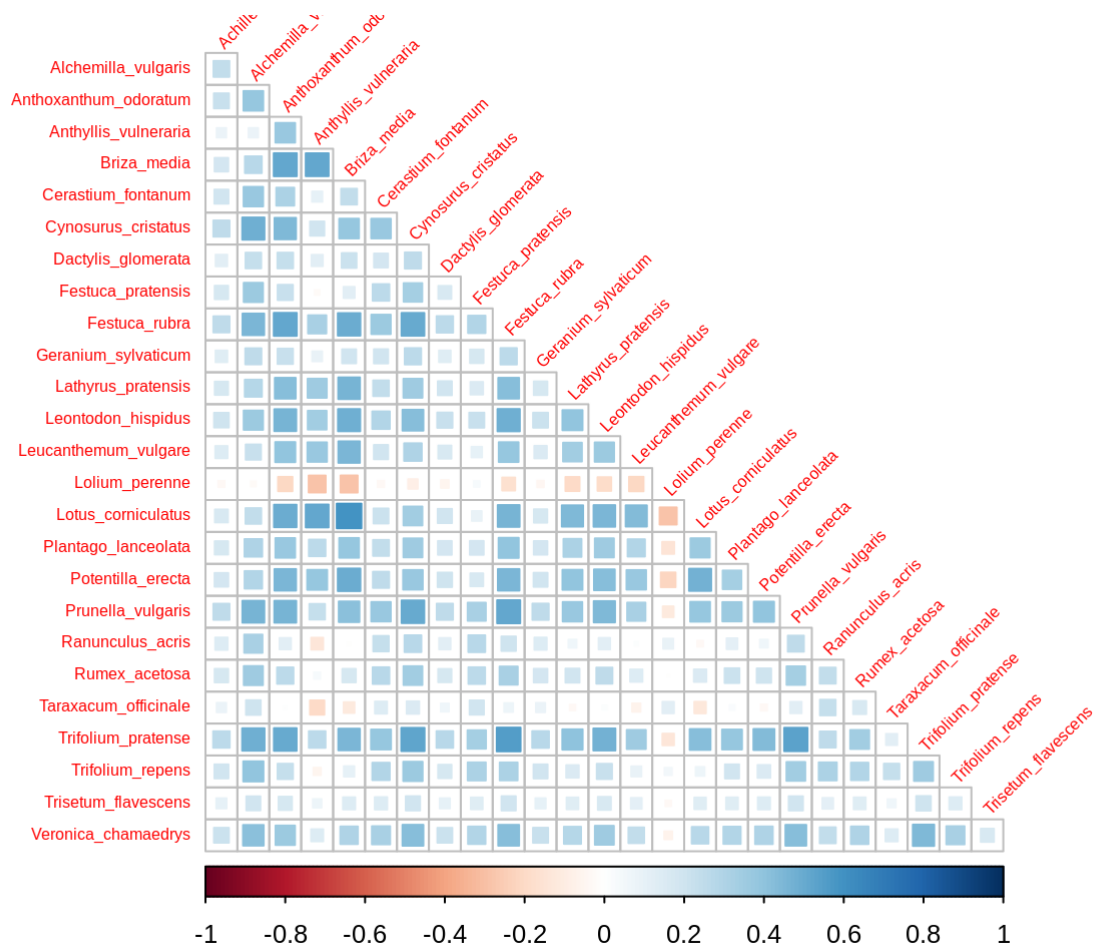
This cooccurrence plot shows associations between different species. Blue values indicate positive co-occurrence values, which indicate the species are likely to co-occur. Larger squares and darker colours indicate stronger relationships. Red values indicate the opposite. What happens when we introduce MIND and DDEG0?

```
cr1 <- getResidualCor(fit_list[[3]])
corrplot(cr1[colline_species,colline_species], diag = FALSE, type = "lower",
         method = "square", tl.cex = 0.5, tl.srt = 45, tl.col = "red")
```



You can see that there's a notable change in the species associations, as variation that was formerly attributed to latent variables (and thus to species associations) is now attributed to the environmental covariates. Let's see what happens when we add in a proxy like elevation.

```
cr2 <- getResidualCor(fit_Elevation)
corrplot(cr2[colline_species,colline_species], diag = FALSE, type = "lower",
  method = "square", tl.cex = 0.5, tl.srt = 45, tl.col = "red")
```



You can see that in many cases the association becomes less pronounced again.

► EXTENDED QUESTION

That's it from me. Have any extra questions? Get in touch with us via the conference app! You can also contact me directly via email at sam.perrin@ntnu.no (mailto:sam.perrin@ntnu.no) or on Twitter at @somperrinNTNU.