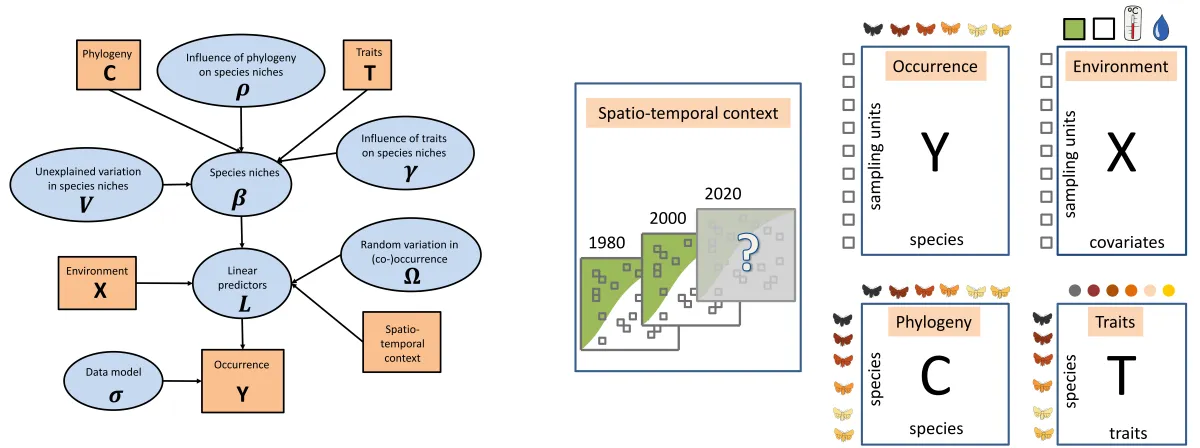
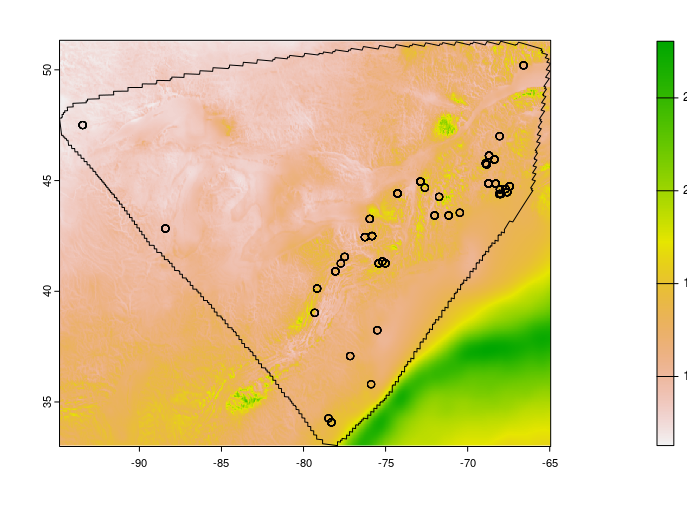
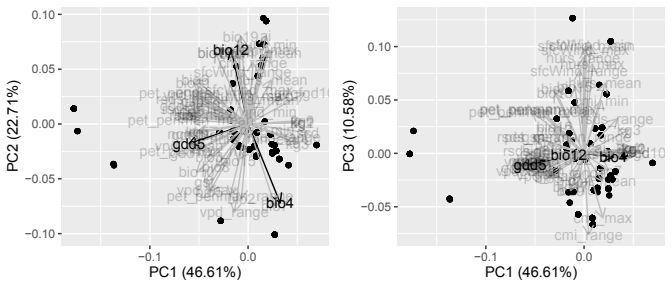
**Hierarchical Modelling of Species Communities (HMSC)**

This model is one of the Joint Species Distribution Models that have been developed in the last decade. It aims to fit multiple species at once, taking into account phylogenies and species functional traits into account when estimating the effect of environment on the different species (betas). It also allow to account for co-occurrence residuals (omegas) once the effect of the environment has been removed. More information here and references therein: <https://www.helsinki.fi/en/researchgroups/statistical-ecology/software/hmsc>

**Model calibration**

I fit the model using climate data from the Chelsa Climate project, and occurrence for the four species, selecting only those from localities in which at least 5 samples were identified, (n=350) regardless of the method used. Climate variables were analyzed in a PCA an the three first axis were selected to summarize main environmental gradients. Selected samples are shown in the following figure.

Figure 1: Map of the study area and samples (n = 350), in the background Bio12 (annual precipitation).

The three first axis of the pca explain more than 80% of the variance in all climatic variables. In the next plot I highlight the role of GDD 5 (Growing degree days over 5 degrees), Bio4 (temperature seasonality as a measure of continentality) and Bio12 (annual precipitations).

The model was fit with a structure in which the occurrence of the four species depends on the 3 main environmental gradients, two habitat variables (Cover and Hydrology), and two random effects: Sample identity itself and the Locality in which multiple samples were colleted. I didn’t use localities as our sites in which species co-occurrence were analysed, because it would make us lost the information about the particular habitat in which each sample was collected. Initially, each species was assumed to have a unimodal (quadratic) response to the 3 main environmental gradients. However, no species had significant coefficients for the quadratic terms of the climate variables. Hence, we rebuilt models considering only monotonic (binomial-linear) responses.

Model was calibrated using 2 MonteCarlo Markov Chains. Each chain was calculated with 10.000 burning samples, and 100.000 posterior samples. To avoid autocorrelation in the posterior samples, I thined the chains using a factor of 100 (selecting 1 sample every 100 samples), resulting in a total of 1.000 posterior samples. Because I fitted 2 independent chains, the total posterior distributions are drawn from a total of 2.000 samples.

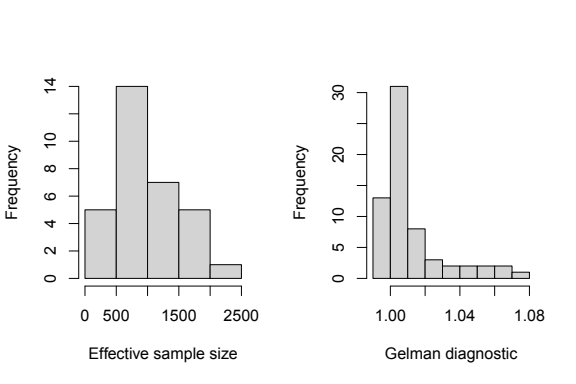
**Model evaluation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species | Training data | | | Testing data – 4 fold CV | | |
| RMSE | AUC | TjurR2 | RMSE | AUC | TjurR2 |
| Sph. diabolicum | 0.080 | 1.000 | 0.898 | 0.323 | 0.924 | 0.575 |
| Sph. divinum | 0.193 | 0.993 | 0.749 | 0.278 | 0.940 | 0.613 |
| Sph. magniae | 0.054 | 1.000 | 0.868 | 0.058 | 0.964 | 0.820 |
| Sph. medium | 0.170 | 0.995 | 0.687 | 0.281 | 0.934 | 0.433 |

RMSE: Root mean squared error. AUC: Area Under the Roc Curve. TjurR2: adjusted R2 for binomial data. In general, the lower the RMSE the better, whereas the higher the AUC and TjurR2 the better. I would claim that the model adjust pretty well the training data. Hence, the model is able to explain the species occurrence pretty well. It also seems to estimate correctly the testing data (using a cross validation with 4 folds).

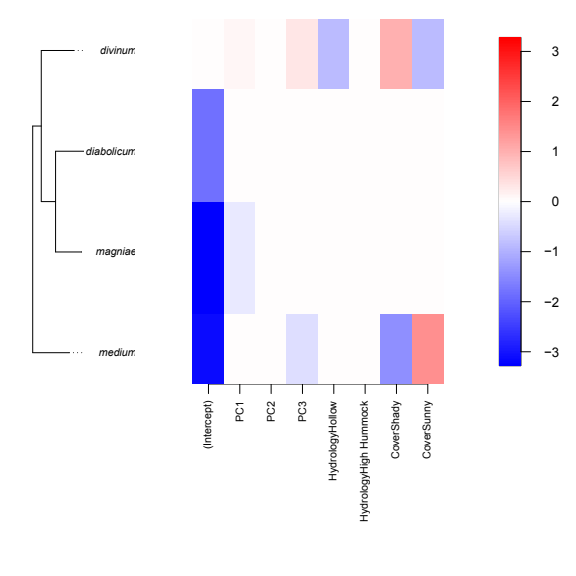
**Model convergence**

Being a bayesian model, its convergence has to be validated. To do so, I used the Effective Sample Size (ESS) and the Gelman diagnostic over the betas parameters estimated by the model (the effect of the environmental variables). Ideally, ESS should be around 2000, the same sample size we have in our bayesian models. Gelman diagnostic should be around 1. Both metrics indicate a fearly well calibration of the beta parameters in the model.

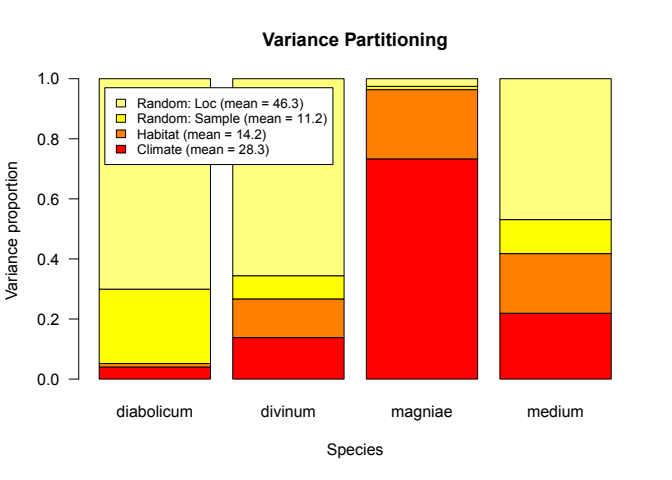
An overview of the chains evolution can be inspected in the supplementary information (attached as an independent pdf file).

**The effect of environmental variables (climate and habitat) on species occurrences**

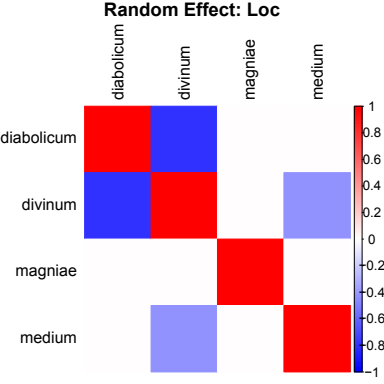
Environmental variables had a contrasting effect among species. *Sphagnum diabolicum* didn’t show a significant response to any environmental variable, whether climatic or habitat. *Sphagnum magniae* and *Sph. divinum* showed contrasting responses to climatic variables. *Sph. magniae* showed a small but significant negative response to the first climatic gradient, and positively to the second and third climatic gradients. On the other side, *Sph. divinum* responded positively to the first climatic grandient and negatively to the second and third climatic gradients. Finally, *Sph. medium* responded positively to the third climatic gradient. Regarding the response to habitat variables (cover and hydrology), *Sph. medium* had a positive response towards open and sunny habitats, and *Sph. divinum* respond positively to close and shady habitats. Furthermore, *Sph. divinum* responded negatively to Hollow habitats.

When analysing the variance explained by each variable, there were again contrasting results among species. Whereas, *Sph. magniae* is explained in a great proportion by climate variables (>75%), for *Sph. divinum, medium* and *diabolicum* the role of climate decrease significantly (<45, 30, and 10%, respectively). Habitat variables explain, in general, a small amount of variation (<20%) for all species. Finally, random variables (Sample and Locality) combined explain the greatest proportion of variation. Although it changed considerable among species. For instance, *Sph. diabolicum* and *medi**um* are mainly explianed by the random variables (>50%), whereas *Sph. magniae* doesn’t varies due to the random effects.

The importance of random variables suggest other unaccounted variables in the distribution of the species and its “community” assemblage. The nature of these unaccounted variables is hard to figure out. It could be some dispersal constrain, other environmental variable whether climatic or habitat variables (pH).

**Species cooccurrences**

Cooccurrences residuals estimated by the HMSC model at the Locality level indicates that once the effect of climate and habitat has been removed, *Sph. diabolicum* co-occur less frequently than expected by random with *Sph. divinum*. *Sph. medium* seems to covary negatively with *Sph. divinum*.

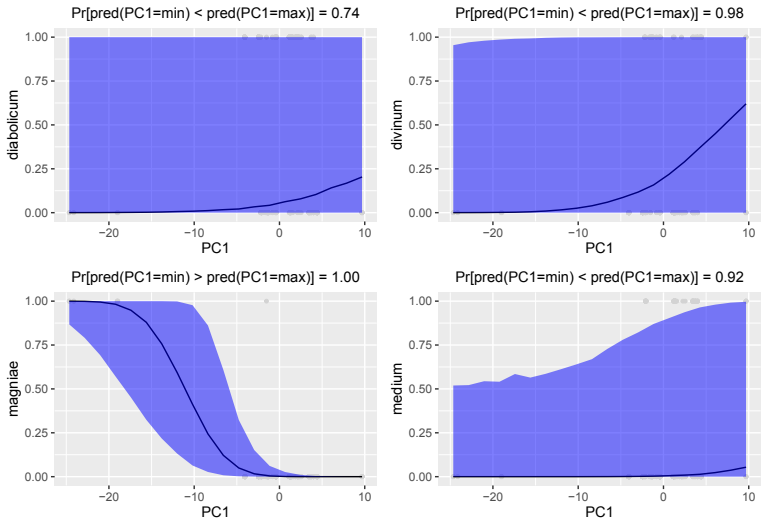
****

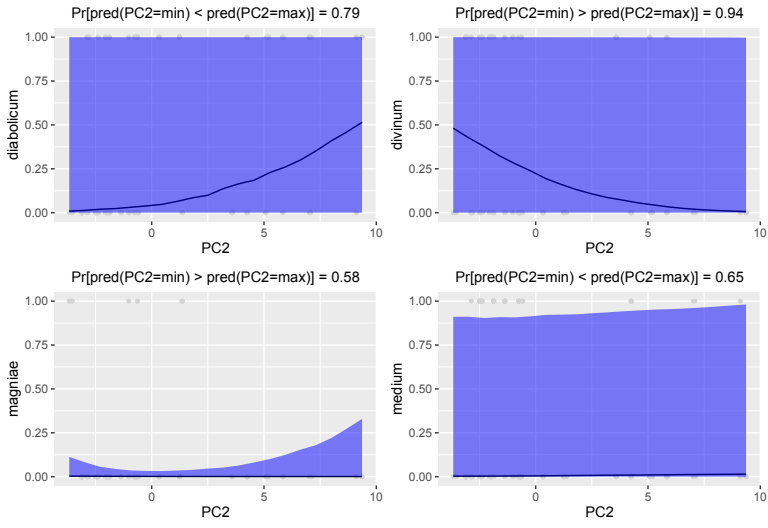
**Species response curves to environmental variables**

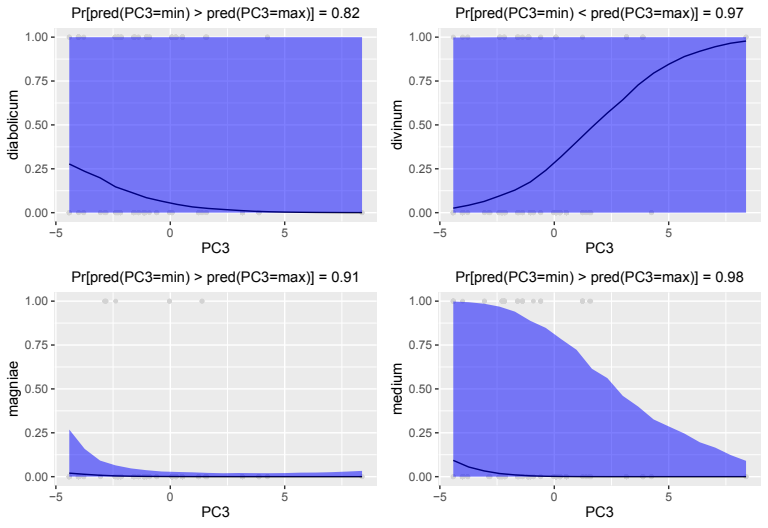
I think these figures are selfexplanatory, just show how the species respond to each environmental variables.

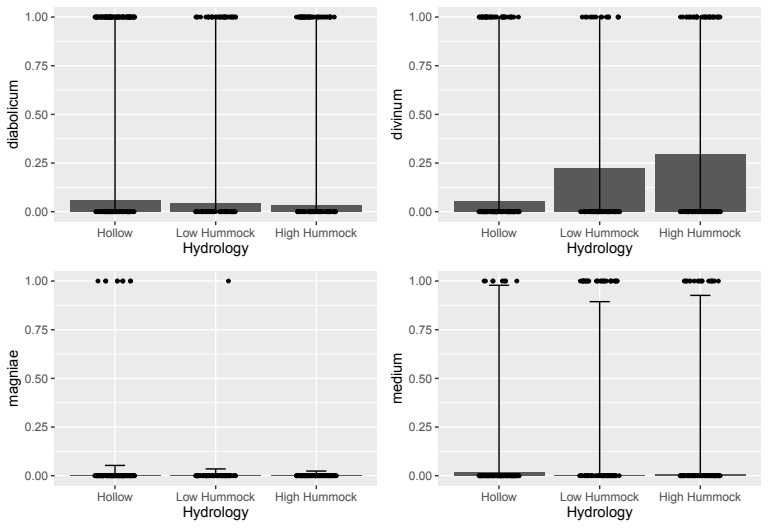
My main comment about these figures is the high confidence intervals, which are in light with the relative small role of climatic variables in explaining species occurrence and the small values of the beta parameters.

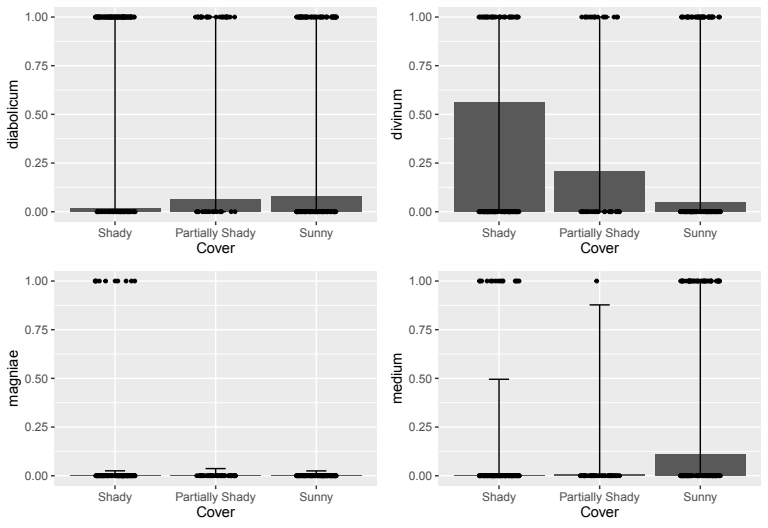
All together combined, I don’t think it make sense to project the models to future climatic conditions. I think the model somehow highlight some differences in the ecology of the species, whether climatically or at the habitat level but suggest that we are missing something else that might be important. It could also be that there are an important idiosyncratic variance at the sample and Locality level.



****

****

****

****