Relating vegetation to environmental variables with multilevel data
Hans Van Calster, Floris Vanderhaeghe, Els De Bie, Maud Raman & Paul Quataert

Context
- Analysis of data from a multilevel study (multilevel in the sense of studies with a nested design, repeated measurements or other forms of clustering) or meta-analysis of the raw data from a compilation of similar individual studies:
  - Plots (relevés) nested within study area (cluster)
  - Response variable: plant species cover or presence-absence, summarizing attributes of vegetation (species richness, ...,)
  - Plot level measurements / covariates: pH, groundwater depth, ...
- Such data should be analysed with a mixed model because of clustering
- However, the effect of a plot-level covariate may be different within versus between clusters or even differ from cluster to cluster. For instance because the effect of the plot-level factor is replaced or modified by the effect of other factors. A within-between formulation of the mixed model is needed in that case.

Adding a random effect
A (generalised) linear mixed model is a flexible method to account for the dependency of plots within the same cluster, through a random effect. The standard linear mixed model:

\[ y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + \epsilon_{ij} \] (1)

where:
- \( x_{ij} \): plot-level covariate
- \( u_j \): random intercept for cluster \( j \)
- \( \epsilon_{ij} \): plot-level residual
with assumptions:
\[ u_j \sim \mathcal{N}(0, \sigma_u^2) \]
\[ \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \]
Account for the possible intercorrelation of plots within the same cluster. The model should be checked for residual patterns against the plot-level covariate!

Adding within - between effects
- Equation 1 does not mean that plot-level covariates can be interpreted as within cluster effects (Van de Pol & Wright 2009, Bell & Jones 2015)!
- The model implicitly assumes that a one unit increase in \( x_{ij} \) has the same effect (\( \beta_1 \)) within (W) as between (B) clusters.
- A better approach is to cluster-mean center the plot-level covariate(s), so they only express variation within the cluster + include the cluster means as new covariate(s), which measure the variation between clusters:

\[ y_{ij} = \beta_0 + \beta_{W}(x_{ij} - \bar{x}_j) + \beta_{W} \bar{x}_j + u_j + \epsilon_{ij} \] (2)

Now, the slope of the plot-level covariate is allowed to be different within versus between clusters.
- To see if the difference is significant from zero, (2) may be rearranged:

\[ y_{ij} = \beta_0 + \beta_{W} x_{ij} + \beta_{W} - \beta_{B} \bar{x}_j + u_j + \epsilon_{ij} \] (3)

The model is easily extendable:
- To include cluster level covariates (\( + \beta_{CL} \))
- To more hierarchical levels, allowing more detailed analysis of spatial scales
- To include random slope effects, i.e. allow the coefficient associated with a within-cluster level covariate to vary by cluster

Conclusion
The within-between mixed model avoids biased conclusions and allows better model validation through clearer interpretation of random effects.

Literature cited

Example
- We used data from the Flanders wetlands sites database, which contains information on vegetation, hydrology and soil. In this example, 322 plots located in 54 nature reserves were analysed.
- We modelled the presence-absence of Molinia caerulea (binomial distribution + logit-link) as a function of the cation exchange capacity of the soil (lme4 package in R). Nature reserve is modelled through the random intercept.

Results
- Only weak evidence for a negative effect of CEC from model 1: -0.1 (-0.17 -0.03).
- A strong and negative effect of high CEC at the between nature reserve scale [-0.41 (-0.76 -0.22)], which differed significantly from the very weak effect at the within nature reserve scale: -0.06 (-0.13 0).