## Independent work 3

Today you will be using the "*saana.csv*" data set. For this data set, vascular plant species and abiotic conditions were surveyed on the Saana massif in north-western Finland (69°3'N 20°51'E) at ca. 700 m a.s.l., ca. 100–200 m above the birch tree line. The data set consists of grid data of 1m x 1m quadrats. Within each quadrat the following variables were surveyed or measured:

mesotopo = mesotopography
soil_moist = soil moisture
soil_temp = soil temperature
soil_ph = soil pH
veg_height = mean vegetation height (cm)
vasc_spr = species richness of vascular plants
Empher_cover = % cover of *Empetrum hermaphroditum*
Betnan = occurrence of *Betula nana* (dwarf birch)
Cashyp = occurrence of *Cassiope hypdnoides* (moss heather)
Empher = occurrence of *Empetrum hermpaphroditum* (crowberry)
Gersyl = occurrence of *Geranium sylvaticum* (woodland geranium)
Salret = occurrence of *Salix reticulata* (snow willow)

You will be utilizing this data set to build GLMs, GAMs, and GBMs in order to model vegetation properties, such as species occurrence (presence/absence), species richness, vegetation height, and the cover of *Empetrum hermaphroditum*, a dominant species. You will be using topography and soil parameters as predictors

**Question 1:** Divide the saana.csv data randomly into two different datasets:

- model calibration data (70%)
- model evaluation data (30%)

Build the models based on the calibration data and test the predictive performance of the models using the evaluation data. What is the predictive performance of the GLM, GAM and GBM models for ***Betnan, Cashyp, Empher*** and ***Salret*** based on AUC-values of the model evaluation data? Use **mesotopo, soil_moist, soil_temp** and **soil_ph** as predictors.

Report the results in one short paragraph (max 5 sentences).

*Note: you can use sample-function to divide the data, e.g.*

```
ncal <- 0.7
s <- sample(nrow(data1), ncal)
data_cal <- data1[s,]
data_eval <- data1[-s,]
```

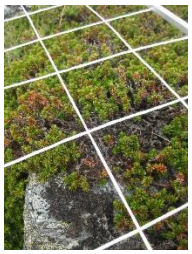
Question 2. What is the predictive performance of the GLM, GAM and GBM models for **veg_height** and **vasc_spr**? Again, build the models using calibration data and test the models using evaluation data. Use Spearman correlation -values as the evaluation metrics. Use the same set of predictors that you used in question 1).

Report the results in one short paragraph (max 5 sentences).

Question 3. Characterize **soil_moist, soil_temp, soil_ph, veg_height** and **vasc_spr** conditions along the **mesotopo**graphic gradient using GAM. Model the values of these five responses at the valley bottom (**mesotopo** 1), mid-slope (**mesotopo** 5) and ridge-top (**mesotopo** 10). Present the results as an informative figure.

Report the results in one short paragraph (max 5 sentences).

Question 4. Does the cover of *Empetrum hermaphroditum* (**Empher_cover**) have an effect on the **vasc_spr** when all other predictors are controlled for? Use the same set of predictors as used in question 1). Use all three modelling frameworks to test the hypothesis.



Report the results in one short paragraph (max 5 sentences).

*The main idea behind this question: as a dominant species **Empher_cover** might have a strong influence on the vegetation properties – can we see the effect? Please, test it!*