

Derivation of Typical Assemblages

Jonathan Jupke

Oktober 01, 2020

Contents

1	Introduction	3
2	Harmonizing taxa names	3
3	What is the optimal taxonomic level?	3

1 Introduction

Here we describe the methods we used to derive typical assemblages of macro-invertebrates and diatoms for selected river types across Europe. We also describe and briefly discuss the results.

2 Harmonizing taxa names

3 What is the optimal taxonomic level?

One result of the last progress review for Get Real (held on the 29.04.2020) was that taxa in the TA should be included on their respective optimal taxonomic level instead of using one level (e.g. Genus) for all. *Oligochaetes*, for example, are usually only determined to subclass level, which should not prevent them to be part of a TA in the case that they are common in a given river type. Thus, the question arises: given a data set, what is the optimal taxonomic level to represent a specific taxon?

To establish the optimal level, we used a hierarchical approach. First, we removed all observations from Phyla and Classes that were not present in all data sets. We assumed that these represented differences in sampling rather than in communities. That left us with the classes Clitellata (Annelida), Insecta, Malacostraca (Arthropoda), Bivalvia and Gastropoda (Mollusca).

In the following, a higher taxonomic level refers to levels with higher resolution, i.e. species is the highest taxonomic level and kingdom the lowest. For each taxon, we calculated the percentage of observations that are represented at each higher level. For example, 4.12% of observations from the order *Lepidoptera* are at the species level, 74.77% at the genus level, 7.75% at the family level, and 13.35% at the order level. Now given a threshold X, which is to be determined, we would call a taxon optimally represented at a certain taxonomic level if less than X% are represented by higher levels. For example, *Lepidoptera* would be represented on order level if $X > 4.12 + 74.77 + 7.75 = 86.64\%$. As there are no theoretical grounds on which to base such a threshold value we searched for noticeable patterns in the data (Figure 1). The most noticeable jump occurs between 85 and 86%. It occurs because for $X > 86$ *Chironomidae* are represented at the family level. Hence, we used 85% as threshold. Observations that were missed by this procedure, e.g. observations of *Chironomidae* at the family level, were included at their respective level.

```
source("../r_code/test.R")
print(vec_test)
```

```
## [1] "C:/Users/jonat/Documents/01_Uni/02_getreal/005_documents/2020_10_Zwischenbericht"
```