Summary

This report is an intuitive introduction to the mixed models regression technique. We will explain the underlying principles by means of a simulation study and illustrate the technique on two examples from the watercourse sediment monitoring network.

As in classical regression, the regression equation of a mixed effects regression model describes the mathematical relationship between an outcome variable (eg concentration of a pollutant in the sediment) and a set of explanatory variables (year, ecoregion, soil type, ...). However, in mixed models some coefficients are composed of the sum of a fixed and a random (stochastic) part. The random term models individual local deviations from the overall effect as quantified by the fixed term. As an example, we can consider a trend model in which the local trend of a pollutant in individual streams is described as random fluctuations around the global trend, representing the average over all streams.

Therefore, mixed models can describe a situation in which the local pattern deviates from the average. Since individuals can be characterized insufficiently, the effect of an individual will be modelled as a random variation around the average. Mixed models then combine fixed parameters, quantifying the global effects of the explanatory variables on the outcome variable (fixed effects), with stochastic parameters, specifying individual effects of the explanatory variable (random effects). In this way, mixed models give great flexibility to add local variations to a general pattern.

To analyze repeated measurements on the same object, the use of mixed models is also necessary. Measurements of the same object will be more alike than measurements from other objects. When the concentration of a substance in a particular measurement site is high, we may expect it to be high again in a next visit, or at least higher than for a location where the original concentration was low. Classical regression models cannot take into account the correlation between repeated observations on the same object. In contrast, random effects in a mixed model characterize the specific characteristics of the measured objects. In this way, the correlation between observations on the same object is automatically accounted for.

This text aims to bring insight into the possibilities and underlying principles of mixed models using several simulation studies. Step by step the models are made more complex and the implications visualized. We start with a simple linear trend model in which an outcome variable changes constantly over time. We show how multiple regression lines can be combined in a single formula by entering an extra categorical variable or factor, allowing for a different effect of the explanatory variable on the outcome of interest (eg, a different trend depending on ecoregion). Mixed models regression is the next logical step. Adding random variables allows the (slope of a) regression line to vary across measurement locations without explicitely specifying the underlying cause. This theory will be illustrated on the evolution of concentrations of two metals (cadmium and arsenic) from the watercourse sediment monitoring network.