UNIVERSITY OF NOVA GORICA
GRADUATE SCHOOL

Uroš ŽIBRAT

DEVELOPMENT OF A NEW BIOTIC INDEX BASED ON HYPORHEIC FAUNA FOR WATER QUALITY DETERMINATION IN LOTIC ECOSYSTEMS

Dissertation

Supervisor: prof. dr. Anton Brancelj

Nova Gorica 2012

Auch dem Wasser darf es in Kanälen
Nie am Laufe, nie an Reine fehlen.

J. W. von Goethe

ABSTRACT

We assess water quality and the status of water bodies with numerous variables, both abiotic and biotic. Up to date, few water quality assessment and ecoremediation measures took the surface water – groundwater ecotone (i.e. the hyporheic zone) into consideration. We developed a new index based on hyporheic copepod fauna. The copepod biodiversity in the hyporheic of 26 Slovenian rivers, belonging to all five river basins, was assessed, and 49 species, belonging to 6 families were found. Species accumulation curves and richness estimators indicated that our sampling effort was sufficient, all estimates pointed toward a richness of 50 species. Two biodiversity hotspots were identified, but one of them was found to have an uneven species distribution, indicating human-impacted sites. One species (*Moraria radovnae)* appears to be endemic. Differences in relative abundances of the most common species accounted for most of the differences between river basins. The hyporheic copepod community, combined with an accurate typology and diversity measures, and while considering the limitations of our model, can be used for water quality assessment. The obtained final model is most accurate in mid-range classes, while being inaccurate at extreme values. By developing this index and promoting its use in water quality assessment and environmental impact studies, we can expand currently applied methods.

Keywords: hyporheic copepods, water quality assessment, biodiversity, data mining.

POVZETEK

Kakovost vode in stanje vodnih teles ocenjujemo s številnimi spremenljivkami, tako abiotskimi kot biotskimi. Doslej je le malo metod za oceno kakovosti vode in ekoremediacijskih ukrepov upoštevalo ekoton med površinskimi in podzemnimi vodami. Razvili smo nov indeks, ki temelji na združbi ceponožnih rakov v rečnem hiporeiku. Raziskali smo biotsko pestrost ceponožnih rakov 26 slovenskih rek, ki iz vseh petih porečij. Našli smo 49 vrst, ki pripadajo 6 družinam. Krivulje akumulacije vrst in ocene vrstnega bogastva nakazujejo, da je bil naš napor vzorčenja zadosten za dobro oceno števila vrst. Vse ocene nakazujejo prisotnost 50 vrst ceponožnih rakov. Odkrili smo dve vroči točki biodiverzitete, vendar smo pri eni ugotovili neenakomerno razporeditev vrst, kar nakazuje na vpliv človeških dejavnosti. Za eno vrsto (*Moraria radovnae)* trenutno izgleda, da je endemična. Razlike v relativnih abundancah najpogostejših vrst so največ prispevale k razlikam med porečji. Ocenjevanje kvalitete površinske vode je možno z uporabo združbe ceponožnih rakov v hiporeiku. Za to so potrebni natančni tipologija površinskih voda in ustrezni diverzitetni indeksi, upoštevati pa je potrebno tudi omejitve našega modela. Končni model je najbolj zanesljiv v srednjih razredih, medtem ko je netočen pri ekstremnih vrednostih. Z razvojem tega indeksa in spodbujanjem njegove uporabe pri metodah ocenjevanja kakovosti površinskih voda in presojah vplivov na okolje, želimo razširiti metodologijo, ki je trenutno v uporabi.

Ključne besede: ceponožni raki, hiporeik, ocenjevanje kvalitete voda, biodiverziteta, biotska pestrost, strojno učenje.

CONTENTS

I LIST OF FIGURES

II LIST OF TABLES

III LIST OF APPENDICES

1 INTRODUCTION

1.1 The importance of subsurface water

Water is essential for the continuous existence and development of all living organisms. Even though we are aware of this fact, conflicts between nature conservation and economic interests are becoming more common. The later frequently lead to environmental pollution and worsening ecosystem health (Costanza *et al.* 1992, Sampat 2000). Superficial freshwater ecosystems are especially exposed and under increasing amounts of stress. These ecosystems are, through their hyporheic zones, connected with aquifers, which hold 97 % of the planet's liquid freshwater. Aquifers are essential to life above ground; they play a vital role in replenishing rivers and wetlands and thus in nurturing the life of the land and air as well. If too much groundwater has been depleted, riverbeds may dry out and wetlands become desiccated. By soaking up excess water after heavy rainfall, aquifers can prevent surface waters from flooding. On a global scale, approximately 2 billion people rely on groundwater as their only source of drinking water. Considering the direct connection between surface and groundwater, the importance of methods for assessing the status of surface waters becomes evident. The significance of the hyporheic zone can be highlighted by the following example: at any point in time only 1 % of the water in the Mississippi River system is running downstream; the other 99 % lies underground (UNEP 1996, Sampat 2000).

Water quality is affected by several complex factors. In addition, we use numerous variables to describe the status of water bodies and their surrounding area. The term “water quality” is therefore comparatively hard to define. “Water quality” can be regarded as a neutral term that relates to the composition of water as affected by natural processes and human activities. It depends on not only water’s chemical, but also its biological, physical, and radiological condition. The quality of water is also related to specific use, and is usually measured in terms of constituent concentrations. The level of water quality is based upon the evaluation of measured quantities and parameters, which then are compared to water quality standards, objectives or criteria (Zucker *et al.* 1998). Most often, we use the term in relation to the water’s suitability for drinking.

1.2 The hyporheic zone

The hyporheic extends vertically and laterally from the river channel. Water in streams and rivers is continuously exchanged between the active channel and subsurface (hyporheic) flowpaths (Fig. 1). This interaction can be fast enough that, within several kilometers, water in streams is completely exchanged with interstitial water of the hyporheic zone. Dissolved material and organisms are also continuously exchanged between surface and groundwater. Thus surface water is in close contact with chemically reactive mineral coatings and microbial communities in interstitial (i.e. below surface) waters. This process has the effect of enhancing biogeochemical reactions and downstream water quality (Williams and Hynes 1974, Gibert *et al*. 1990, Vervier *et al*. 1992, Gibert *et al*. 1994, Harvey *et al.* 1996, Orghidan 2010).

Several models emphasize the “storage zone” function of the hyporheic. This zone delays downstream transport (i.e. hydraulic conductivity is comparatively low, approximately 10^{-6} ms^{-1}, and water retention time increases) and enhances biogeochemical reactions. Water retention time in interstitial habitats is under the influence of particle size of the substrate. Small particles (sand and clay) can clog the space between larger particles (pebbles and gravel) and hydraulic conductivity decreases. Glacial clay in alpine valleys can almost completely block the hyporheic zone, hydraulic conductivity is below 10^{-8} ms^{-1} and water exchange at a minimum (Runkel *et al.* 1996; Mullholland *et al*. 1997).

Interstitial pore size and hydrological exchange between river and subsurface flowpaths shape the hydrological, physical, and chemical conditions in the hyporheic zone (Marmonier and Creuzé Des Châtelliers 1991, Boulton and Stanley 1995, Bork *et al.* 2009). The hyporheic zone can be viewed as the subset of fine-scale (within one reach) interactions between surface water in the channel and groundwater, within the context of large-scale exchanges in drainage basins. Large scale exchanges account for water loss and gain between the main channel and aquifers further inland. Nevertheless, the greatest amount of water is exchanged in smaller, fine-scale, hyporheic flowpaths (centimeters to tens of meters). The size of these flowpaths is subject to seasonal and storm-influenced fluctuations. Comparatively rapid inputs of oxygen and organic carbon enhance rates of microbially mediated chemical processes. Examples of this effect include nitrification, microbial uptake of dissolved organic carbon and oxidation of manganese in the hyporheic zone (Wondzell and Swanson 1996; Angradi and Hood 1998).

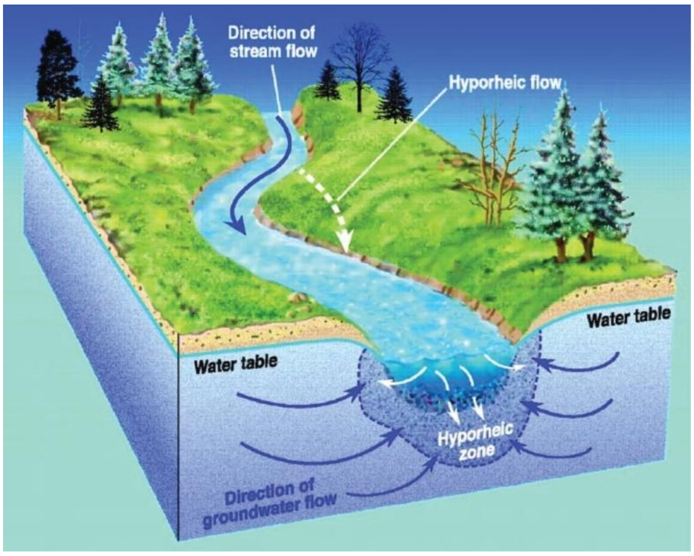


Figure 1: *Schematic illustration of the hyporheic zone (after Alley et al. 2002).*

The degree to which groundwater affects in-stream ecology is governed by the so-called base flow index (BFI) (Sear *et al.* 1999). A high BFI represents a flow regime of surface water, which is volumetrically dominated by groundwater discharge. In such cases, groundwater is the predominant influence on in-stream ecology at almost all times. Flood events present a special case, where even in streams with an otherwise high BFI, the flow regime can be reversed, and surface water discharges into subsurface flow paths. In middle- to low-BFI streams, the variations in flow regime are more dynamic, with changes occurring over very short timescales (Maddock *et al.* 1995, Malcolm *et al.* 2003). Relatively modest fluctuations in stream discharge can give rise to profound changes in groundwater movement in the hyporheic (Alden and Munster 1997).

Subsurface hydrological exchange in the hyporheic can be divided into three modes (after Kaplan and Newbold 2000):

- Upwelling: groundwater discharges from an adjoining aquifer into the main channel.
- Downwelling: surface water flows through the hyporheic into the underlying aquifer.
- Shallow exchange: surface waters enter the bed sediment and exit back into the channel a short distance away. For example, water enters the subsurface through the upstream face of a stony riffle (i.e. small-scale down-welling), re-surfacing through the downstream edge of the riffle (i.e. small-scale upwelling), flowing into the adjoining pool (Younger 2007).

A switch between up- and down-welling can occur momentarily when a flood wave passes through the stream. Therefore, water in the hyporheic will reflect the recent history of stage fluctuations in the river.

1.2.1 Ecology of the hyporheic

Microbes govern the major part of organic matter decomposition in interstitial waters (Pusch 1996, Craft *et al.* 2002). However, some studies showed that invertebrates contribute greatly to hyporheic nutrient cycling (Marshall and Hall Jr. 2004). Microcrustaceans (Copepoda and Ostracoda) are a diverse and abundant component of interstitial communities in gravelbed rivers and deeper groundwater, many of them being true groundwater species (stygobionts) (Marmonier and Creuzé Des Châtelliers 1992, Pospisil 1994, Rouch and Danielopol 1997, Hakenkamp and Palmer 2000, Galassi *et al.* 2009, Stoch *et al.* 2009). These crustaceans are highly mobile through a wide range of sediment pore sizes and some species are tolerant to low oxygen concentrations. Knowledge of their trophic role is lacking; probably they are preyed upon by macroinvertebrates and stimulate microbial growth by excrement deposition (Hakenhamp and Palmer 2000).

Oxygen concentrations, temperature, and nutrient availability are the main limiting factors for hyporheic organisms. Generally, depth in the hyporheic zone is characterized by a gradient from high oxygen, high nutrient availability and unstable temperatures to low oxygen, fewer nutrients and more stable temperatures. The "ideal" niche for animals would therefore be in the zone of approximately equal mixing of surface and groundwater, with still enough oxygen for most organisms, stable temperatures throughout the year and a comparatively constant supply of nutrients (Younger 2007). Downwelling waters supply nutrients from the breakdown of plant and animal debris. These nutrients are usable by animals inhabiting the hyporheic zone, such as copepods. For the microbial community such debris is too large and cannot be utilized. Instead, they feed on dissolved organic compounds and assimilate inorganic molecules containing oxygen (such as phosphate, nitrate and sulphate). Upwelling groundwater usually contains relatively high concentrations of inorganic nutrients (Findlay and Sobczak 2000, Younger 2007).

The diversity of organisms in the hyporheic is governed by micro-scale distributions in temperature, dissolved oxygen, and nutrient availability. For larger animals, such as crustaceans, pore size distribution of the sediment also exerts an important control. Hakenkamp and Palmer (2000) report that crustaceans are more common in gravels than in fine-grained sand. Figure 2 summarizes the controls on the hyporheic fauna.

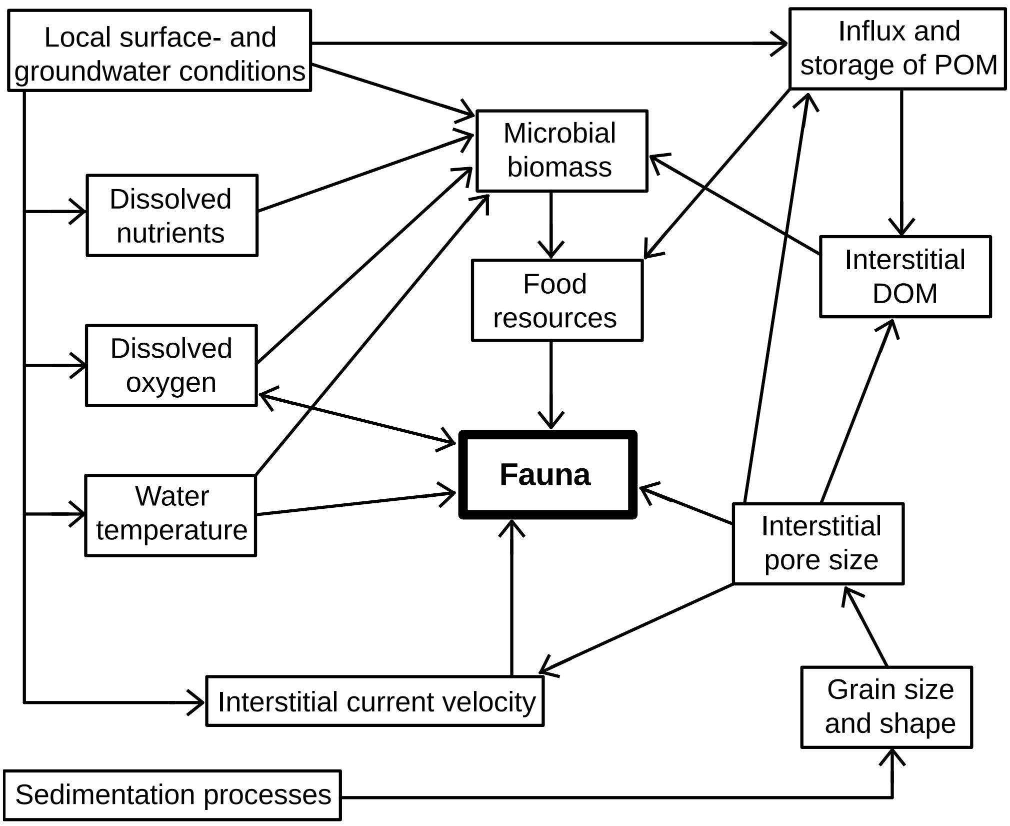


***Figure 2**: The principal factors governing the macrofauna of the hyporheic zone (after Boulton 2000). POM: particulate organic matter.*

1.3 Water quality assessment

Many parameters affect water quality and only a combination of different methods can yield reliable results. The best results can be obtained by combining physical, chemical and biological methods, and evaluation of the status of entire ecosystems, the functional association of abiotic and biotic factors. Water quality is directly connected to ecosystem health and several methods were developed to determine the status of different ecosystems (Urbanič and Toman 2003, Jørgensen *et al.* 2005). Each of these methods has its benefits and weak points (Table 1).

Table 1*: Benefits and weak points of different water quality assessment methods (Urbanič and Toman 2003, Jørgensen et al. 2005)*

Physical and chemical	Biological
Benefits	
Accurate detection of changes over time Accurate detection of pollutants and the direction of their flows Can be used on all water bodies, even groundwater Methods are standardized	Response of organisms to long term, low concentration pollution Basis for assessing the physical degradation of the environment Bioaccumulation and biomagnifications
Weaknesses	
High detection threshold* Possible contamination of samples Limited usefulness for continuous surveys** Expensive	Cannot detect the direction of flow of pollutants Undefined or unknown for groundwater Limited standardization***

* Too low concentrations of pollutants are not detected. Certain methods (e.g. HPLC and GC-MS) omit this limitation,
** Time windows for detecting pollution events are comparatively short; one-time pollution events are rarely detected.
*** Differences in community compositions among similar ecosystems from different geographical areas (for example the hyporheic from alpine and lowland rivers) make standardizations difficult, if not impossible.

Both groups of methods for assessing water quality of surface waters (Table 1) have a common weakness, i.e. water retention time. Physical and chemical methods specially suffer from this deficiency. With this group of methods, chance becomes important, since detection of pollutants depends on time of sampling. If sampling is performed too late, a pollutant spill can be missed or pollutants become diluted. The importance of stochastic events, combined with high costs, also limits the usefulness of these methods for continuous surveys. Biological methods are more appropriate or accurate, because effects of pollutants can still be seen even weeks after their initial introduction. Yet, in fast flowing rivers, such disturbances can have a comparatively low impact, since animals are exposed to pollutants for a short time span (Hoehn 2001, Danielopol *et al.* 2009, Schmid-Araya 2009).

Among biological methods, the most often used method is assessment based on benthic macroinvertebrates. Methods based on macrophytes, fish, algae and protozoans are also employed (Jiang 2005, Jiang and Shen 2005, Haunschmid and Jagsch 2006, Orlando-Bonaca *et al*. 2008). Lately, ecomorphological methods are gaining in importance. This set of methods also assesses the surrounding area, its structure and use (Urbanič and Toman 2003, Jørgensen *et al.* 2005). Even though these methods assess different parameters, they all consider only surface dwelling organisms. Subsurface organisms are ignored, even though they are hypothetically more sensitive to certain ecological disturbances.

Surface water bodies are connected to subsurface ecosystems further inland. The transition area between two ecosystems is called an ecotone. Between surface water bodies and groundwater (the phreatic zone), this ecotone is referred to as the hyporheic. The hyporheic zone is directly under the influence of surface and groundwater. Therefore, both surface and subsurface dwelling animals can be found there. In the phreatic zone, on the other hand, surface animals can survive for a limited amount of time or are absent, while the influence of surface water is also constrained or zero. In these two zones water retention time is longer than in the river. Since animals are exposed to pollutants for longer periods, a stronger effect would be expected (Newman 2009).

1.4 The concept of ecosystem health

Ecosystem health is a metaphor. And therein lays its main appeal. We apply the notion of human health to ecosystems and gain a paradigm for viewing ecological policy questions. People have an inherent understanding of personal health, and by extension, most people instinctively envision a "healthy" ecosystem. Such an ecosystem usually encompasses pristine conditions or is at least minimally disturbed by human activity (Ryder 1990, Meyer 1997) (Figure 3). But, since the term "ecosystem health" is used by scientists and stake holders of many different fields of expertise, we lack a universal definition of this concept. Karr and Chu (1999) define ecosystem health as the preferred state (i.e. what experts think a modified ecosystem should at its best look like) of ecosystems disturbed by human activity (e.g. farm land, urban environments, airports, managed forests). They also propose the use of the term ecosystem integrity, denoting an unimpaired condition, with no or little influence from human activity. Natural ecosystems, with a high degree of integrity, would continue to function in the same way if humans were removed (Anderson 1991). Since the vast majority of ecosystems are not pristine, most ecological policy debates concern ecosystem health, rather than integrity. Costanza *et al.* (1992) emphasize that it is necessary to consider all or at least most of the definitions simultaneously. They summarize the concept definition of ecosystem health as: (1) homeostasis; (2) absence of disease; (3) diversity or complexity; (4) stability or resilience; (5) vigor or scope for growth; (6) balance between system components.

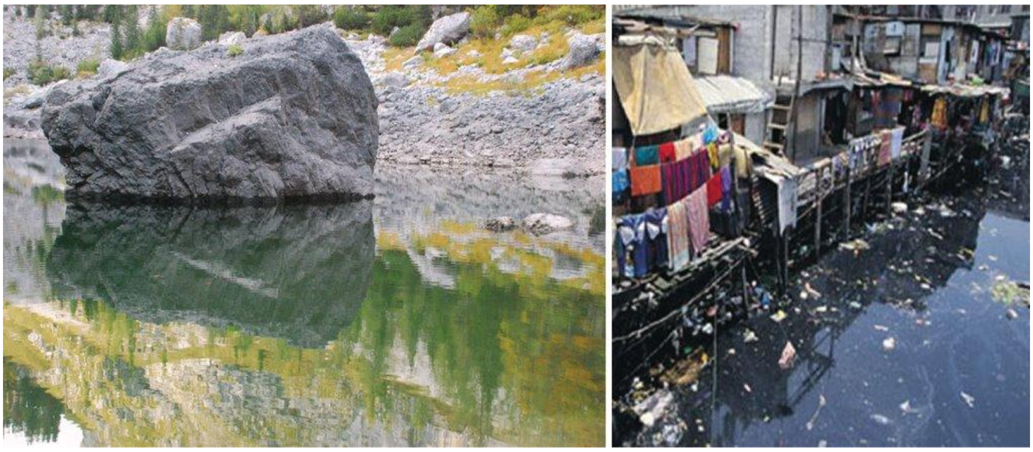


Figure 3: *The ecosystem health metaphor. Most people see the left hand picture as depicting a healthy, pristine ecosystem. The right hand picture is understood to present an unhealthy or polluted system (left photo: U. Žibrat; right photo: www.supergreenme.com).*

1.4.1 Ecosystem health indicators

There are several indices in use for each group of parameters for surface water quality assessment. Most work has been done on biological parameters and therefore most indices are concerned with this group. Biological indices can be divided into six groups, according to the amount and type of information they consider (Fath *et al.* 2001, Fath *et al.* 2004, Jørgensen *et al.* 2005, Orlando-Bonaca *et al.* 2008, Steube *et al.* 2009):

- Indices based on indicator species (Bellan's pollution index, Marine Biotic Index - AMBI, Bentix, Benthic Response Index - BRI):
  The presence or absence of indicator species, and their dominance are associated with an environmental deterioration;

- Indices based on ecological strategies (Nematodes/Copepods Index - NCI, Infaunal index):
  Environmental stress assessment accounts for the ecological strategies followed by different organisms;
- Diversity indices (Shannon-Wiener, Simpson, Margalef):
  Changes in diversity reflect disturbances;
- Indices based on species biomass and abundance (Abundance Biomass Comparison - ABC Method):
  The variance of organism's biomass and abundance are applied as a measure of environmental disturbance;
- Indices integrating all environment information (Trophic index, Index of Biotic Integrity - IBI, Estuarine Ecological Index - EBI, Saprobia):
  These indices try to integrate several metrics, for example diversity, dominance, ecological strategies, indicator species, sediment particle size and sand/silt ratio;
- Holistic or thermodynamic indices (Exergy, Ascendency, Dissipation, Kullback diversity measure):
  Assessment is based on the entire ecological network, individual species and their ecological preferences are not considered.

Each of these indices has its strong and weak properties. Experience has taught us to use several indices to assess an ecosystems status. By combining different methods more information on each ecosystem can be gathered and better and more accurate conservation or restoration measures undertaken.

Compared to surface waters, groundwater ecosystems are depleted in certain bioindicator organisms, such as macroinvertebrates, or completely lack those in the case of macrophytes and algae. In addition, knowledge on the drivers for the patchy distribution of organisms is poor, which limits predictive conclusions (Steube *et al.* 2009). There is also a significant lack in knowledge on the taxonomy, autecology and physiology of groundwater organisms. Nevertheless, several authors have stressed the importance of groundwater invertebrates for biomonitoring purposes and assessment of ecological status of the hyporheic zone (Mösslacher 1998; Matzke *et al.* 2005; Hahn 2006; Schmidt *et al.* 2007).

1.5 Study organisms

With more than 2800 species, Copepoda represent the most species-rich group of freshwater crustaceans (Boxshall and Defaye 2008). The taxonomic diversity of copepods in surface water is reflected in groundwater, where copepods are represented by approximately 1000 species and subspecies. Together with Ostracoda and Amphipda they outnumber all other invertebrate groups living in subsurface environments (Galassi *et al.* 2009).

The Copepoda include 10 orders, only 4 of which (Harpacticoida, Cyclopoida, Calanoida and Gelyeloida) include free-living (i.e. non-parasitic) freshwater representatives. Of these 4, only Harpacticoida and Cyclopoida are widely distributed in subsurface freshwater habitats. Together they represent more than 70% of all freshwater copepod species. Copepods show distinct differences in microhabitat preference (Galassi 2001) and sensitivity to human-induced disturbances, such as changes in water chemistry and hydrological regime of groundwater (Rundle and Ramsay 1997, Paran *et al.* 2005). Furthermore, copepods may be useful as biological indicators of groundwater – surface water connectivity (Malard *et al.* 1994, Di Lorenzo *et al.* 2005). Meiofauna (i.e. organisms living in spaces among gravel and sand) are known to be a significant component of the heterotrophic assemblage (Giere 1993, Freckman *et al.* 1997). Biological activities (feeding, excretion, movement) can influence the transfer and cycling of material and energy through the ecosystem (Aller and Aller 1992). Although the knowledge of copepod ecological functions in groundwater is lacking, there are indications that they play a significant role in groundwater ecosystems (Gibert and Deharveng 2002).

1.5.1 Copepods in ecotoxicological studies

Amphipod and isopod crustaceans, oligochaetes and chironomid larvae are usually used in sediment toxicity assessment studies (USEPA 2000a). Yet, to the best of our knowledge, only poor information is available on the effects of pollutants on species living in the hyporheic zone (Williams and Fulthorpe 2003). These pollutants include pesticides, persistent organic pollutants and heavy metals. In agricultural areas groundwater may be enriched with nutrients (such as nitrogen and phosphorous), leading to eutrophication of surface waters. On the other hand, polluted surface water can alter hyporheic communities (Mösslacher *et al.* 2001). Harpacticoid copepods (Figure 4) are almost omnipresent in hyporheic communities, often in comparatively large abundances. Bengtsson (1978) used them for assessing acute effects of marine pollutants. Several authors have shown harpacticoid copepods to be sensitive to environmental pollution, especially to metal pollution (Van Damme *et al.* 1984; Burton *et al*. 2002; Brown *et al.* 2005).

Figure 4: *The harpacticoid copepod* Bryocamptus zschokkei *(Schmeil, 1893). The photograph is of an ovigerous female, with protozoan epibionts attached to pereiopods P3 and P4 (photo: U. Žibrat).*

1.6 The reference condition

A critical element in the process of biological assessment of the effect of human activity is estimating biological status in the absence of human disturbance. Most biological assessments are based on the concept of comparing the current state to conditions in the absence of human disturbance or alteration. The state under study is compared to a pristine and unpolluted state (Jackson and Davis 1995, Davies and Jackson 2006), which is used to gauge the effects of human activity (Stoddard *et al.* 2006). This pristine state is referred to as *reference condition* or *reference state*. Reference or benchmark locations are chosen to be as close as possible to the state of an environment undisturbed by human activity. These locations are often chosen without a particular impact in mind, but to represent what a water body could be like in the absence of human activity.

The main problem with this approach is the definition of a reference state, since it can be defined for a variety of purposes related to biological assessment. For example, it can mean the absence of any human disturbance, the best remaining condition in a specific region or the upstream condition when assessing impacts of a point-source discharge into a stream. In this regard, four different aspects of reference condition can be defined (Stoddard *et al.* 2006, Hawkins *et al.* 2010): historical condition, least disturbed condition (LDC), minimally disturbed condition (MDC) and best attainable condition (BAC) (for details see below).

The Water Framework Directive (WFD) (Directive 2000/60/EC) defines reference conditions for surface waters in terms of "no or minimal anthropogenic stress" and satisfying the following criteria: (1) reflecting totally, or nearly, undisturbed conditions for hydromorphological elements, general physiochemical elements, and biological-quality elements; (2) having concentrations of specific synthetic pollutants close to zero or below the limit of detection of the most advanced analytical techniques in general use; (3) exhibiting concentrations of specific non-synthetic pollutants within the range normally associated with background levels. Reference conditions are to be linked to stream typologies, and the population of reference sites should represent, as well as possible, the full range of conditions that are expected to occur naturally within the stream type. A reference condition can be spatially based (the condition of existing sites meeting the above criteria), based on modeling or developed using a combination of these. Using these, all surface water bodies should be classified into ecological-quality classes.

The Water framework directive requires the establishment of a water body typology based on abiotic factors. In Slovenia, such a typology was published in 2008 (MOP 2008). This particular typology is based on ecoregions, bioregions, average yearly discharge, catchment size and spring influence. In accordance with WFD guidelines, reference conditions for each river type were also defined. This typology is to be used in all water quality assessment schemes.

1.6.1 Defining the reference condition

It is becoming increasingly difficult to locate sites that could represent the undisturbed or natural state and be used for estimating the reference condition. Often sampling sites are chosen for estimating reference conditions because they are "the best of what's left". The entire population of possible sites has been degraded by human activity; therefore, these sites bare only a passing resemblance to natural conditions. Consequently, the reference condition departs from the natural for an unknown amount.

During monitoring programs, we measure and interpret multiple ecological indicators. These indicators need to be compared to a reference condition, which is most commonly defined by a range of indicator or index scores. In this sense, "reference condition" describes a distribution, rather than a single absolute value (Figure 5) (Stoddard *et al.* 2006, Hawkins *et al.* 2010). For any given index, the range of values is defined by sampling error and natural variability. A set of sites, disturbed or undisturbed, will always exhibit a range of biological attributes. By sampling several sites used for estimating reference conditions, this distribution can be described for a particular location or habitat type. Parameters in this distribution can then be selected as criteria for classifying and assessing the condition of individual sites. The tails of this distribution and the final thresholds are under strong influence of the methods chosen to measure and describe the reference distribution. By being restrictive in the level of acceptable disturbance, the distribution can be constrained and criteria increased. This approach is applicable in all habitats, including both surface and groundwater. The main drawback for groundwater (including the hyporheic) is lack of quality data.

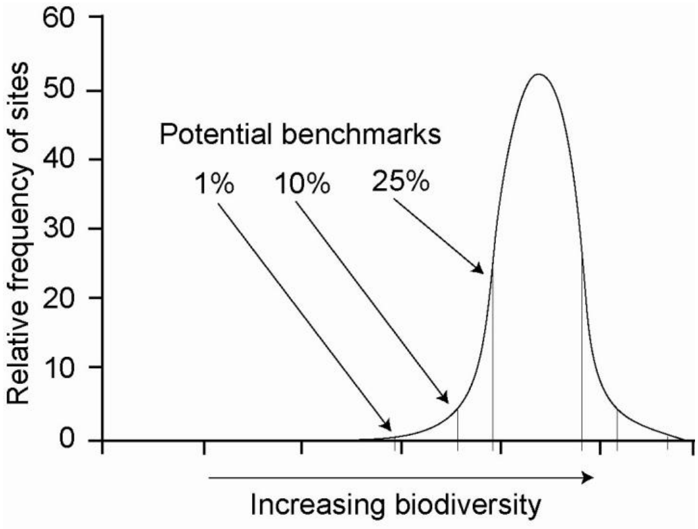


Figure 5: *The range of ecological conditions found under reference conditions. In this case the distribution is represented by a theoretical biological index. The range of this distribution results from natural variability and can be used to set criteria for defining classes of ecological condition. Benchmarks can be below (bad status) or above (good status) the selected thresholds (e.g. 1 %) (after Stoddard* et al. *2006).*

The definitions of reference condition in use today most often denote a state that is different from a natural one. Stoddard *et al.* (2006) therefore propose the use of a modified term, reference condition for biological integrity or RC(BI). This term would be reserved for “naturalness” or “biological integrity” (in accordance with the U.S. Clean Water Act) (USEPA 1972), even though it can be only approximated in most parts of the world. They also suggest the use of the following specific terms for reference conditions other than in relation to biotic integrity: Minimally Disturbed Condition (MDC), Historical Condition (HC), Least Disturbed Condition (LDC), and Best Attainable Condition (BAC).

1.6.1.1 Minimally Disturbed Condition (MDC)

This term describes the condition of sites in the absence of significant human disturbance. It is our best approximation of biotic integrity, since finding sampling sites that are undisturbed by human activity is impossible. Therefore, the concept of minimal disturbance is defined as “condition in the presence of atmospheric contaminants well below the threshold for effects, but nonetheless present”. MDC also recognizes that some natural variability in indicators will always occur. One important aspect of a MDC distribution is its consistency over time. Even though geologic, climatic, and ecological fluctuations will change (in a geological sense in a short time span) the characteristics of individual sites, the range of MDC should be nearly invariable. This distribution can therefore serve as a nearly invariant benchmark by which to judge current condition.

1.6.1.2 Historical Condition (HC)

This term describes the condition of sites in their history. If the chosen point in history is before the start of human disturbance, it can be our best approximation of RC(BI). However, other historical reference points are possible (e.g. pre-industrial, pre-settlement, pre-regulation) and can be chosen in regard to local needs. The EU WFD defines reference conditions in this sense, as a “state in the present or past corresponding to very low pressure, without the effects of major industrialization, urbanization and intensification of agriculture, and with only very minor modification of physiochemistry, hydromorphology and biology.“ This definition implies no fixed date, but the condition is in general believed to have existed in Great Britain before 1850, i.e. prior to the industrial revolution (Wallin *et al.* 2003). In North America, a different approach is used, as the HC is defined as a period before settlement. It includes impacts of indigenous people, but excludes impacts of European immigrants (Hughes *et al.* 1998). This definition also does not imply a fixed date. Yet, in contrast to the WFD definition, the time period used as a benchmark, changes with the westward migration of settlers. Nevertheless, if no historical data exists, then this method is not applicable for determining reference conditions. Such is the case with aquifers and the hyporheic.

1.6.1.3 Least Disturbed Condition (LDC)

The LDC is in conjunction with the best available physical, chemical, and biological habitat conditions given today's state of the landscape. To describe the LDC, data, collected at sites selected according to certain criteria denoting what is "best" or least disturbed by human activity, needs to be evaluated (Hughes *et al.* 1986). These criteria are developed with the aim to establish the least amount of ambient human impact and vary from region to region, as the characteristics of the landscape and human use vary. For the same reason LDC varies in time, unlike MDC. As the ecological conditions of the best available sites changes through time, so does our measure of LDC.

1.6.1.4 Best Attainable Condition (BAC)

The BAC is equivalent to the expected ecological condition of LDC sites, if the best possible management practices were in effect for a certain period. Thus, BAC sites would be locations, where the impact of land use on biota is minimized. This definition is therefore somewhat theoretical, as it encompasses management goals, best available technology, prevailing use of landscape and public interest and commitment to achieving environmental goals. MDC and LDC, respectively, define upper and lower limits of BAC (Figure 6). Similar to LDC, BAC also varies over time.

Differing levels of human disturbance in different ecological regions lead to a situation where LDC sites remaining in a region describe different definitions of reference condition (Fig. 5). In Stream Group C the level of degradation is higher than in Stream Group B. For Group C a reference condition might be a condition that does not exist, but could be achieved with adequate management (illustrated as "Best attainable"). In contrast, the condition at the LDC sites for Group B might be a reasonable reference condition for these streams (after Stoddard *et al.* 2006).

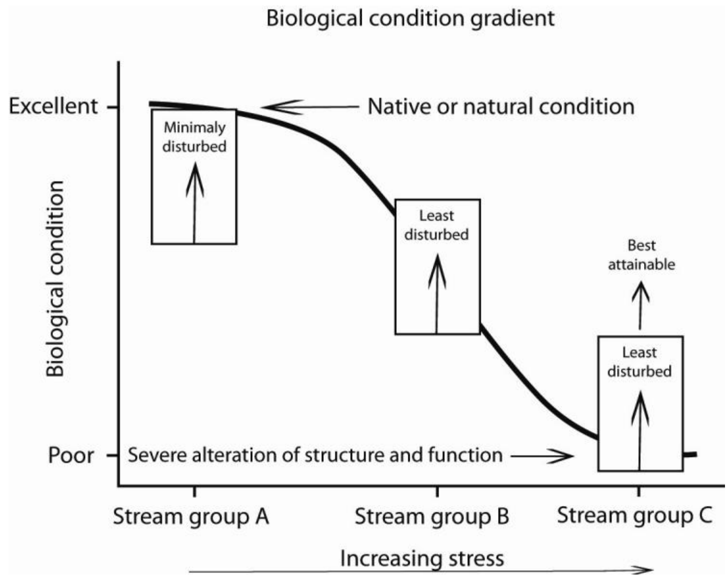


Figure 6: *Three types of reference conditions along a biological condition gradient (after Stoddard* et al. *2009).*

1.6.2 Methods of estimating reference conditions

1.6.2.1 Reference-site approach

The most common approach for estimating reference conditions is the so-called "reference-site approach" (Bailey *et al.* 2003). This method quantifies the biological condition at a set of sites that are either minimally or least disturbed by human activity. This method requires the form of reference condition (MDC, LDC, HC and BAC) the reference sites represent to be clearly defined. The preferred approach for estimating MDC and LDC is to use a set of criteria for site selection that exclude data on resident biota. Bailey *et al.* (2003) argue that the structure of the biotic assemblage should not be used to classify sites as reference or non-reference. By following this guideline, we avoid any preconceived notions about the structure and composition of biotic assemblages at a reference site. The main goal of the reference-site approach is to describe the amount of variability present at sites in the absence of human disturbance, and we cannot know in advance how much variation is typical for any assemblage among a population of reference sites.

Certain exceptions to the non-use of biotic data for reference-site selection are warranted. Organisms can function as indicators of the presence of certain stressors, such as toxic compounds. The use of organisms to assess the presence of toxics can be more cost effective, than direct chemical measurements. Alien or introduced species can also act as stressors. Therefore, one may wish to focus on those reference sites with all native assemblages.

1.6.2.2 Interpreting historical condition

Sometimes records from earlier times are on hand, whether from stream samples, museum collections, journals, photographs or personal notes (Hughes *et al.* 1998). Historical conditions can sometimes be inferred by analyzing indicators that maintain a record of the past. Such are lake sediments, where historical lake conditions can be inferred from the composition of deposited planktonic organisms (Brancelj *et al.* 2000a, 2000b, 2002). Pollen profiles in these sediments can be used to interpret watershed vegetation cover.

1.6.2.3 Extrapolating from empirical models

Often MDC sites cannot be defined or are absent from the chosen area. In such situations empirical models derived from associations between biological indicators and human-disturbance gradients can be extrapolated to infer conditions in the absence of human disturbance (Karr and Chu 1999). Even though these models can be multivariate and based on large amounts of data, the extrapolations are outside the model calibrations. Therefore, caution is advised in using these models to MDC sites.

1.6.2.4 Ambient distributions

LDC can be estimated trough various interpretations of the range of index or metric values currently observed in a region. For example, in the absence of reliable estimates of reference condition, the 5th and 25th percentile of current nutrient concentrations can be used as a criterion to separate acceptable from unacceptable values (USEPA 2000b). This approach has several issues: (1) an a priori decision about what proportion of the population is to be in LDC; (2) it assumes that higher index scores represent better conditions, rather than different environments (e.g. smaller vs. larger streams) that require separate estimates of expectations; (3) it is dependent on the distribution of sites sampled relative to the range of the indicator.

Another approach is to use indices of biotic integrity, for example the Index of biotic integrity or IBI (Karr 1981). This variant involves the assumption that, for any stream type or size, LDC is represented by the highest species richness. This assumption is used to develop maximum species richness (MSR) lines. The MSR line can be used to set reference conditions for each sampled stream. The expectation would be equal to the maximum species richness for a site of equal catchment area. The main concern is when the MSR line describes the condition of sites that are not in LDC. One of the possible explanations of this pattern is the intermediate-disturbance hypothesis (Connel 1978).

1.6.2.5 Best professional judgment

This method of estimating reference conditions is based on experience and knowledge of aquatic biologists. Even though quantification is in most cases not possible, professional judgment can provide additional insights into MDC and LDC. In order to avoid best personal judgment to become an educated guess, several criteria have to be considered: (1) justification in ecological theory; (2) comparable experience of others; (3) documentation of the "rules" by which experts develop MDC or LDC; (4) a description of how the expert reached the conclusions. By the development and inclusion of fuzzy set logic, best professional judgment has gained in importance and has been included in the development of LDC sites and multimetric indices (USEPA 2000c; Siligardi *et al.*2010).

In general it is advisable to integrate several approaches for estimating reference conditions. This may lead to firmer and scientifically more justifiable conclusions about MDC and LDC sites, especially if conclusions derived from different approaches are consistent.

1.7 Aim of this study

Our main goal was to develop a new biotic index for surface water quality assessment. This new index was to be based on hyporheic metazoan fauna, where copepods are the dominant group. In this regard, our main working hypothesis was that the hyporheic copepod community alone can be used for accurate water quality assessment. Individual species can also act as bioindicators, indicating a certain water quality status (good or bad). A water quality assessment method which would take into account the hyporheic community would be advantageous over other freshwater fauna based indices, as it could detect short term and low concentration pollution.

Even though we have become aware of this ecotones importance for surface and groundwater, our knowledge of the community is limited. The ground basis for developing such an index is knowledge of the biogeographical distribution and diversity of the animals under study. In Slovenia, there were only a few studies on hyporheic communities, all of them limited to just one or two rivers. A comprehensive study, which would encompass all major rivers, is therefore lacking. We remedied this by performing a survey of several rivers of all river beds in Slovenia, thus acquiring data on species composition and biogeography of the hyporheic community. We hypothesized that the hyporheic copepod community would conform to an upstream – downstream gradient (highest diversities would be achieved in the middle part of rivers), with deviations being the consequence of human activity.

Up to date, few ecoremediation measures took the surface water – groundwater ecotone into consideration. By developing this index and promoting its use in water quality assessment and environmental impact studies, we intend to expand currently applied methods. Evaluation of the hyporheic zone could be implemented into national water quality assessment methods. O'Connor and Dewling (1986) laid down five general guidelines for suitable indices to evaluate ecosystem degradation. An index should be:

- Relevant,
- Simple and easily understood by laymen,
- Scientifically justifiable,
- Quantitative,
- Acceptable in terms of cost.

From a more scientific point of view, the following characteristics define a good ecological indicator (Jørgensen *et al.* 2005):

- Ease of handling,
- Sensibility to small variations of environmental stress,
- Independence of reference states,
- Applicability in extensive geographical areas and in the greatest possible number of communities or ecological environments,
- Possible quantification.

Fulfilling all of these requirements is probably impossible or rather highly unlikely. To the best of our knowledge, no index has yet been developed, which would adhere to all these criteria. These deficiencies of biological indices are usually overcome by using different methods. Our index also does not fulfill all of the mentioned criteria. It is meant to be used together with other indices and thus give us a better picture of the particular ecosystem under study.

1.8 Study area

Slovenia has five major river basins (Jadran, Soča, Sava, Drava and Mura) (Figure 7), belonging to two catchments: Adriatic (Jadran and Soča) and Black Sea (Sava, Drava and Mura). Jadran is the coastal area, but is considered a river basin, since the rivers in this region flow directly into the Adriatic Sea. We selected 26 larger rivers (discharge > $1m^3s^{-1}$) of all five river basins. In addition to large enough discharge, the rivers had to have gravel beds of adequate size (surface area at least 20 m^2). The gravel beds themselves had to be easily accessible over land and 5 to 8 km apart. Using national digital orto photographs (supplied by the Slovenian environmental agency), we selected which gravel beds to sample. Final selection was done in the field, as some gravel beds have been shifted from the time when the aerial photographs were taken. Several rivers in the south-central, karst-dominated part of Slovenia (Sava Basin) did not pass these criteria.

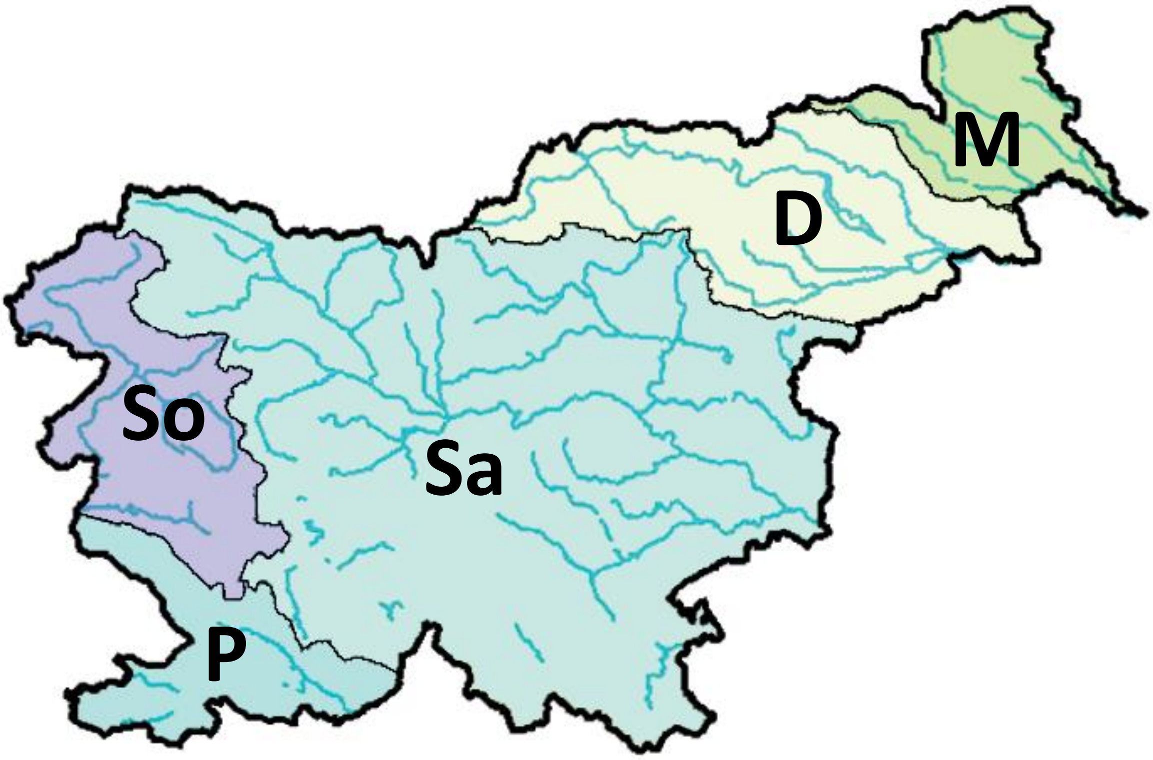


Figure 7*: Map of Slovenia with the five basins. M – Mura, D – Drava, Sa – Sava, So – Soča, P – Primorska. Blue lines – larger rivers and streams.*

2 MATERIAL AND METHODS

2.1 Sampling

Altogether, we sampled 92 gravel beds and obtained 273 samples (Figure 8). Sampling was performed within a period of 30 days (starting on 20th August 2008), during a drought. Hence, discharge in all rivers was comparatively low. Each gravel bed was divided into three equally large sections: upper (upstream), middle and lower (downstream). In each section, approximately one meter inland from the water line, the hyporheic zone was sampled according to the Karaman-Chappuis method (Delamare Debouttville 1954). The hyporheic zone of each gravel bed was sampled three times, on the same day. For each Karaman-Chappuis sample, we dug a pit approximately 20 cm below the water line and 20 cm wide (Figure 9). Exact depth and width of each pit were measured. From each pit we took a 10 liter sample (a mixture of fine organic and inorganic sediment, and interstitial fauna) which we filtered through a 60 µm net. The samples were immediately stored in 70 % alcohol. Discharge was measured as the time needed for filling a 10 liter bucket (we used a ½ l container for emptying the pits) and used as an estimate of hydraulic conductivity (HCond in s). After obtaining biological samples, temperature (T in °C), oxygen concentration (O_2 in mgl^{-1}) and saturation (Sat in %), and electric conductivity (Cond in μScm^{-1}) were measured in each pit. In addition, we also took samples (one for each gravel bed) for chemical analyses (see chapter 2.2 Laboratory analyses). For this purpose, a 250 ml bottle (for total nitrogen, total phosphorous, and alkalinity) was filled with water from the pit. In addition, 50 ml of water were filtered through a 10 µm sieve (for chemical composition of the water). Samples for chemical analyses were stored in the dark at 5 °C.

In each pit gravel size distribution, divided into 4 groups (1 cm – 3 cm, 3 cm – 6 cm, 6 cm – 10 cm, > 10 cm), was estimated. Furthermore, the amount of sediment in a 10 l sample (in ml) was measured, and sand / stone ration (in %) calculated.

At each sampling site altitude (as height above sea level – masl) was measured using a GPS device (Garmin XC60). The distance from the spring (river_km in kilometers) was measured subsequently from publicly accessible online maps (www.geopedia.si). The same source also supplied data on geological characteristics for each location. These were divided into six classes: clastites, gravel deposits, clay-gravel deposits, tertiary sediments, tonalite, and carbonates.

Parallel to the hyporheic, the benthic community of the main river channel was also sampled. Since these samples were used for determining surface water quality, we worked according to the guidelines for assessing the quality of surface rivers in Slovenia (MOP 2009b). For each gravel bed we took one sample of the bethic community. One such sample consisted of several smaller subsamples, obtained by kick-sampling into a 0,25 x 0,25 m net with a mesh size of 500 µm. All subsamples were gathered along a 100 m stretch, encompassing as many microhabitats as possible. The samples were stored in 70 % alcohol. Samples for chemical water quality assessment were taken according to the same guidelines (MOP 2009b). Altogether, we collected 184 surface water samples (92 for benthic community and 92 for water chemistry).

Figure 8: *Map of sampling sites. Each red point denotes one sampling site. The names of sampling sites are given next to the red dots. River basins are color-coded the same way as in Figure 7.*

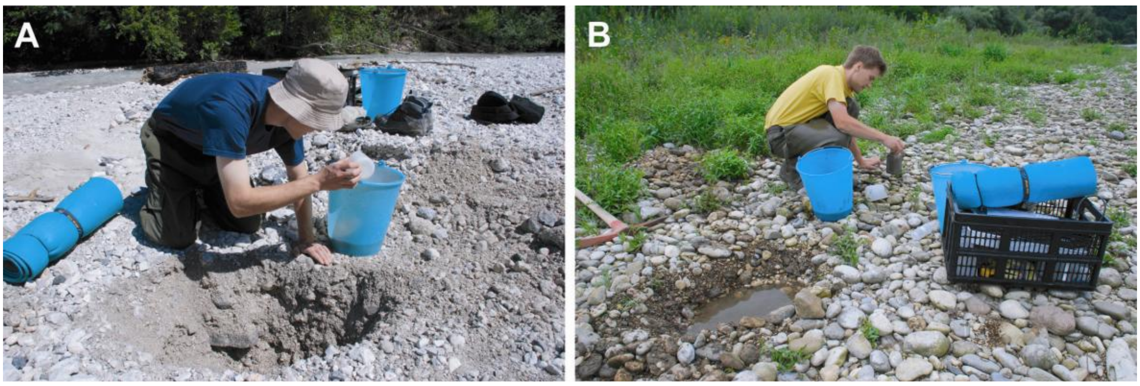


Figure 9: *Photos of sampling. A: Emptying of pit into a 10 l bucket. B: Filtering of biological samples (photo: U. Žibrat).*

2.2 Laboratory analyses

We measured the volume of each biological sample from the hyporheic. The samples were then thoroughly mixed and subsampled into 1ml units. The number of subsamples varied, but total volume of subsamples had to amount to 10 % of sample volume. Animals from all samples were determined to order or family level. Copepods were divided into Cyclopoida, Harpacticoida and Calanoida. Adult harpacticoids were determined to species level. Juvenile copepods were counted and later added to adults, according to relative abundances of each species. Absolute abundances were then calculated as number of individuals per 10 liters.

Benthic samples were analyzed in accordance with the national guidelines for surface water quality assessment (MOP 2009a). Each sample was emptied into a quadrangular dish, which we separated into four equal parts. Once the suspended material in the water settled, one of the four squares was selected at random and all animals from this particular square removed. These animals were then determined to the taxonomic level required by the national guidelines. Water quality values for each location were then calculated based in these findings.

In the laboratory the following parameters were measured: NO_2^-, NO_3^-, SO_4^{2-}, PO_4^{3-}, K^+, Cl^-, NH_4^+, Na^+, Ca^{2+}, Mg^{2+} (all in mgl^{-1}), alkalinity (Alka in $mmoll^{-1}$), total phosphorous (TP in μgl^{-1}) and total nitrogen (TN in mgl^{-1}). We used ion chromatography (Metrohm, 761 Compact IC) to determine the chemical composition of the water (anions and cations). A Perkin Elmer-Lambda 25 spectrophotometer was used to determine total nitrogen and total phosphorous. Alkalinity was measured with a WTW pH 540 GLP laboratory meter. Data on biological oxygen demand (BOD in mgl^{-1}), chemical oxygen demand (COD in mgl^{-1}) and total organic carbon (TOC in mgl^{-1}) was obtained from the Slovenian environment agency (ARSO 2008).

2.3 Statistical analyses

2.3.1 Environmental variables

Environmental data (both physical and chemical) were normalized using the Box-Cox transformation (Osborne 2010). Altogether we considered 37 environmental variables, divided into three groups: water chemistry (WC; all chemical data including temperature), physical habitat (PH; 4 classes of gravel size distribution, amount of sediment and sand, hydraulic conductivity), and geological characteristics and geography (GG; all six classes of geological characteristics, distance from spring, altitude and river membership) (Appendix A). Since the variables were on different scales (e.g. chemical parameters in mg L^{-1}, and conductivity in $\mu S\ cm^{-1}$) they were standardized to a range of 0 – 1, using the formula (Legendre and Legendre 1998):

$$\frac{Xi - \min(X)}{\max(X) - \min(X)} \qquad (1)$$

where Xi is the individual value, min(X) the minimum and max(X) the maximum value in the entire dataset for each environmental variable.

By using this standardization procedure all variables have the same range (the effect of magnitude is nullified), while retaining differences in means and variance. Data on geological characteristics and sediment pore size was fuzzy coded to the same range (0 – 1), and dummy coding (presence – absence) was used for assigning membership to each location. If a particular variable was not present at a location, the value 0 was assigned. When only one variable was present at a given location, it was given the value 1. When there were two or more variables present at a given location, the number 1 was divided by the amount of present variables and each of these assigned the same fraction. Depending on the material, a certain location could have only one variable present (the others would have values 0) or several, as long as the sum of all present variables is 1 (Table 2) (Appendix B). By using this technique, all environmental variables could be used together. Thus, normalized and standardized variables were used in principal component analysis and neighbor joining cluster analysis. For the latter Mahalanobis distances were used as a distance measure.

Table 2: *Example of fuzzy coded dummy variables. Subset 1 consists of 5 variables (E1 to E5), Subset 2 of 4 (G1 to G4). Each of these subsets can be used separately for analyses. The sum of values for each location for each subset is always 1.*

Location	Subset 1					Subset 2			
	E1	E2	E3	E4	E5	G1	G2	G3	G4
1	0	0	0	0.5	0.5	0	0	1	0
2	0.33	0.33	0	0.33	0	0.5	0	0	0.5
3	0.2	0.2	0.2	0.2	0.2	0.33	0.33	0	0.33
4	0.25	0.25	0	0.25	0.25	1	0	0	0

2.3.2 River typology and defining reference states

The Slovene river typology is based on ecoregions, bioregions, average yearly discharge, catchment size and spring influence (MOP 2008). Since this typology is to be used in all water quality assessment schemes, we implemented it into our new index.

Four approaches for defining reference states were used (reference-site approach, ambient distributions, predictive modeling and expert judgment). The national guidelines for surface water quality assessment specify reference values of nitrate and biological oxygen demand for each river type (MOP 2009c). But the national guidelines values are based on data from surface water only and therefore not to be used directly as references for interstitial waters. For that reason, we calculated average values from all sites for each water body type, for surface and hyporheic water. Average values from hyporheic waters were divided by average values from surface water. Water-body type-specific reference values were then multiplied by this number (Appendix C). For most water body types nitrate concentrations in the hyporheic were higher than in surface waters. Hence, after the transformation, reference conditions were also higher.

Once reference values were determined for each river type (one for good status, one for bad), they were used together with nitrate and biological oxygen demand data in a principal component analysis (PCA). In the PCA plot, the value 1 was assigned to the reference denoting a high status, and the value 0 to the “bad” reference point. Thus, samples close to 1 would be deemed to have a high status and those close to 0, a bad one. Based on the guidelines for establishing reference conditions for inland surface waters of the EU (REFCOND 2003) the range values for the chemical status classification were:

- High: 1 – 0,83;
- Good: 0,82 – 0,62;
- Moderate: 0,61 – 0,41;
- Poor: 0,40 – 0,20;
- Bad: < 0,20.

In the theoretical example shown in Figure 10, the high status reference is in the negative part of component 1 and positive on component 2. It is thus characterized by low nitrogen and BOD values. On the other end of the spectrum, the “bad” reference point is characterized by higher nitrogen concentration and BOD. The sample symbols can be projected perpendicularly onto the line connecting both reference points. Samples are thus located within one of the aforementioned classes. The projection of sampling stations out of range between high and bad reference values, were assigned to “high” and “bad” status, respectively.

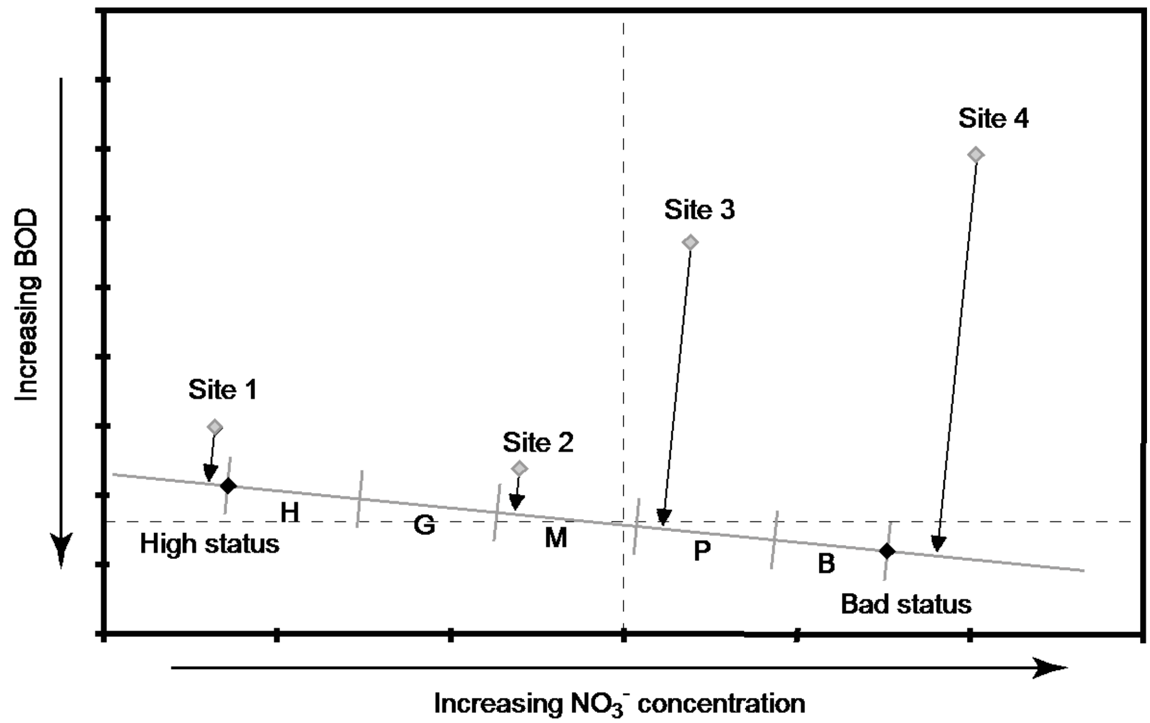


***Figure 10**: Scheme for status determination. The possible statuses are defined as: H – high, G – good, M – moderate, P – poor, B – bad. Grey diamonds: theoretical sampling sites; black diamonds: high and bad end of the quality spectrum.*

2.3.2 Copepod biodiversity calculation

A common approach to estimating species richness is the calculation of species accumulation curves (Soberón and Llorente 1993, Ugland *et al.* 2003, Colwell *et al.* 2004). For a better estimate several methods for calculating accumulation curves were used: Mao Tau (incidence based rarefaction curve), Chao1 (based on the number of rare species in a sample), Chao 2 (an instance based version of Chao1), ACE (abundance based coverage estimator) and ICE (incidence based coverage estimator) (Chao 1984, 1987; Chazdon *et al.* 1998, Chao *et al.* 2000, Colwell *et al.* 2004).

Six diversity indices and one modified index were calculated to assess copepod biodiversity (Table 3). The accuracy of certain indices was checked by plotting their agglomerative values against the number of samples. In addition, a rank/abundance plot was used to assess whether a log series distribution or log normal distribution is a good descriptor of underlying species abundance patterns. The differences in diversity indices among river basins were determined by generating a 1000 random pairs (sampled from each river basing were pooled), and diversity indices calculated for each pair. From this data, probabilities that the observed differences could have occurred by random sampling were calculated.

To achieve a better picture of the spatial distribution of the calculated indices a gridding method of spatial interpolation was used. This method interpolates scattered 2D data points onto a regular grid. We used a jackknifed inverse distance weighting algorithm and a 100x100 cells map (Davis 2002).

***Table 3**: List of the calculated diversity indices with according equations. The legend is united for all indices. Some indices use the same parameters.After Magurran (2004).*

Index name	Equation	Legend
Margalef's richness	$Mg = \frac{S-1}{\ln(n)}$	Mg – Margalef's richness S – number of taxa n – number of individuals
Fisher's alpha	$S = \alpha \ln(1 + \frac{n}{\alpha})$	α – Fisher's alpha
Shannon-Wiener diversity	$H' = -\sum p_i \, ln p_i$	H' – Shannon diversity p_i – relative abundance of species i
Modified Shannon-Wiener	$H^{1'} = e^{H'}$	$H^{1'}$ – Modified H'
Simpson's diversity	$D = 1 - \sum \frac{n_i(n_i-1)}{N(N-1)}$	D – Simpson's diversity n_i – nr. of individuals of the i-th species N – total nr. of individuals
Simpson's evenness	$E_{1/D} = \frac{1/D}{S}$	$E_{1/D}$ – Simpson's evenness
Berger-Parker dominance	$d = \frac{N_{max}}{N}$	d – Berger-Parker dominance N_{max} – nr. of individuals in most abundant species

To evaluate the contribution of each species to the differences among river basins and other sample clusters, we used the similarity percentage analysis (SIMPER). This method calculates the average Bray-Curtis dissimilarity between all pairs of sample groups. Samples can be grouped in accordance with the study's requirements (e.g. according to habitat, river type or altitudinal range). The species which consistently contribute greatly to the average similarity between sites are considered characteristic of the habitat (Clarke 1993).

2.3.3 Species-environment relations

Distance based redundancy analysis (db-RDA) (Legendre and Anderson 1999) to search for patterns in species-environment relations. As the distance measure we calculated the zero-adjusted Bray-Curtis dissimilarity (za-BC) on 4th-root transformed species abundance data (Clarke *et al.* 2006). The za-BC was calculated using the following equation:

$$za - BC\ (i,j) = \frac{\sum_{k=0}^{n-1} |y_{ik} - y_{ik}|}{2 + \sum_{k=0}^{n-1} (y_{ik} + y_{ik})} \qquad (2)$$

where za-BC is calculated as the distance between locations *i* and *j*, *k* is the index of a species and *n* is the total number of species *y.*

The za-BC dissimilarity matrix was square-root transformed and then used as input data in a principal coordinates analysis. The derived components were then used in the db-RDA. To present the spatial distribution of za-BC dissimilarities a non-metric multidimensional scaling (NMDS) procedure was used.

A variance decomposition procedure of three sets of explanatory variables was also performed. These three sets were: WC (including temperature), PH (amount of sediment and gravel size distribution) and GG (geological characteristics and river membership for each location).

Statistically significant differences among clusters were evaluated with non-parametric Manova (also referred to as Permanova) (Anderson 2001, McArdle and Anderson 2001). As input data we used the za-BC dissimilarity.

2.3.4 Machine learning methods

"Essentially all models are wrong, but some are useful."
George E.P. Box

The 26 rivers, included in this study, were assigned into 13 river types, according to national guidelines (MOP 2008, Appendix E). These types were used in all data mining procedures. In order to reduce effects of rare and less abundant species a principal component analysis on 4th root transformed species abundance data was performed. Two diversity estimates, the Simpson's index (1-D) and Simpson's evenness ($E_{1/D}$) were also included. Each model was developed using 92 instances (each sampling location was regarded as an instance). For model validation an additional dataset from 8 sampling locations distributed along the entire length of the river Kamniška Bistrica was used. These samples were gathered using the same technique as samples for model development and surface water quality determination according to WFD guidelines. A total of 32 additional samples (24 hyporheic and 8 benthic community) were gathered.

For the development of the final model three widely used algorithms were chosen: Regression and classification trees M5P and J48 (both based on the algorithm C4.5), and the neural network algorithm called multilayer perceptron (Quinlan 1992, Wang and Witten 1997).

Regression trees predict the value of a dependent variable (referred to as class) from the set of values of independent variables (called attributes). Data are organized as an attribute-by-location matrix (species or environmental variables can be attributes). One of the columns in the data matrix has to be the dependent variable. In regression trees the dependant variable has to be in integer form, whereas in classification trees a division into classes is obligatory. Tree construction starts with the entire data table. At each step, the most discriminating attribute is selected as the root of the subtree and the current data set is split into subsets according to the values of the selected attribute. If the attribute is discrete (divided into classes), a branch of the tree is created for each possible value. With continuous attributes, a threshold is created and the tree is split into two branches based on that threshold. The most discriminating attribute is the one that reduces most the variance of the dependant variable. Tree construction stops when the variance of the dependant variable values of all examples in a node (i.e. tree branch) is small enough. These final nodes are called leaves and represent a model (either a constant for classification trees or linear equation for regression trees), predicting the value of the dependant variable.

Artificial neural networks (ANN) are mathematical models inspired by the structure and functional aspects of biological neural networks. Mostly, neural networks are adaptive systems; they change their structure based on information that flows through the network during the learning phase (Lawrence 1994). A multilayer perceptron is an example of a widely used artificial neural network. Generally, ANNs are input/output devices with the neurons organized into layers. Simple perceptrons consist of a layer of input neurons, a single layer of weights and a layer of output neurons (Figure 11). The learning process consists of finding the correct values for the weights between the input and output layers. The data is presented at the input layer; the network processes the input by multiplying it with the weights. The results are interpreted by the output nodes, using a function to determine whether the output nodes fire (in analogy to excited biological neurons). During the training phase, the difference between the guess made by the network and the correct value for the output is assessed and the weights changed, in order to minimize the error. The main benefit of multilayer perceptrons is that they can distinguish data that is not linearly separable.

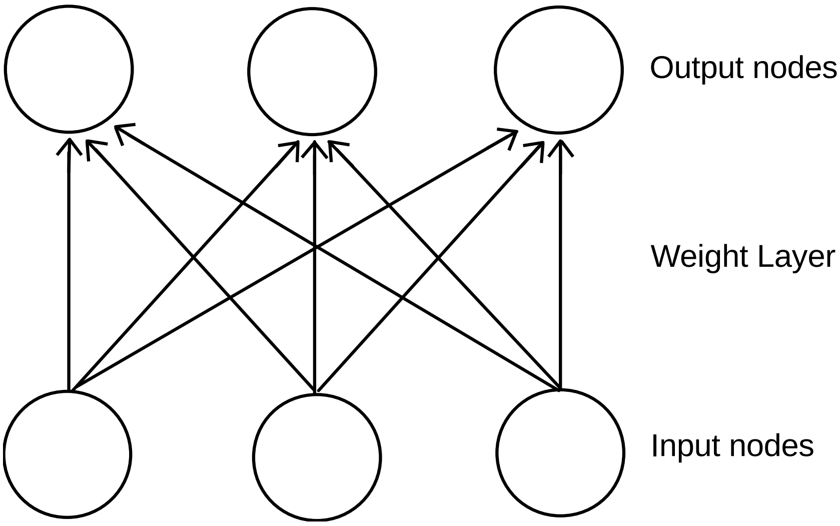


***Figure 11**: Structure of a simple perceptron.*

For general statistics, diversity and cluster analyses, we used the PAST program (Hammer *et al.* 2001), while Canoco was used for other multivariate analyses (dbRDA and models based on this analysis). Regression and classification tree models (based on algorithms M5' and J48, respectively) and neural network models (multilayer perceptron) (Quinlan 1992, Wang and Witten 1997) were developed in the Waikato Environment for Knowledge Analysis or WEKA (Witten and Frank 2005). All models were validated using 10-fold cross validation.

3 RESULTS

3.1 Environmental parameters

Most of the measured variables were significantly correlated with distance from the spring (Table 4; Appendix D). Oxygen content and saturation, and potassium concentrations are the only chemical variables negatively correlated with distance from spring. Only nitrate and ammonia concentrations, hydraulic conductivity, and biological oxygen demand were correlated with less than five variables. This suggests that a strong gradient, represented by the distance from the spring, can be expected. Significant correlations among the other variables (e.g. chloride and sodium, potassium, oxygen content and conductivity) reflect this main gradient. The four less often correlated variables (BOD, Hcond, NH_4^+ and NO_3^-) can then influence the deviation of sampling sites from this gradient.

To achieve a better picture of the spatial distribution of this gradient, a principal component analysis was performed. Strong correlations among certain variables (distance from spring/headwater and altitude, Cl^-, NO_2^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+, K^+, TP, T, O_2, O_2 saturation, conductivity, sediment, COD and TOC) enabled us to reduce their number prior to the analysis, and thus avoid artificially strengthened gradients, which could obscure other patterns. Dummy and fuzzy coded variables were excluded from the analysis in a backward selection procedure as they produced only noise and obscured patterns. The first two axes of the principal component analysis explain 61,2 % of the variance inherent in environmental data (Figure 12). The first axis represents an upstream-downstream gradient and reflects results of the correlation analysis. The second axis shows the effect of the less correlated variables. Ammonium (NH_4^+), an indicator of organically polluted waters, has the highest correlation with the second axis. Biological oxygen demand and hydraulic conductivity have approximately the same influence on both components, while nitrogen prevails in the second axis (even though it is the weakest of all variables). The second axis thus represents a pollution component, independent of the location along the river. Most of the sampling sites are arranged along axis 1, upstream on the right-hand side, downstream on the left. In the upper right are locations close to the spring or headwater with higher nitrate concentrations, while in the upper left downstream locations with ammonia and higher biological oxygen demand reside.

***Table 4**: Pearson correlations between environmental variables. Dummy and fuzzy coded dummy variables are not included. Upper triangle: probabilities that the variables are uncorrelated; lower triangle: correlation values. Rkm: distance from spring/headwater (km), masl: meters above sea level (m), TN: total nitrogen (mgl^{-1}), TP: total phosphorous (μgl^{-1}), Alka: alkalinity ($mMoll^{-1}$), T: temperature (°C), sat: oxygen saturation (mgl^{-1}), sed: amount of sediment in a 10 l sample (ml), Hcond: hydraulic conductivity (s), Cond: electric conductivity (μScm^{-1}), BOD: biological oxygen demand (mgl^{-1}), COD: chemical oxygen demand (mgl^{-1}), TOC: total organic carbon (mgl^{-1}).*

	rkm	masl	Cl	NO2	NO3	SO4	NH4	Ca	Mg	Na	K	TN	TP	Alka	Ca/Mg	T	O2	sat	cond	sed	Hcond	BOD	COD	TOC
rkm		0.00	0.00	0.00	0.46	0.00	0.12	0.02	0.02	0.00	0.00	0.15	0.00	0.11	0.99	0.00	0.00	0.00	0.00	0.05	0.39	0.07	0.00	0.00
masl	-0.49		0.00	0.07	0.69	0.02	0.76	0.00	0.09	0.00	0.00	0.14	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.12	0.11	0.44	0.03	0.00
Cl	0.57	-0.46		0.00	0.29	0.00	0.88	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.46	0.00	0.00	0.00	0.00	0.00	0.26	0.02	0.00	0.00
NO2	0.37	-0.19	0.43		0.04	0.03	0.29	0.24	0.78	0.01	0.00	0.01	0.00	0.87	0.33	0.00	0.00	0.00	0.05	0.22	0.81	0.12	0.04	0.01
NO3	0.08	0.04	-0.11	-0.21		0.14	0.69	0.15	0.04	0.51	0.92	0.00	0.76	0.12	0.12	0.88	0.25	0.28	0.01	0.51	0.14	0.99	0.13	0.08
SO4	0.36	-0.25	0.57	0.23	-0.15		0.08	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.05	0.08	0.00	0.00	0.00	0.01	0.83	0.53	0.01	0.00
NH4	0.16	0.03	-0.02	0.11	0.04	-0.18		0.88	0.77	0.04	0.88	0.30	0.27	0.77	0.62	0.01	0.20	0.72	0.52	0.85	0.51	0.31	0.27	0.36
Ca	0.25	-0.55	0.63	0.12	-0.15	0.33	0.02		0.00	0.00	0.00	0.01	0.07	0.00	0.03	0.00	0.00	0.00	0.00	0.02	0.96	0.19	0.74	0.01
Mg	0.24	-0.18	0.28	0.03	-0.21	0.36	0.03	0.32		0.30	0.22	0.01	0.08	0.00	0.00	0.11	0.01	0.25	0.00	0.63	0.74	0.17	0.55	0.75
Na	0.44	-0.38	0.83	0.26	-0.07	0.51	-0.21	0.58	0.11		0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.24	0.21	0.00	0.00
K	-0.53	0.40	-0.83	-0.38	0.01	-0.54	-0.02	-0.45	-0.13	-0.76		0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.34	0.07	0.00	0.00
TN	-0.15	0.16	-0.49	-0.28	0.64	-0.31	-0.11	-0.28	-0.28	-0.36	0.35		0.02	0.01	0.18	0.01	0.09	0.19	0.00	0.05	0.61	0.01	0.47	0.42
TP	0.52	-0.41	0.53	0.36	-0.03	0.25	0.12	0.19	0.18	0.44	-0.50	-0.24		0.08	0.45	0.00	0.00	0.00	0.00	0.18	0.48	0.44	0.00	0.00
Alka	0.17	-0.40	0.47	-0.02	-0.16	0.33	0.03	0.76	0.68	0.34	-0.30	-0.28	0.18		0.00	0.03	0.00	0.01	0.00	0.15	0.99	0.55	0.43	0.28
Ca/Mg	0.00	0.17	-0.08	-0.10	-0.16	0.21	-0.05	-0.23	0.81	-0.19	0.12	-0.14	0.08	0.31		0.22	0.69	0.27	0.04	0.65	0.88	0.79	0.15	0.04
T	0.77	-0.52	0.49	0.41	-0.02	0.19	0.28	0.37	0.17	0.41	-0.44	-0.27	0.41	0.22	-0.13		0.00	0.00	0.00	0.79	0.55	0.01	0.00	0.03
O2	-0.61	0.48	-0.69	-0.31	-0.12	-0.40	-0.13	-0.48	-0.25	-0.56	0.59	0.18	-0.48	-0.40	0.04	-0.57		0.00	0.00	0.02	0.16	0.35	0.00	0.00
sat	-0.51	0.38	-0.62	-0.30	-0.11	-0.30	-0.04	-0.40	-0.12	-0.51	0.56	0.14	-0.46	-0.28	0.12	-0.39	0.80		0.00	0.02	0.03	0.02	0.00	0.00
cond	0.38	-0.49	0.71	0.21	-0.28	0.51	0.07	0.80	0.64	0.53	-0.54	-0.46	0.31	0.83	0.21	0.43	-0.57	-0.43		0.00	0.86	0.11	0.57	0.01
sed	0.20	-0.16	0.36	0.13	-0.07	0.27	-0.02	0.25	0.05	0.36	-0.38	-0.20	0.14	0.15	-0.05	0.03	-0.24	-0.24	0.30		0.58	0.28	0.10	0.00
Hcond	0.09	0.17	0.12	0.03	0.15	0.02	0.07	0.00	-0.03	0.12	-0.10	-0.05	0.07	0.00	0.02	0.06	-0.15	-0.22	-0.02	0.06		0.00	0.02	0.26
BOD	0.19	0.08	0.25	0.16	0.00	0.07	0.11	0.14	0.15	0.13	-0.19	-0.26	-0.08	0.06	-0.03	0.26	-0.10	-0.24	0.17	-0.11	0.32		0.00	0.79
COD	0.49	-0.23	0.46	0.22	0.16	0.25	-0.12	0.03	-0.06	0.45	-0.48	-0.08	0.30	-0.08	-0.15	0.31	-0.43	-0.54	0.06	0.17	0.25	0.33		0.00
TOC	0.39	-0.34	0.60	0.27	0.18	0.36	-0.10	0.28	-0.03	0.56	-0.57	-0.09	0.40	0.11	-0.21	0.23	-0.60	-0.57	0.26	0.39	0.12	0.03	0.62	

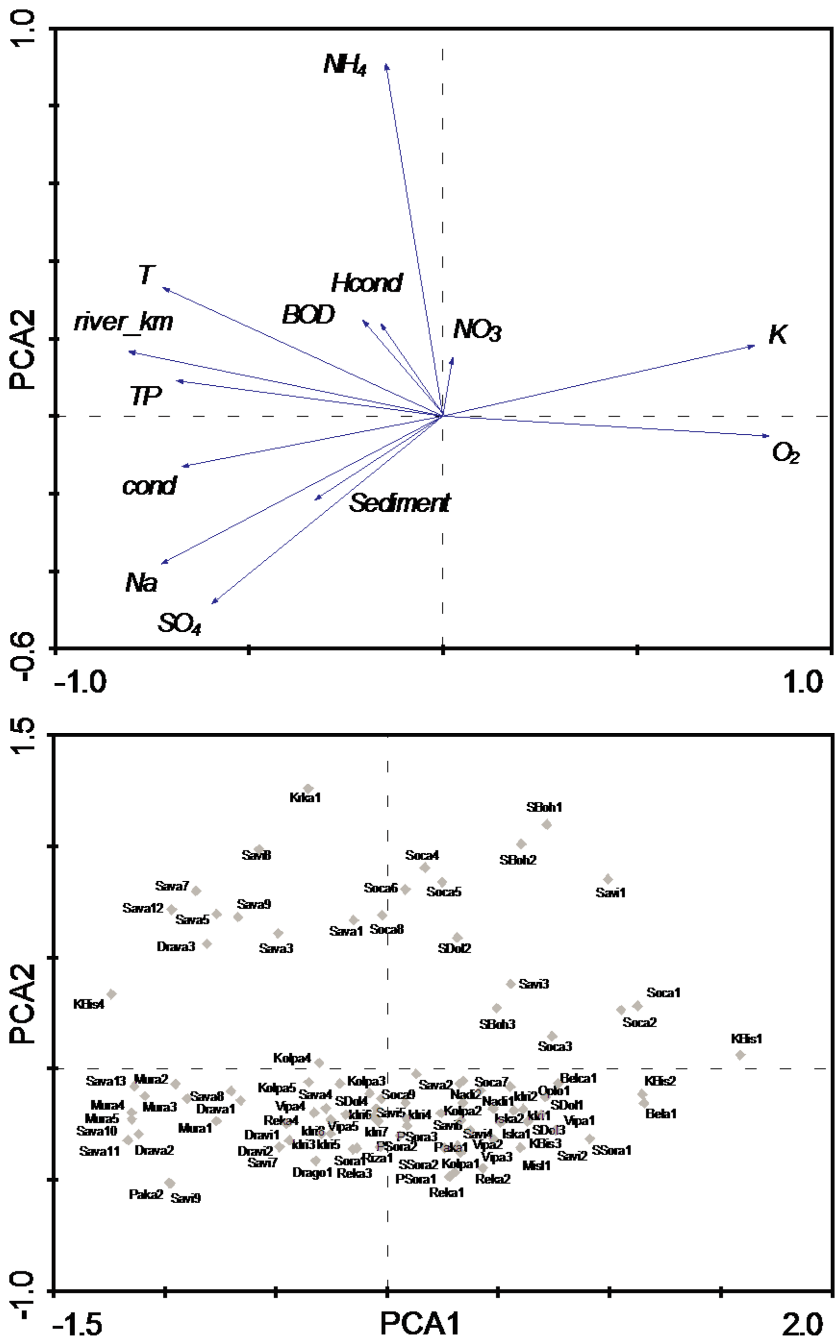


Figure 12: *Principal component analysis of environmental data. Variables, included in the analysis, were determined with a backward selection procedure.*

The same variables, used for the PCA, were then used for neighbor joining cluster analysis. The analysis produced five distinct clusters (Figure 13). A non-parametric Manova test showed statistically significant differences among all clusters (sum of squares 46,5; F: 17,82; p<0,001). Clusters 2, 3 and 4 mostly contain middle and lower reach locations. Within cluster 3 the Mura, lower part of the Sava and Drava are grouped together. The first cluster combines sampling sites close to springs of rivers, which rise in the Kamniško-Savinjske Alpe mountain region (Kamniška Bistrica, Savinja, Bela). These sampling sites are characterized by low temperatures, high oxygen content and low resource availability. The second cluster contains some of the most impacted sites in Slovenia (Paka 2, KBis4, Savi7 and Savi9; Appendix E). Higher temperatures, better resource availability, and less available oxygen are characteristic of this cluster. The amount of total phosphorous was highest in cluster 2, while ammonia reached highest concentrations in cluster 4. In cluster 3, hydraulic conductivity and the amount of sediment were highest. This cluster contains mostly low-reach locations (e.g. the river Mura and downstream locations of the Sava). The 2nd and 3rd cluster also have the lowest O_2 concentrations. Clusters 1 and 5 envelop upper and middle reach locations. In the fifth cluster, upper reaches from all rivers rising from the Pohorje mountain region are grouped together (Oplotnica, Mislinja and Paka). Characteristics of these locations are similar to those in cluster 1. The greatest differences were in the amount of total phosphorous and sediment (Table 5).

Table 5*: Average values of environmental variables for the five clusters identified in the neighbor-joining cluster analysis.*

Taxon	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Masl (m)	422	219	202	291	369
river_km (km)	16,6	38,5	123	78,1	28,8
Hcond (s)	224	169	286	209	212
Sediment (ml)	32	39,6	47,2	23,6	13
T (°C)	10,9	13,1	15,9	16,5	12,3
O_2 (mgl^{-1})	8,8	2,66	3,67	5,21	9,42
Saturation (%)	84,1	25,8	37	56,1	90,9
Conductivity (µScm^{-1})	314	489	396	362	295
TP (mgl^{-1})	1	107	26,8	24,1	14,8
TN (mgl^{-1})	1,15	3,31	1,34	1,54	1,01
Ca^+ (mgl^{-1})	58,3	86,4	71,3	62,6	49,6
NH_4^+ (mgl^{-1})	0,04	0,02	0,00	0,46	0,00
Na^+ (mgl^{-1})	2,40	11,30	6,52	2,60	1,72
Mg^+ (mgl^{-1})	11,6	13,9	13,6	10,6	14
K^+ (mgl^{-1})	0,50	2,83	1,38	0,86	0,49
Cl^- (mgl^{-1})	2,21	11,20	6,23	3,65	1,92
SO_4^{2-} (mgl^{-1})	9,62	31,6	20,2	10,5	9,16
NO_2^- (mgl^{-1})	0,004	0,04	0,02	0,02	0,02
NO_3^- (mgl^{-1})	3,15	12,20	4,57	2,76	2,80
Alkalinity (mmoll^{-1})	3,04	4,03	3,44	3,19	2,94
TOC (mgl^{-1})	1,1	1,91	1,79	1,19	0,944
BOD(mgl^{-1})	1,14	1,05	1,03	1,11	0,85
COD(mgl^{-1})	5,07	7,76	9,20	5,48	5,06

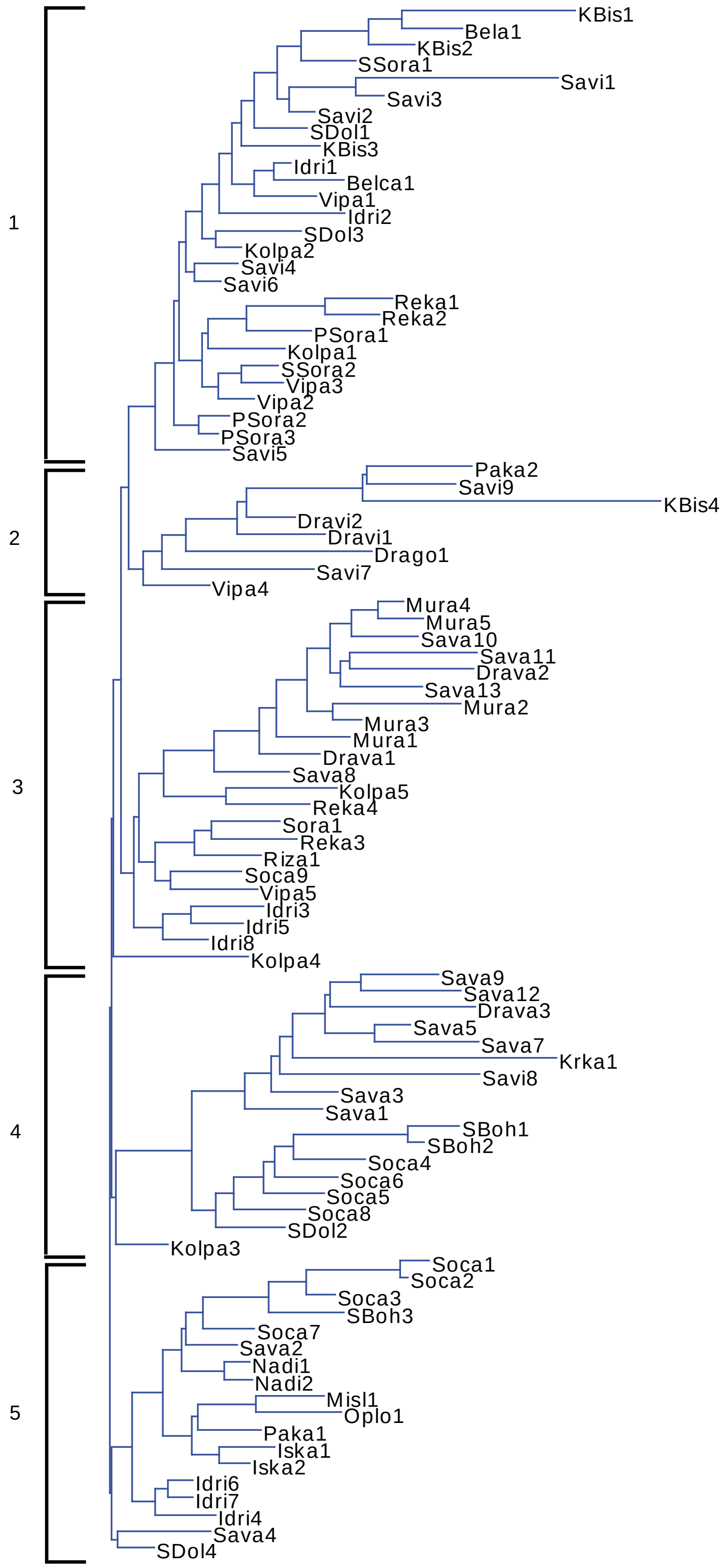


***Figure 13**: Neighbor joining cluster analysis on environmental data. The numbers on the left (1 to 5) indicate the cluster number, referred to in the text.*

3.2 Copepod biogeography and biodiversity

In total 49 copepod species, belonging to 23 genera and 6 families were found. Of these, 11 species were found in only one location (five Harpacticoida and six Cyclopoida species). Three species were found in more than 40 % of locations (*Bryocamptus dacicus, Bryocamptus zschokkei* and *Elaphoidella elaphoides*), and three more species (*Diacyclops clandestinus*, *Echinocamptus pilosus* and *Paracamptus schmeili*) were found in approximately a third of all locations. Seven species (*Bryocamptus mrazeki, B. vejdovskyi, B. hoferi, B. echinatus, B. minutus, Elaphoidella phreatica* and *P. vigueri*) were found exclusively in the rivers Soča and Idrijca (in the Soča Basin in the west part of Slovenia). *Atheyella wierzejskii* and *Moraria radovnae* were restricted to locations at higher altitudes, with high oxygen content and lower temperatures. They were found in only four locations, yet spread across the northern part of the entire sampling area. Another geographically constrained species is *Maraenobiotus insignipes* that we found only in two tributaries of the river Sava (Sava Dolinka and Sava Bohinjka) (Tables 6 and 7).

Table 6: *List of copepod species found in the hyporheic zone. Abb.: abbreviation of the species name, used in further analyses. Nr.: number of locations in which the species was found. T: temperature range (min – max). Alt.: altitudinal range (min – max) in meters above sea level. O_2: oxygen range in mg l^{-1} (min – max). Sat.: saturation range in % (min – max). Con.: conductivity range in µS cm^{-1} (min – max). Species are arranged in order of frequency of occurence.*

Species	Abb.	Nr.	Alt. [m]	T [°C]	O_2 [mg/l]	Sat. [%]	Con. [µS/cm]
Bryocamptus dacicus (Chappuis, 1923)	Bryda	43	77-780	6-20	1-6	13-114	88-707
Bryocamptus zschokkei (Schmeil, 1893)	Bryzs	41	24-780	7-19	1-15	11-134	88-707
Elaphoidella elaphoides (Chappuis, 1923)	Elael	40	24-516	8,5-23	1-14	10-129	244-825
Diacyclops clandestinus Kiefer 1933	Diacl	30	36-780	7-22	1-14	15-134	119-701
Paracamptus schmeilii (Mrázek, 1893)	Parsc	29	78-719	8-23	1-13	7-130	226-623
Echinocamptus pilosus (van Douwe, 1910)	Echpi	26	36-604	9-22	1-14	13-80	266-707
Bryocamptus pygmaeus (Sars, 1862)	Brypy	19	77-530	9-22	1-14	13-129	219-304
Nitocrella hirta (Chappuis, 1923)	Nithi	15	24-719	8-19	1-11	10-80	235-529
Athyella crassa (Sars, 1862)	Athcr	14	24-436	11-22	1-2	7-80	266-528
Acanthocyclops hispanicus Kiefer, 1937	Acahi	12	78-719	8-18	2-14	19-134	235-451
Epactophanes richardii Mrázek, 1893	Epari	12	107-697	9-22	1-15	11-134	88-528
Megacyclops viridis (Jurine 1820)	Megvi	12	150-493	10-22	2-9	19-83	289-497
Diacyclops bisetosus (Rehberg 1880)	Diabi	10	213-718	8-18	2-11	15-107	246-700
Diacyclops languidoides (Lilljeborg, 1901)	Dialg	9	90-638	9-17	2-11	19-114	220-479
Eucyclops serrulatus (Fischer 1851)	Eucse	9	77-434	10-23	2-11	19-118	215-515
Acanthocyclops kieferi (Chappuis, 1925)	Acaki	8	77-418	10-23	2-10	19-118	278-513
Canthocamptus staphylinus (Jurine, 1820)	Canst	7	150-269	13-22	1-8	13-96	288-629
Bryocamptus mrazeki (Minkiewizc, 1916)	Brymr	6	107-328	11-14	1-11	10-99	266-596
Diacyclops languidus (Sars 1863)	Diala	6	141-403	9-16	2-14	19-131	241-451
Bryocamptus minutus (Claus, 1863)	Brymi	5	223-328	10-14,5	1-8	13-75	370-472
Bryocamptus typhlops (Mrázek, 1893)	Bryty	5	263-660	9-19	8,5-14,5	96-134	276-428
Diacyclops crassicaudis (Sars 1863)	Diacr	5	24-604	9-14	5-14	43-134	276-428
Atheyella wierzejskii (Mrázek, 1893)	Athwi	4	288-780	7,5-11	5-12	50-113	119-413
Bryocamptus vejdovskyi (Mrázek, 1893)	Bryvej	4	288-436	10-14,5	1-8	14-73	322-472
Diacyclops zschokkei (Graeter, 1910)	Diazs	4	77-324	12-13	6-8,5	61-84	349-440
Moraria poppei (Mrázek, 1893)	Morpo	4	78-530	10-20	1-8	12-82	219-446

Paracylops fimbriatus (Fischer 1853)	Parfi	4	69-328	11-14	3-12	31-110	349-387
Parastenocaris nolli (Kiefer, 1938)	Parno	4	171-363	12-21	7-9,5	66-92	298-392
Bryocamptus echinatus (Mrázek, 1893)	Bryec	3	314-452	10-13	6-11,5	65-110	294-388
Maraenobiotus insignipes (Lilljeborg, 1902)	Marin	3	417-660	14-18	7-10	82-112	208-284
Mesochra aestuari (Gurney, 1921)	Mesae	3	36-229	13-16	1-11	13-114	356-533
Speocyclops infernus (Kiefer 1930)	Spein	3	328-719	6-11	5-15	43-135	170-387
Acanthocyclops venustus (Norman&Scott 1906)	Acave	2	213-406	13-16	4-6	39-61	351-423
Bryocamptus rhaeticus (Schmeil, 1893)	Bryrh	2	335-378	12-13	3-6	29-59	343-394
Elaphoidella gracilis (Sars, 1862)	Elagr	2	446-516	9-10	6-9	5-8	279-300
Moraria radovnae (Brancelj, 1988)	Morbr	2	697-719	8	9-11	83-104	88-235
Moraria varica (Graeter, 1911)	Morva	2	303-530	17-18	7-8	82-86	219-339
Phyllognathopus vigueri (Maupas, 1892)	Phyvi	2	239-287	13-14	9-10	96-97	278-296
Acanthocyclops sambugarae Kiefer, 1981	Acasa	1	406	16	4	39	423
Acanthocyclops sensitivus (Graeter&Chappuis, 1914)	Acase	1	304	14	2	19	497
Bryocamptus cuspidatus (Schmeil, 1893)	Brycu	1	436	11	8	84	316
Bryocamptus hoferi (Douwe, 1907)	Bryho	1	436	11	8	74	316
Elaphoidella phreatica (Chappuis, 1925)	Elaph	1	304	14	2	25	450
Graeteriella unisetigera (Graeter 1908)	Graun	1	406	16	8	39	423
Macrocyclops albidus (Jurine 1820)	Macal	1	213	14	2	17	533
Nitocra psammophila (Noodt, 1952)	Nitps	1	210	18	1	11	512
Paracyclops affinis (Sars 1863)	Paraf	1	213	13,5	2	17	533
Parastenocaris gertrudae (Kiefer, 1968)	Parge	1	697	8,5	9	8	88
Tropocyclops prasinus (Fischer 1860)	Tropr	1	275	11	2	17	282

With pooled data from the river basins, all five species richness estimators draw near an asymptote after approximately a 100 samples. A comparison between Chao1 with Chao2 and ACE with ICE indicates that the samples are homogenous (Figure 14).

Table 7: *Occurences of species in each river basin. Species abbreviations are in Table 2. The presence of a species is marked with an X.*

River basin / Species	Acahi	Acaki	Acasa	Acase	Acave	Athcr	Athwi	Brycu	Bryda	Bryec	Bryho	Brymi	Brymr	Brypy	Bryrh	Bryty	Bryvej	Bryzs	Canst	Diabi	Diacl	Diacr	Dialg	Diala	Diazs	Echpi	Elael	Elagr	Elaph	Epari	Eucse	Graun	Macal	Marin	Morpo	Morra	Morva	Megvi	Mesae	Nithi	Nitps	Parsc	Paraf	Parfi	Parge	Parno	Phyvi	Spein	Tropr
Sava	X	X	X		X	X	X		X	X				X	X	X		X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X	
Drava									X					X				X	X	X	X						X															X			X				X
Mura																											X																						
Soča	X	X		X		X	X	X	X		X	X	X	X			X	X			X	X	X		X	X	X		X	X	X				X			X	X	X		X		X		X	X	X	
Primorje									X	X								X			X					X	X	X			X				X			X											

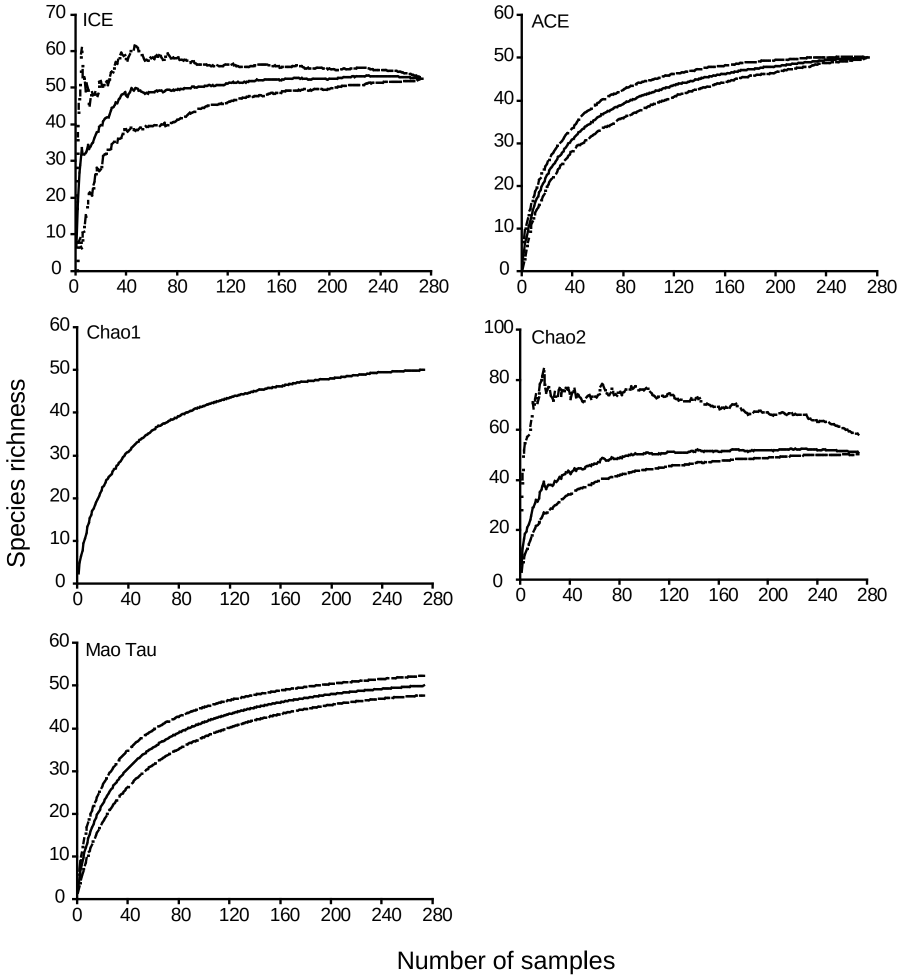


Figure 14: *Species richness estimators. Data were pooled from all river basins. The estimated species richness (averages of 50 runs) is shown as a solid line, confidence limits (Mao Tau and Chao2) and standard deviations (ACE and ICE) are shown with a dashed line.*

Species accumulation curves calculated for each basin separately show a different pattern. For the Sava basin, all five estimates approach an asymptote, suggesting the presence of approximately 40 species (Figure 15). Both incidence-based estimators reach an asymptote at 30 species in the Soča basin, while the other three only just begin leveling out. The differences between Chao1/Chao2 and ACE/ICE suggest more heterogenous samples (Figure 16). The picture becomes more apparent in the Drava and Primorje basins. Only ICE and Chao2 begin approaching an asymptote, while ACE and Chao1 keep rising steeply. The differences between incidence and abundance-based estimators again indicate heterogenous samples (Figure 17; Figure 18). The Mura river basin presents a special case, where all five estimators fail. They all reach an asymptote, estimating 1 species. Yet the variance in all cases is zero and all estimators exhibit exactly the same shape (Figure 19).

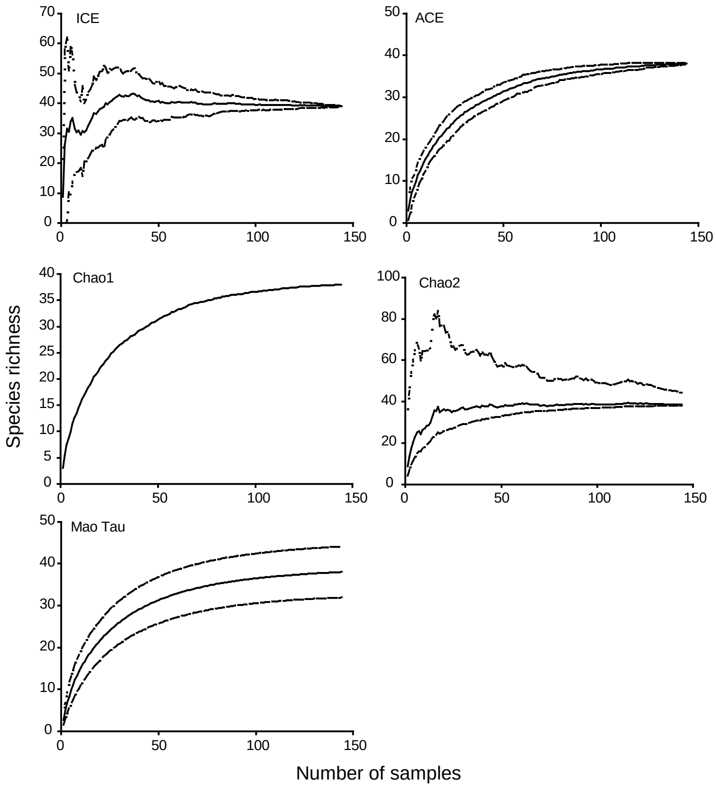


Figure 15: *Species accumulation curves for the Sava basin. Solid lines denote means of 50 runs; dashed lines signify confidence intervals (Mao Tau and Chao2) and standard deviations (ACE and ICE).*

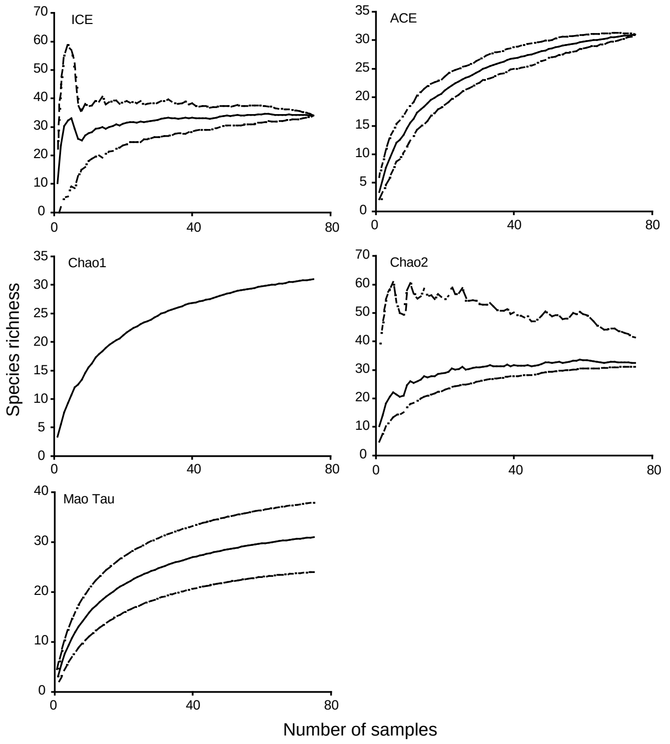


Figure 16: *Species accumulation curves for the Soča basin. Solid lines denote means of 50 runs; dashed lines signify confidence intervals (Mao Tau and Chao2) and standard deviations (ACE and ICE).*

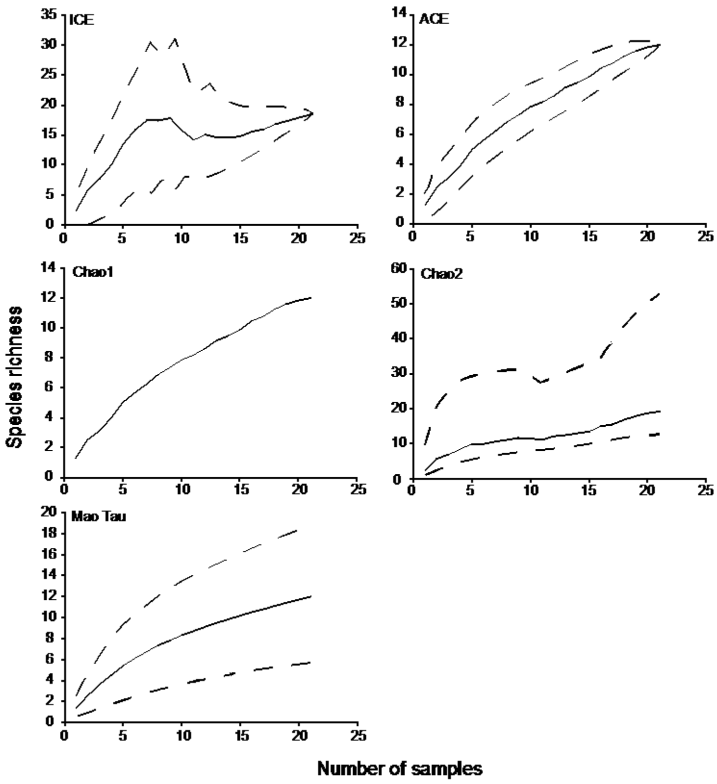


Figure 17: *Species accumulation curves for the Drava basin. Solid lines denote means of 50 runs; dashed lines signify confidence intervals (Mao Tau and Chao2) and standard deviations (ACE and ICE).*

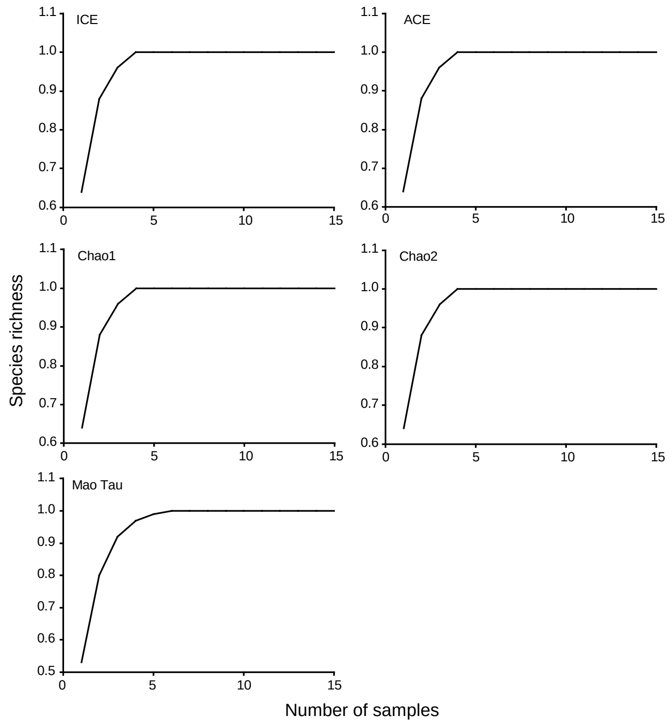


Figure 18: *Species accumulation curves for the Mura basin. Solid lines denote means of 50 runs; dashed lines signify confidence intervals (Mao Tau and Chao2) and standard deviations (ACE and ICE).*

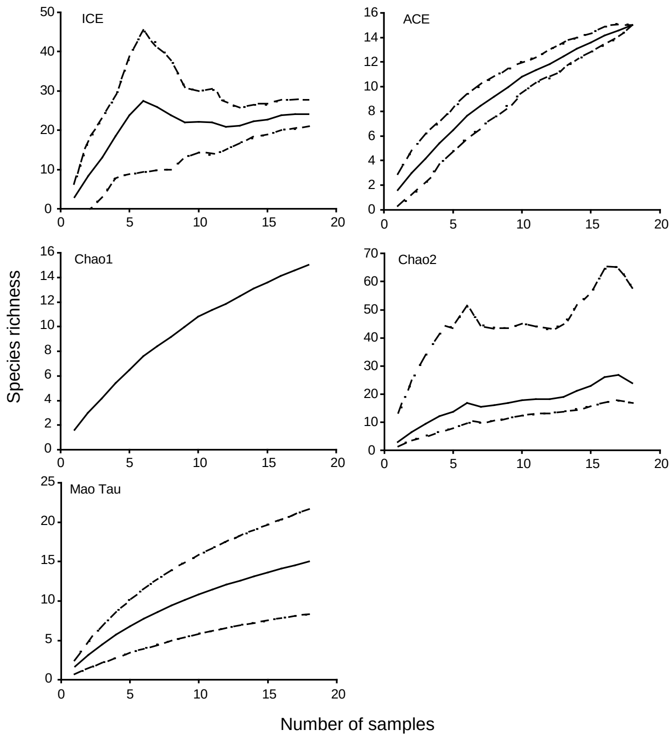


Figure 19*: Species accumulation curves for the Primorje basin. Solid lines denote means of 50 runs; dashed lines signify confidence intervals (Mao Tau and Chao2) and standard deviations (ACE and ICE).*

Four of the chosen diversity indices were tested, whether they encapsulate the diversity of the assemblage, as measured by each index. Their agglomerative values were plotted against the number of samples (similar in vein as species accumulation curves). Fisher's alpha, H' and $e^{H'}$ reach an asymptote, thus indicating that their measure of diversity is reliable. All three indices started approaching the asymptote after approximately 100 samples. The Simpson index behaved differently. After the initial rise, it levels of at 40 samples and begins dropping off after 60. After 130 samples it evens out again at values of (1-D)≈0 (Figure 20). The applicability of Fisher's alpha was further tested by fitting log normal and log series distributions, and testing their respective fit with our observed data (Figure 21; Table 8).

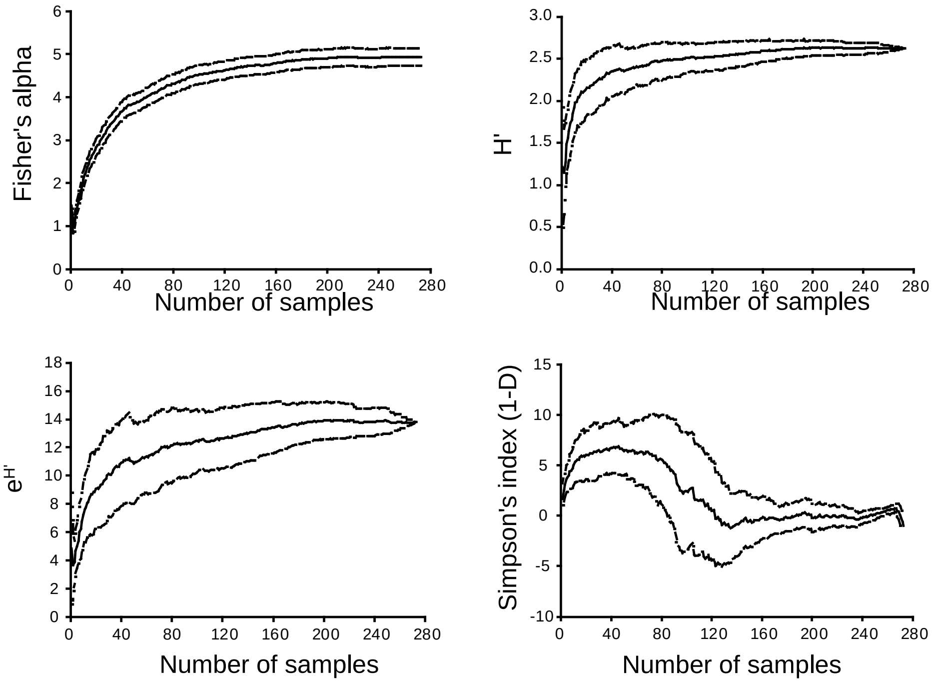


Figure 20*: The reliability of four diversity measures. The solid line represents the average of 50 runs of the analysis. Dashed lines denote standard deviations.*

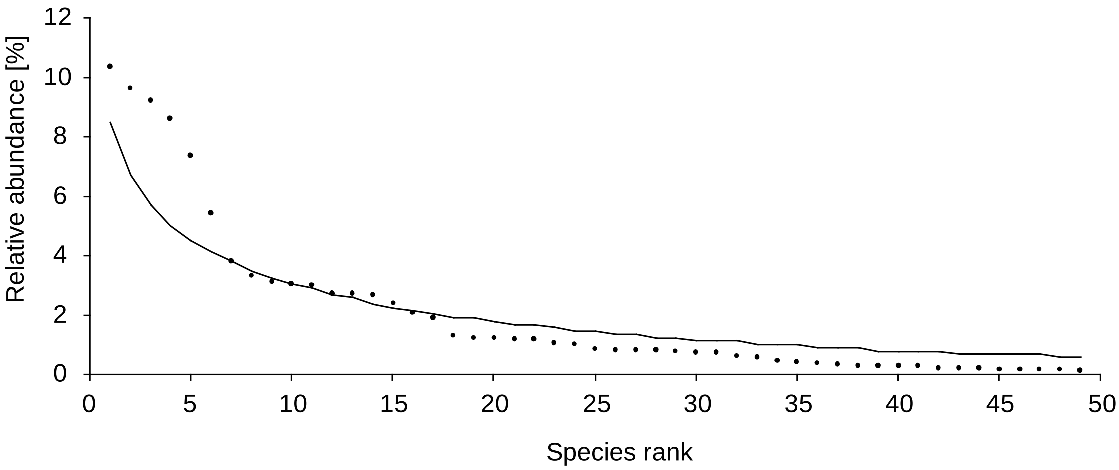


Figure 21*: Rank-abundance plot. Dots represent species' relative abundances; the solid line denotes a log series model.*

***Table 8**: Results of log series and log normal distribution model fitting.*

	Chi^2	p(same)	α/mean	x/variance
Log series	8,74	0,89	37,97	0,72
Log normal	6,58	0,04	0,46	0,11

There were significant differences in biodiversity indices only between the Mura and the others basins (Figure 22, Table 9). Differences in species richness and evenness of their distribution among the Sava, Drava, Soča and Primorje basins were not significant. Since in the Mura only one species was found, the calculations for this basin are of exploratory value only.

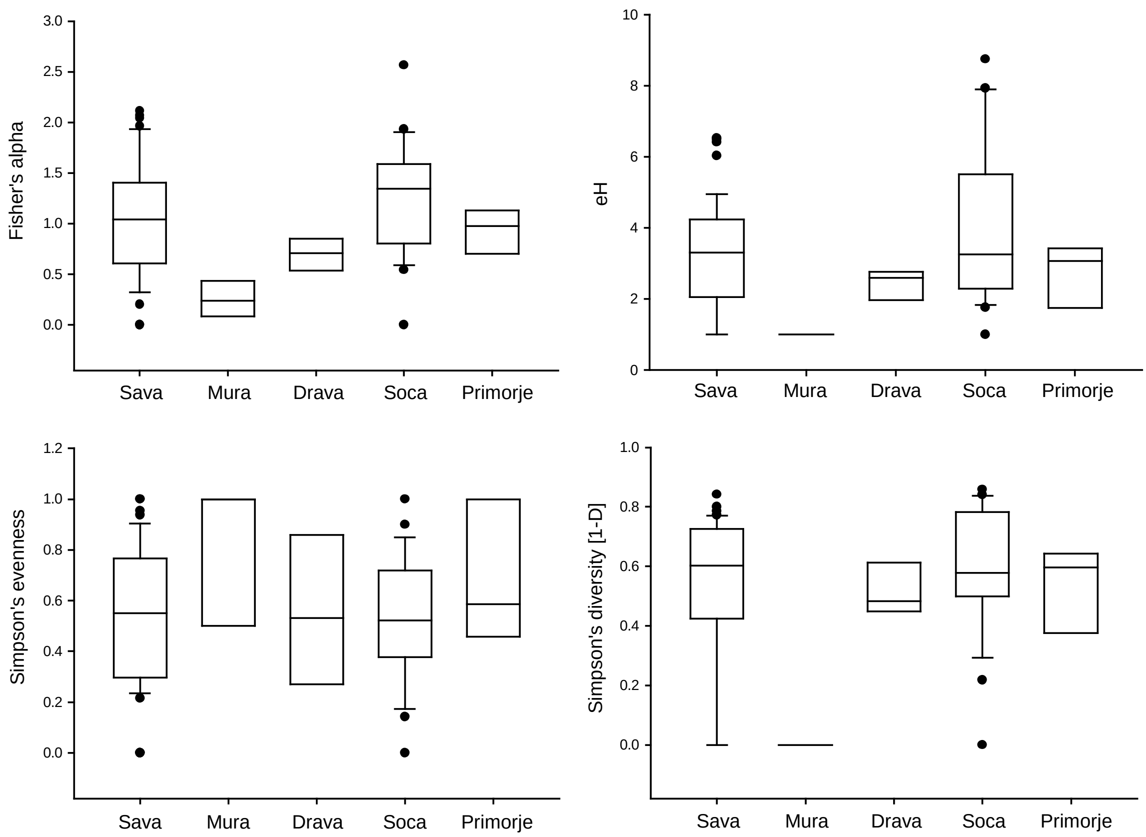


***Figure 22**: Box-Plot comparison of biodiversity among five river basins. Points denote outliers.*

***Table 9**: Results of permutation tests to compare biodiversity indices among the river basins. Above the diagonal are p-values for α and (1-D), below are p-values for $e^{H'}$ and $E_{1/D}$.*

	Sava	Mura	Drava	Soča	Primorje
Sava		0,549 0,001	0,531 0,083	0,537 0,529	0,824 0,139
Mura	0,166 0,001		0,2 0,001	0,542 0,001	0,234 0,001
Drava	0,412 0,134	0,29 0,001		0,77 0,051	0,658 0,504
Soča	0,097 0,897	0,151 0,001	0,616 0,582		0,911 0,072
Primorje	0,259 0,135	0,544 0,001	0,695 0,391	0,587 0,103	

The spatial distributions of hyporheic copepod biodiversity as calculated by the chosen indices are presented in Figure 23. Three indices (Fisher's alpha, modified Shannon and Simpson's) indicate two hotspots; the first in central-western Slovenia, and the second one in the central-eastern part of Slovenia. The eastern part belongs to the Soča basin (the hotspot is centered on the middle part of the river Idrijca), the western to the Sava (the center of the hotspot is around the confluence of the Sava and Savinja). A smaller hotspot is indicated in the southern part of Slovenia, in the middle part of the river Kolpa. The Simpson's evenness index indicates that even though diversity in the second hotspot is comparatively high, the abundance distribution of species is uneven. Unevenness of species distributions reaches maximum values in the far east of Slovenia (the river Mura) and the far west (rivers Rižana and Dragonja, and lower reaches of the river Reka). The same holds true for the lower part of the river Paka.

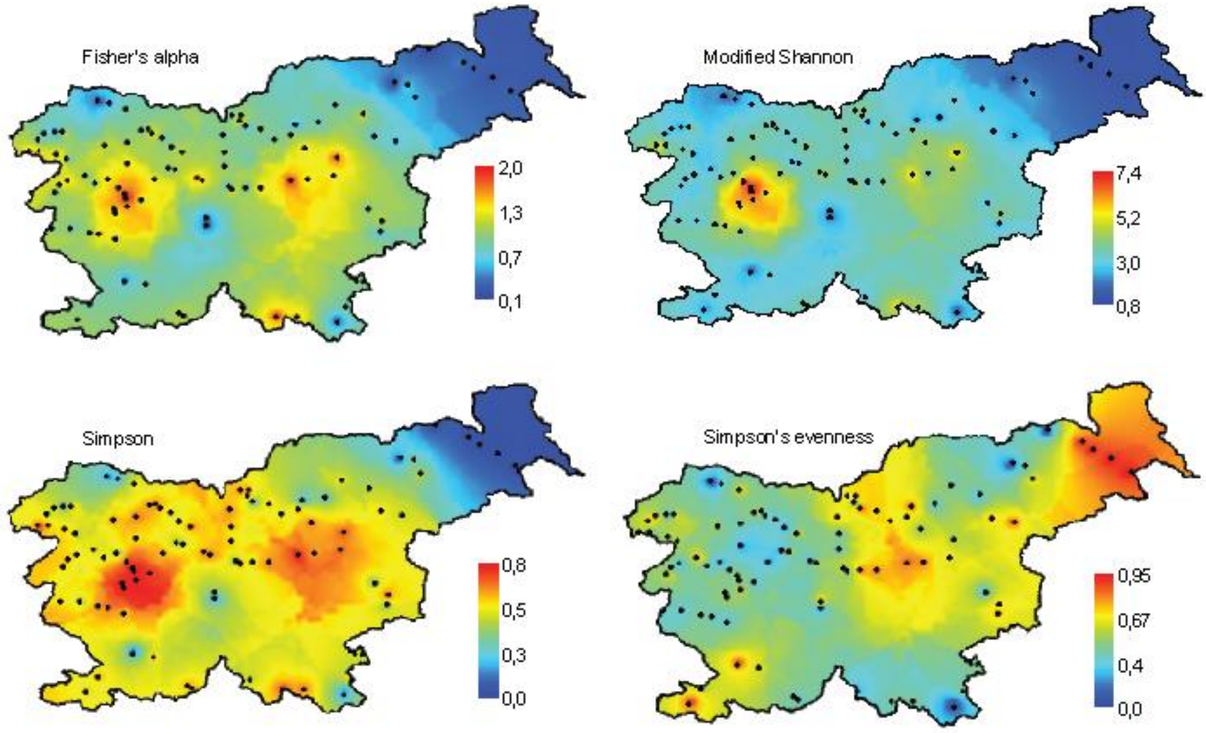


Figure 23*: Jackknifed spatial interpolation of four biodiversity indices. Black dots denote sampling sites.*

Since a biodiversity index which combines both partitions of diversity doesn't exist, we used principal component analysis as a surrogate. The first two extracted axes correspond to richness and evenness estimators, respectively. Both together explain 93,5 % of the variation in species diversity data (Table 10). Richness indices form a stronger gradient, than evenness. A distinct cluster of samples with low richness and high dominance (and therefore low evenness) is formed by the river Mura with locations from lower parts of the rivers Paka and Reka, and one location in the upper reaches of the river Savinja. The last location of the river Kamniška Bistrica is also in the vicinity of this distinct cluster. Species richness increases toward the previously identified hotspots. The main gradient on Axis 1 corresponds with a gradient from upper to lower reaches of the rivers (Figure 24). This finding was confirmed with a Pearson correlation analysis (Table 11) of diversity indices and environmental variables. Species richness was significantly correlated with nitrate, potassium and sodium concentrations, total phosphorous, oxygen saturation, hydraulic conductivity, and chemical oxygen demand. Species evenness/dominance was correlated with chlorine, sodium and potassium concentrations, oxygen saturation, and total organic carbon.

Table 10*: Principal component analysis of biodiversity indices. Correlations between the individual indices and both axes are given in the second and third column. α: Fisher's alpha diversity; H': Shannon-Wiener diversity; $e^{H'}$: modified Shannon-Wiener diversity; (1-D): Simpson's index (D); $E_{1/D}$: Simpson's measure of evenness; D_{Mg}: Margalef's diversity index; d: Berger-Parker index.*

Index	Axis 1	Axis 2
α	-0,93	-0,2
d	0,07	-0,92
$e^{H'}$	-0,94	0,19
$E_{1/D}$	0,35	-0,84
D_{Mg}	-0,97	0,03
H'	-0,98	0,001
(1-D)	-0,95	-0,06
eigenvalue	0,86	0,08
variance explained	85,6	7,9
Sum of squares	50,86	

Table 11: *Pearson correlations between PCA axes and environmental variables. Upper number: Pearson correlation coefficient; Lower number: p-values; Hcond: hydraulic conductivity, COD: chemical oxygen demand, TOC: total organic carbon, T: temperature, river_km: distance from spring, Sediment: amount of sediment in a 10 liter sample, TP: total phosphorous, TN: total nitrogen.*

	Axis 1	Axis 2		Axis 1	Axis 2
River_km	0,26 0,01	-0,11 0,3	Sat	-0,24 0,02	0,28 0,007
Sediment	-0,03 0,77	-0,03 0,77	O_2	-0,1 0,32	0,2 0,05
T	0,09 0,35	-0,1 0,31	Hcond	0,35 0,000	-0,01 0,89
Conductivity	-0,08 0,46	-0,06 0,56	BOD	0,2 0,06	-0,08 0,5
NO_2^-	0,15 0,14	-0,1 0,34	COD	0,46 0,000	-0,2 0,06
NO_3^-	0,22 0,03	-0,2 0,06	TOC	0,2 0,07	-0,29 0,004
SO_4^{2-}	0,04 0,73	-0,13 0,23	Cl^-	0,199 0,06	-0,24 0,02
NH_4^+	-0,03 0,8	-0,04 0,7	TP	0,21 0,04	-0,13 0,2
Na^+	0,23 0,02	-0,29 0,005	TN	0,08 0,44	-0,15 0,16
K^+	-0,23 0,03	0,3 0,003			

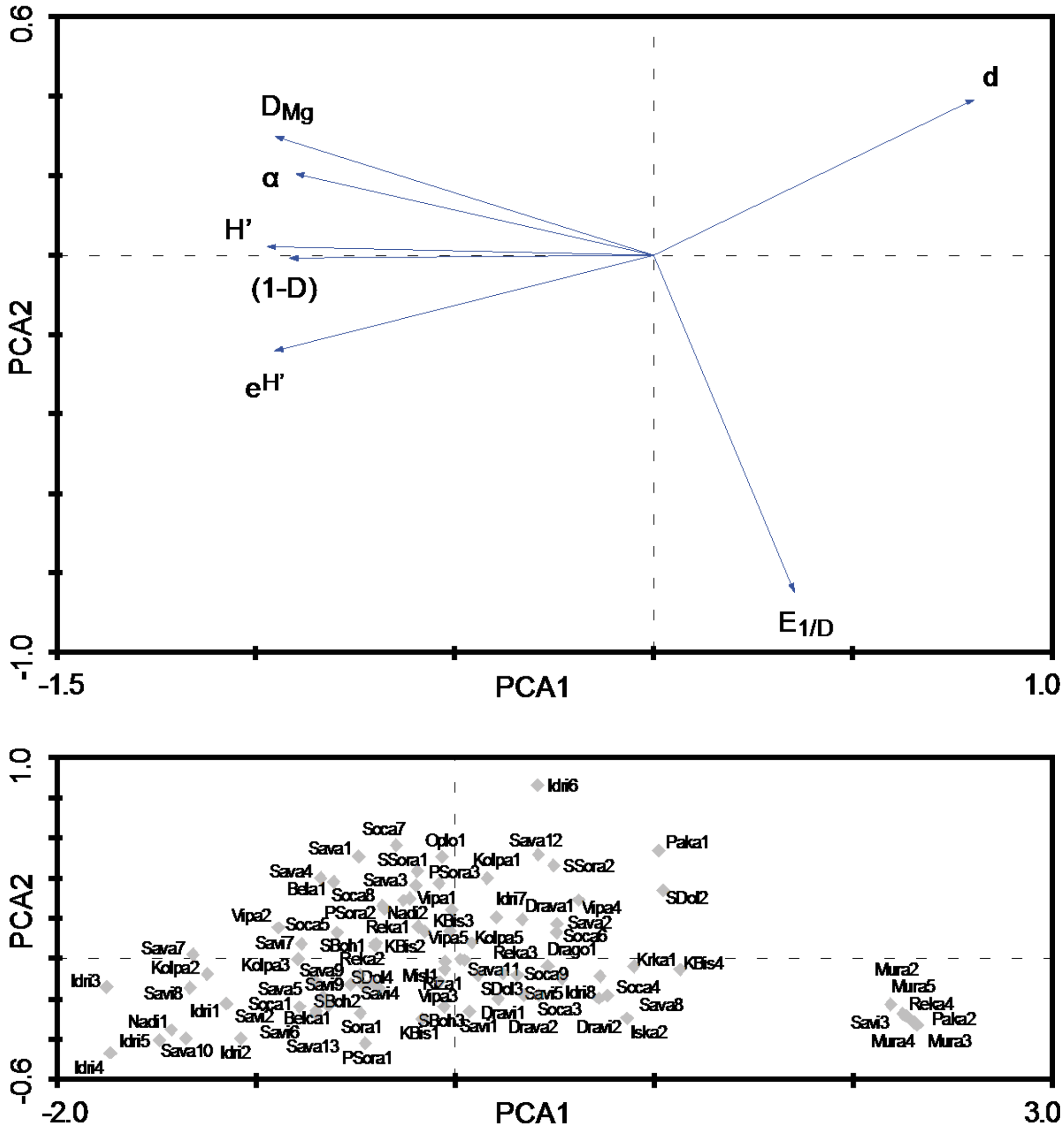


***Figure 24**: Principal component analysis on biodiversity indices. α: Fisher's alpha diversity; H': Shannon-Wiener diversity; $e^{H'}$: modified Shannon-Wiener diversity; (1-D): Simpson's index (D); $E_{1/D}$: Simpson's measure of evenness; D_{Mg}: Margalef's diversity index; d: Berger-Parker index.*

3.2 Copepod species assemblies

None of the found species can be described as being characteristic of any river basin. The only possible exception would be *E. elaphoides*, which was found in all basins, but was also the only species we found in the Mura basin. Since the distribution of this species is likely ubiquitous, its presence in the Mura cannot be described as characteristic. Because *E. elaphoides* was the only species in the Mura, it contributes most to differences between the Mura basin and the rest (Table 12). The only statistically significant differences we found were between the Sava and Drava, Mura and Soča basins (sum of squares $2,65*10^{-5}$; F: 2,63; for all three pairs p<0,001). These differences are mostly accounted for by *D. clandestinus, B. dacicus, B. zschokkei,* and *E. elaphoides*. *D. clandestinus* was most abundant in the Sava basin, and it contributes most to differences between Sava and Primorje basins, although these were not significant. Comparing the far west (Soča and Primorje basins) with the far east (Drava and Mura basins), differences are characterized by the same three species (*B. dacicus, D. clandestinus* and *B. zschokkei*). Another important species is *C. staphylinus*, found only in the Drava basin, but in such small relative abundances, that its contribution is overshadowed by more abundant species.

Table 12: *Results of SIMPER analysis of species assemblages in river basins. Shown are only those species whose cumulative contribution exceeds 50%. Above the diagonal are species abbreviations, below the individual contribution with cumulative values (%) in brackets.*

	Sava	Drava	Mura	Soča	Primorje
Sava		Diacl Bryda Elael Parsc Bryzs Canst Diabi	Elael Diacl Bryda Parsc Bryzs Diabi	Diacl Bryzs Bryda Elael Parsc Echpi Dialg Nithi	Diacl Elael Bryzs Bryda Parsc Echpi Epari
Drava	10,5 (12,2) 7,50 (20,9) 6,27 (28,2) 5,73 (34,9) 5,65 (41,4) 5,64 (48) 4,73 (53,5)		Elael Diacl Canst	Bryda Diacl Bryzs Echpi Dialg Canst Elael Athwi	Diacl Canst Bryzs Elael Bryda Echpi
Mura	13,1 (14,8) 10,3 (26,5) 7,47 (35) 7,04 (43) 6,26 (50) 4,55 (55,1)	24,4 (26,4) 17 (44,7) 11,42 (57)		Elael Bryzs Bryda Dialg Echpi	Elael Diacl Epari Bryzs
Soča	7,05 (8,5) 6,43 (16,2) 5,61 (23) 5,20 (29,3) 5,13 (35,5)	7,94 (8,8) 7,82 (17,5) 7,56 (26) 5,65 (32,3) 5,55 (38,4)	13,9 (15,5) 9,61 (26,2) 8,5 (35,7) 7,2 (43,7) 6,86 (51,4)		Bryzs Diacl Echpi Bryda Elael

	4,83 (41,3) 4,48 (46,7) 3,54 (51)	5,43 (44,5) 4,87 (50) 4,27 (54,7)			Dialg Nithi
Primorje	9,70 (11,5) 6,97 (19,9) 6,38 (27,5) 5,57 (34) 5,35 (40,5) 4,85 (46,2) 3,99 (51)	12 (13,8) 7,05 (21,9) 6,94 (29,9) 6,74 (37,6) 6,33 (44,8) 5,94 (51,6)	24,4 (26,1) 11,55 (39) 8,61 (48,4) 7,93 (57)	7,51 (8,9) 7,26 (17,5) 6,29 (24,9) 6,2 (32,3) 6,17 (39,6) 5,16 (45,7) 4,94 (51,6)	

The first division of a neighbor joining cluster analysis on zero-adjusted Bray-Curtis dissimilarities of species data generated two clusters (npManova, sum of squares $2{,}66 \cdot 10^{-5}$, F: 8,22; p<0,0001): the larger one containing sampling sites from all river basins, and the smaller cluster (designated as cluster 7) includes exclusively locations east of the river Sava (Figure 25). The locations in cluster belong to all three eastern basins (Sava, Soča and Primorje). Eight species (*B. zschokkei*, *E. pilosus*, *B. dacicus*, *D. languidus*, *P. schmeili*, *E. elaphoides*, *A. crassa* and *D. clandestinus*) contribute more than 50 % to differences between the two clusters. Seven of these species are more abundant in cluster 7, while *D. clandestinus* is completely absent. The second split (clusters 1 to 4, and 5 & 6) separates lower-reach locations from the rest. Sampling sites where no copepods were found (SDol1, Paka2, Savi3, Iska1, Kolpa4, Mura1, Drava3 and Soca2) are included in cluster 5. In all of the locations in this cluster, very few species were found (see Appendix E). Cluster 6 appears to be dominated by the Mura, since *E. elaphoides* (the only species found in the Mura) contributes approximately 30 % to the difference with cluster 5. Clusters 1 to 4 are a mix of upper and middle reach locations. Differences between them are characterized by seven of the most common species in varying proportions (Table 13). The dissimilarities between pairs of clusters are predominantly not significant (Table 14, Figure 26). Cluster 7 is significantly different from all others, except clusters 2 and 3. These three clusters contain sampling sites with some of the most diverse assemblages. The differences between cluster 1 and 2 & 3 can be characterized by *E. elaphoides* and *D. bisetosus*, the former in small abundances and the later missing from cluster 1.

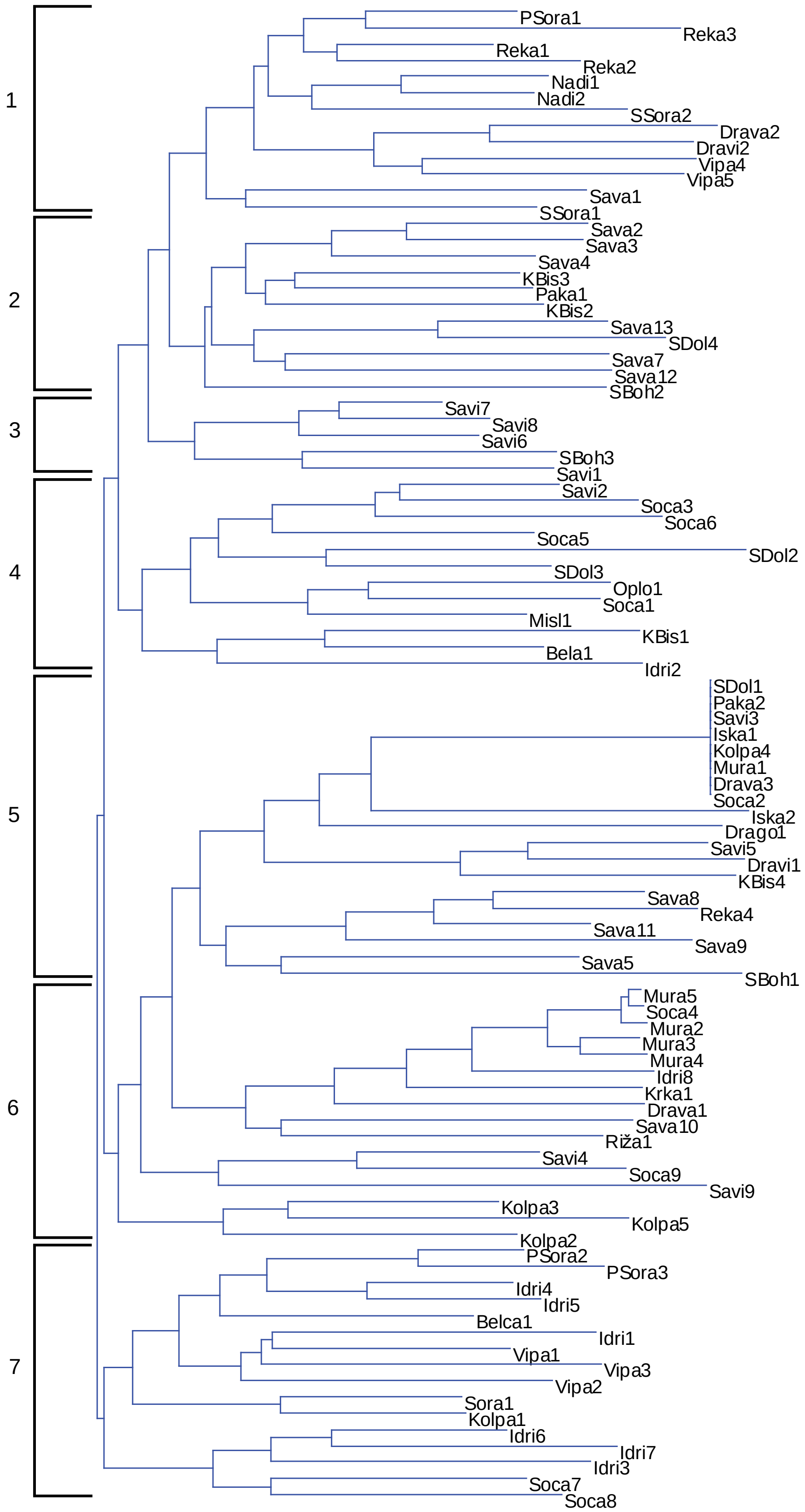


***Figure 25**: Neighbour joining cluster analysis on zero-adjusted Bray-Curtis dissimilarities of species data. The left-hand numbers indicate the clusters, referred to in the text.*

Table 13: *SIMPER analysis on the seven clusters, identified in neighbor joining cluster analysis on za-BC dissimilarity matrix. Shown are only those species whose cumulative contribution exceeds 50%. Above the diagonal are species abbreviations, below the individual contribution with cumulative values (%) in brackets.*

cluster	1	2	3	4	5	6	7
1		Elael Diacl Parsc Bryzs Echpi Acahi Bryda	Diabi Bryzs Bryda Diacl Elael	Diacl Bryda Echpi Bryzs Nithi Dialg	Diacl Echpi Nithi Bryzs	Diacl Elael Echpi Nithi Bryzs	Diacl Bryzs Echpi Bryda Parsc Dialg Elael Athcr
2	7,94 (11,2) 7,3 (21,5) 7,08 (31,4) 4,25 (37,4) 3,94 (43) 3,51 (48) 3,49 (52,9)		Diacl Diabi Bryzs Parsc Bryda Elael	Diacl Elael Parsc Bryzs Bryda Acahi	Diacl Elael Parsc Bryda Bryzs	Diacl Parsc Elael Bryda Acahi Bryzs	Diacl Bryzs Parsc Echpi Elael Dialg Bryda Athcr
3	12,14 (17,2) 7,85 (28,3) 7,72 (39,3) 6,48 (48,5) 5,84 (56,8)	8,84 (13,4) 8,29 (25,9) 6 (35) 5,32 (43,1) 4,62 (50,1) 4,24 (56,5)		Diabi Bryzs Diacl Elael Bryda	Diabi Bryzs Bryda Diacl	Diabi Bryzs Bryda Diacl	Diabi Diacl Bryda Parsc Echpi Elael Bryzs Dialg
4	15,78 (18,6) 8,09 (28,2) 6,63 (36) 6,62 (43,8) 5,42 (50,2) 3,82 (54,7)	17 (19,8) 9 (30,3) 7,59 (39,2) 4,65 (44,6) 4,26 (49,6) 3,83 (54)	13,38 (17,5) 7,89 (27,9) 7,25 (37,4) 6,5 (45,9) 6,09 (53,9)		Bryda Bryzs Dialg Athwi Bryty Epari	Bryda Elael Bryzs Dialg Parsc Athwi	Echpi Bryzs Dialg Bryda Parsc Elael Athcr Acaki
5	24,54 (25,7) 9,28 (35,4) 8 (43,8) 7,38 (51,5)	21,38 (22,7) 11,55 (34,9) 9,39 (44,9) 5,26 (50,5) 4,16 (54,9)	15,49 (16,5) 15 (32,5) 14,17 (47,6) 9,12 (57,3)	18,1 (18,5) 11,77 (30,5) 7,58 (38,2) 5,2 (43,6) 5 (48,7) 4,55 (53,3)		Elael Parsc Epari Diabi Nithi	Bryzs Echpi Bryda Dialg Parsc Elael Athcr
6	19,4 (20,8) 10,6 (32,1) 7,5 (40,2) 7,11 (47,8) 6,12 (54,4)	19,17 (23,5) 7,51 (32,7) 5,48 (39,5) 4,4 (44,9) 3,99 (49,8) 3,78 (54,4)	14,21 (16,9) 12,2 (31,4) 11,1 (44,6) 7,89 (54)	12,72 (13,6) 12,54 (27) 8,54 (36,3) 5,41 (42,1) 4,93 (47,4) 4 (51,2)	29,91 (30,6) 7,47 (38,3) 5,63 (44,1) 5 (49,2) 4,38 (53,7)		Bryzs Echpi Bryda Dialg Parsc Elael Athcr
7	9,64 (11,9) 6,7 (20,2) 5 (26,4) 4,97 (32,5) 4,74 (38,4) 4,67 (44,2) 4,28 (49,5) 4 (54,5)	12,2 (16) 5,26 (22,9) 4,54 (28,9) 4,47 (34,7) 4,15 (40,2) 3,7 (45) 3,5 (49,6) 3,39 (54,1)	8,82 (12,8) 4,85 (19,8) 4,61 (26,4) 4,31 (32,7) 4,19 (38,7) 4 (44,5) 3,97 (50,3) 3,81 (55,8)	6,78 (8,3) 6,72 (16,6) 5,95 (24) 5,33 (30,6) 5,14 (36,9) 4,65 (42,6) 4,4 (48,1) 3,79 (52,7)	12,55 (12,8) 8,82 (21,8) 7,38 (29,3) 6,49 (35,9) 6,43 (42,5) 5,78 (48,4) 5,42 (53,9)	10,28 (12,1) 7,52 (20,9) 6,2 (28,2) 5,54 (34,7) 5,41 (41) 5,31 (47) 4,89 (53)	

***Table 14**: Results of a npManova, comparing the seven clusters. P-values for each pair are given. Sum of squares: 2,66*10^{-5}, F: 9,64.*

cluster	2	3	4	5	6	7
1	0	0.0004	0.99	0.99	0.95	0.0001
2		0.99	1	1	1	1
3			0.45	0.99	0.47	0.49
4				0.99	0.9	0.0006
5					0.62	0.0057
6						0
7						

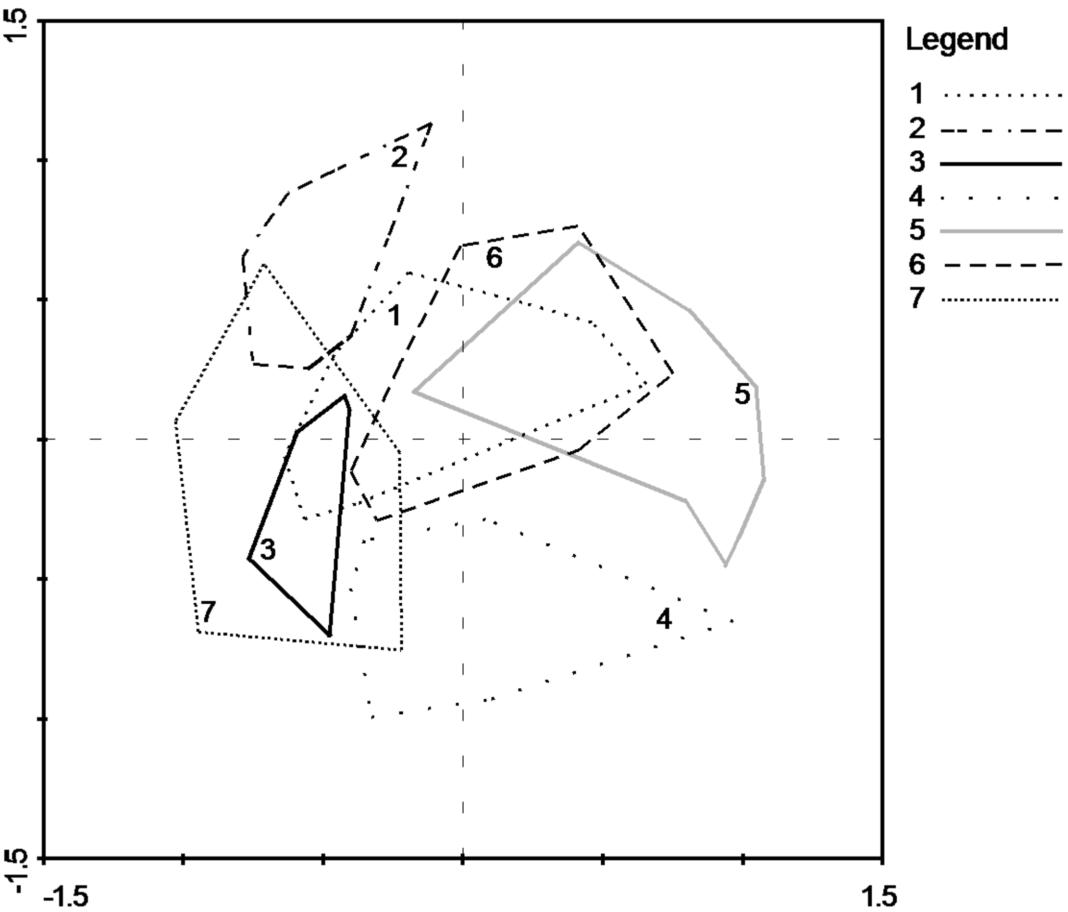


***Figure 26:** Non-metric multidimensional scaling plot of zero-adjusted Bray-Curtis dissimilarities. Clusters are the same as in the neighbor joining cluster analysis. The legend for cluster line styles is shown on the right. Shepard and stress diagrams are in Appendix F.*

3.3 Species – environment relations

Three groups of explanatory variables (water chemistry, physical habitat and geology & geography) together accounted for 73,4 % of species variance (Figure 27). WC and PH together explain approximately 40,8 % of species variance, GG 41 %. Of the two groups of attributes in the GG set, geographical location is the most important. Of the 41 % of overall variance which GG accounts for, geographical location explains 38,5 %.

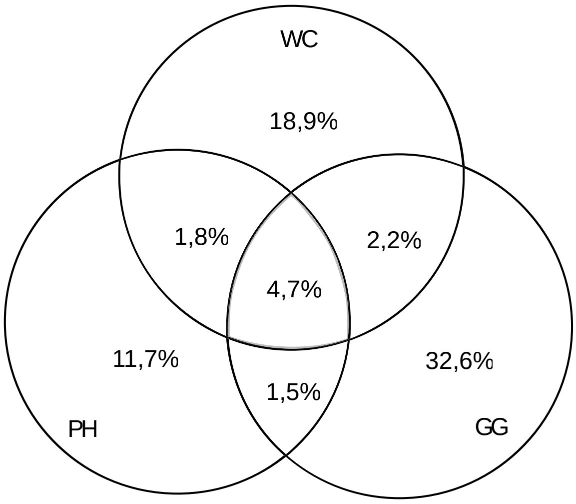


***Figure 27**: Venn diagram of variance partitioning analysis. WC – water chemistry, PH – physical habitat, GG – geology and geography.*

Since individual attributes of the environmental axes one and two are highly correlated, we used a forward stepwise selection procedure with Monte-Carlo permutation tests (full model, 499 permutations) for calculating the dbRDA. The tests revealed five statistically significantly related environmental variables: distance from the spring (river_km), temperature (T), conductivity (cond), oxygen concentrations (O_2) and hydraulic conductivity (Hcond). The remaining variables are correlated with these five. In the final analysis we also included those environmental variables, which were shown to have a significant correlation with biodiversity. An upriver-downriver gradient dominates the first axis of the distance-based redundancy analysis (Figure 28). The three main factors in this analysis are distance from spring, temperature, and oxygen concentrations. The second axis is dominated by conductivity, sulfate concentrations, and sediment on the positive end of the gradient, while total nitrogen and hydraulic conductivity are on the negative. Sampling sites are distributed along the first axis, with downstream locations prevailing in the positive side of the spectrum. Deviations from this pattern are due to the second axis. The river Mura is characterized by high COD and hydraulic conductivity, while species diversity is low. On the other side of the second axis' spectrum, upper to middle-reach locations prevail. The copepod communities of these locations are characterized by *D. clandestinus*, *P. schmeili*, *E. pilosus* and *B. pygmaeus*. The dbRDA explained 20,4 % of species variance, with the first two axes accounting for 51,4 % of species-environment relations. A Monte Carlo permutation test (full model, 499 permutations) showed that all canonical axes are significant (trace=0,204; F-ratio: 1,535; p=0,002). Biodiversity (species richness as described by 3 indices) increases from the lower right to the upper left (Figure 29). Evenness generally shows the same pattern, with a stronger increase toward the lower left. All four general additive models are statistically significant (Table 15). Most species show preference of locations in middle-reach areas, with higher food availability. In these areas, biodiversity was also the highest. *A. wierzejskii*, *S. infernus*, *M, radovnae* and *P. gertrudae* display a profound preference of higher-reach areas, close to the spring. Low-reach areas are preferred by *N. psammophila*, *M. albidus* and *P. affinis*. Locations with higher chemical oxygen demand, hydraulic conductivity and nitrogen concentrations are preferred only by *T. prasinus*.

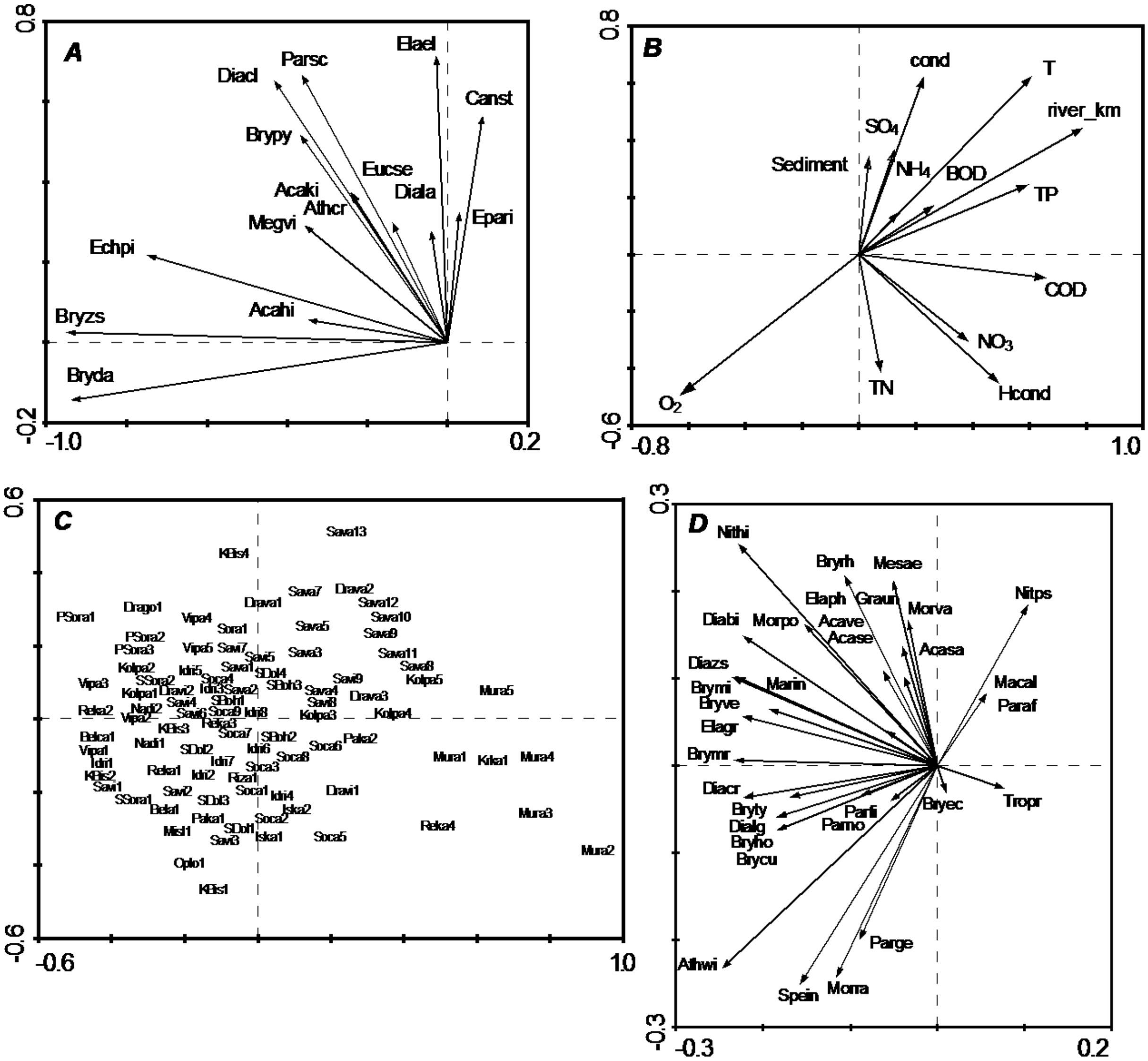


***Figure 28**: Distance-based redundancy analysis. A: common species (included are only species with correlations greater or smaller than 0,2); B: environmental variables; C: distribution of sampling sites; D: less common to rare species (correlations between -0,2 and 0,2; please note the difference in scales with A).*

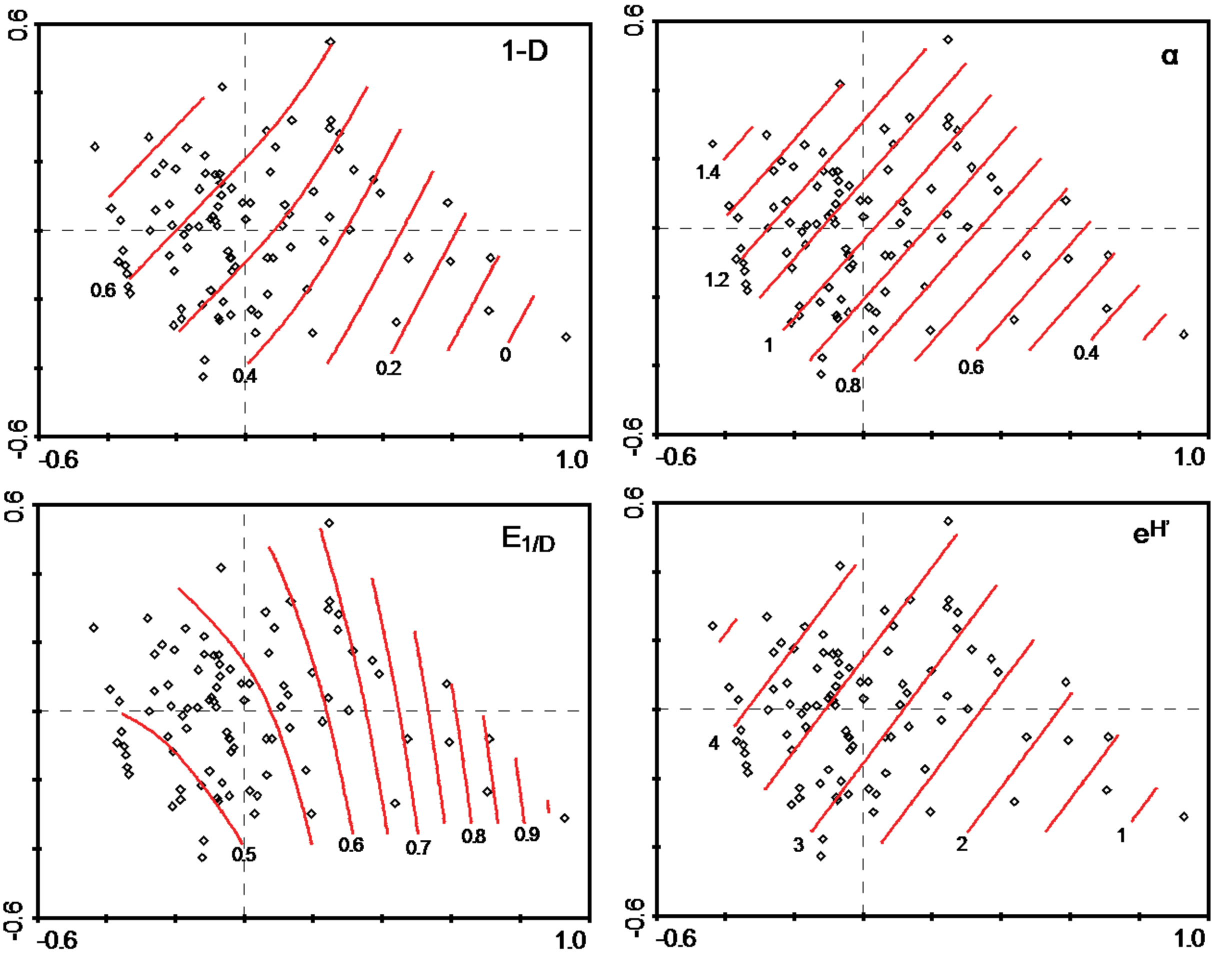


Figure 29: *General additive models based on dbRDA. 1-D: Simpson's index; α: Fisher's alpha; $E_{1/D}$: Simpson's evenness; $e^{H'}$: modified Shannon's index. Empty diamonds denote sampling sites. The distribution of sampling sites is the same as in Figure 28C.*

Table 15: *Significance tests for four general additive models of diversity indices. AIC: Akaike information criterion (Akaike 1974).*

	Index			
	1-D	$E_{1/D}$	$e^{H'}$	α
F	12,25	3,71	7,02	6,14
p	<0,0001	0.013	0,001	0,002
AIC	4,591	6,974	251,6	27,573

3.5 Model development

A principal component analysis produced 12 species with correlations higher than ±0,2 (Table 16). Together with two diversity estimates (Simpson's index and Simpson's evenness), the total number of attributes used in all data mining procedures was 14.

The M5 model tree with a minimum of 8 instances beneath each node is the best model we were able to obtain (Table 17). This model showed the best compromise between accuracy and complexity (i.e. number of rules). At lower minimal number of instances (MNI) settings, the left-hand side of the model tree was reduced to a sequence of river type nodes, ending in an equation. Based on this sequence, water quality could be assessed by river type only. This is a clear case of the model overfitting the data. At MNI set to eight, this problem does not appear anymore, even though correlations and both error estimates do not differ significantly. In regard to correlations and correctly classified instances, the M5 model tree learner performed best. However, when error estimates (root mean square and mean absolute error) are considered, the M5P's performance is not as clear-cut. Both J48 and Multilayer perceptron (MP) performed better in regard to error estimates, even though their percentage of correctly classified instances was lower than for the M5. A generally accepted rule of thumb is that correlations or percentage of correctly classified instances should exceed 0,61 or 61%, respectively, for a model to be accepted. If we apply this to our models, then only M5 can be considered as being acceptable.

All three algorithms performed best in mid-range and worst in upper extremes (classes 4 and 5, i.e. "Poor" and "Bad" status according to the WFD). Confusion matrices for three chosen models showed their performance to be best in classes 1, 2 and 3 (approximately 60 % accuracy). Performance in classes "Poor" and "Bad" (designated as 4 and 5) was negligible. None of the samples belonging to these two classes were classified correctly (Table 18). And yet, differences between the M5 model and actual data were not statistically significant (Mann-Whitney, U=4246, median M5=2,111, median actual=2, p=0,97) (Figure 30).

Table 16: *Principal component analysis of copepod species assemblages. Shown are only those species, whose correlation with the first or second axis was higher than* ±0,2. *Correlations between the individual species and both axes are given in the second and third column*

Index	Axis 1	Axis 2
E. elaphoides	-0,02	0,53
P. schmeili	-0,27	0,50
B. dacicus	-0,70	-0,11
B. pygmaeus	-0,27	0,38
B. zschokkei	-0,70	0,02
E. serrulatus	-0,10	0,22
A. crassa	-0,18	0,28
E. pilosus	-0,56	0,16
D. clandestinus	-0,32	0,49
D. languidus	-0,03	0,20
D. languidoides	-0,24	-0,04
A. kieferi	-0,18	0,27
eigenvalue	0,182	0,152
variance explained	18,2	15,2
Sum of squares	2281,6	

***Table 17:** Summaries of generated models. M5P, J48 and multilayer perceptron were the algorithms used for model creation. All models were tested using 10-fold cross-validation. MNI: minimum number of instances at each leaf/node; RMSE: root mean squared error; MAE: mean absolute error.*

M5P

MNI	Nr. of rules	correlation	RMSE	MAE
2	32	0,68	0,62	0,49
4	32	0,68	0,62	0,49
8	20	0,72	0,57	0,45
16	9	0,71	0,57	0,46
32	4	0,68	0,60	0,49
64	3	0,58	0,66	0,54

J48

MNI	Nr. of rules	Correctly classified	RMSE	MAE
2	27	40,21 %	0,42	0,26
4	17	42,39 %	0,39	0,26
8	14	51,1 %	0,38	0,25
16	5	43,48 %	0,37	0,26
32	2	39,13 %	0,37	0,27
64	1	31,5 %	0,38	0,29

Multilayer perceptron

Nr. of neurons	Nr. of sigmoid nodes	Correctly classified	RMSE	MAE
2	6	44,6 %	0,4	0,23
4	8	41,3 %	0,44	0,25
8	12	34,78 %	0,46	0,26
16	20	38 %	0,46	0,25

***Table 18**: Confusion matrices of M5P, J48 and Multilayer perceptron (MP) models. For the M5 model the results were rounded to the nearest integer and those compared with water quality classes. N: neurons.*

M5P at MNI=8						J48 at MNI=8						MP at N=2					
a	b	c	d	e		a	b	c	d	e		a	b	c	d	e	
10	14	1	0	0	a =1	14	4	5	0	0	a =1	13	9	1	0	0	a =1
8	25	10	0	0	b = 2	8	14	7	0	0	b = 2	11	11	6	1	0	b = 2
0	4	19	0	0	c = 3	2	9	19	0	0	c = 3	5	8	17	0	0	c = 3
0	0	1	0	0	d = 4	1	1	2	0	1	d = 4	2	0	3	0	0	d = 4
0	0	0	0	0	e = 5	0	2	3	0	0	e = 5	0	1	3	1	0	e = 5

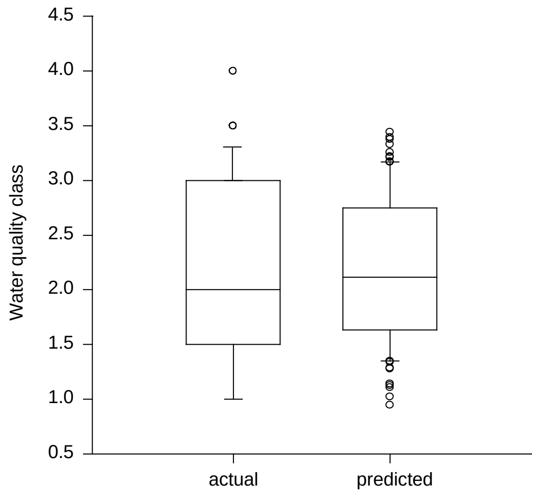


Figure 30: *Comparison of differences between actual and model-predicted water quality. Empty circles are outliers.*

Figure 31 shows the decision tree of the most reliable model we were able to obtain. In this model river type presents the most important attribute, as it divides the tree at the first node. On the right-hand side (type 8,9,11,13 = TRUE) are rivers designated as “large” (e.g. VR1), with lower water quality. The results of the equations in this part of the tree will result in highest numbers (i.e. water quality classes) (Table 19). *B. pygmaeus* either is absent or reaches low abundances, and species evenness is comparatively high. This indicates that in this part of the decision tree copepod assemblages are dominated by few species. On the far left are also locations with higher evenness values, but the river types (codes 1,2,3 and 4; Appendix E) are all in upper parts of rivers. The middle of the model corresponds with middle reaches of rivers. In this part of the model one of the biggest problems becomes apparent. Two consecutive nodes, leading to equations 6, 7 and 8, indicate that the model is overfitting the data.

A direct comparison of surface water quality as determined in accordance with WFD guidelines, and our model for the hyporheic zone, is shown in Figure 32. The failure of the model in higher classes (i.e. bad water quality) becomes more apparent. Generally, locations from classes 1 and 2 were assigned one class higher. Especially noteworthy is the difference in the lower part of the river Sava. According to the WFD the water quality in this part was moderate to poor (classes 3 and 4, respectively). However, the model assigned them to classes 1 and 2. In samples where species are distributed unevenly, the model begins struggling and becomes increasingly less accurate.

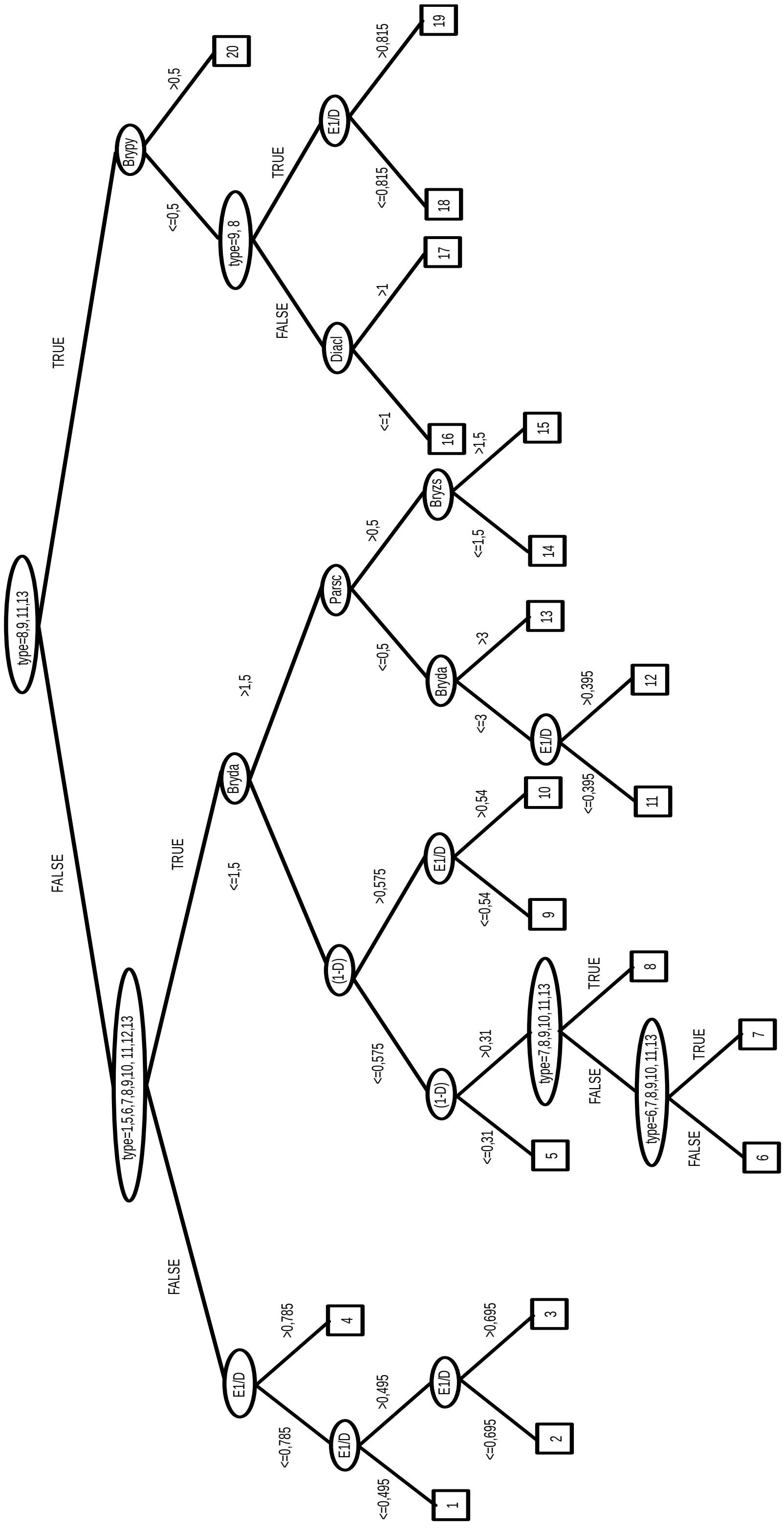


Figure 31*: Decision tree for predicting water quality from copepod species data (method M5P, MNI = 8). Ellipses: decision nodes with the discriminating parameter; Squares: final nodes with* equations for calculating water quality classes*. The equations are given in Table 18 and river types in Appendix E.*

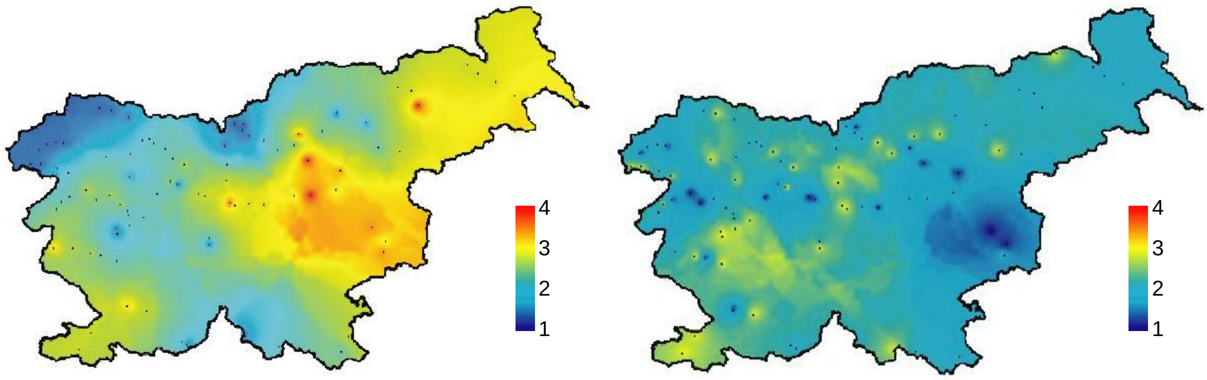


Figure 32: *Spatial interpolation of water quality classes. Left: surface water quality as determined according to WFD guidelines; Right: hyporheic zone water quality as determined by our model.*

Table 19*: Equations for predicting water quality based on the copepod community. In the first column are the numbers of each equation, corresponding with the numbers in Figure 31. The equations are in the third column. WQ: predicted water quality; type: river type. If the location under study belongs to the types listed at the specific node, then the type is substituted by 1. Otherwise, it's replaced by 0. Species abbreviations are substituted by absolute abundances. If a species is absent, the abbreviation is replaced by 0.*

1	WQ=	0,2708 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0948 * Parsc - 0,1115 * Bryda + 0,3682 * E1/D + 1,3069
2	WQ=	0,2708 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0948 * Parsc - 0,1192 * Bryda + 0,3448 * E1/D + 1,3587
3	WQ=	0,2708 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0948 * Parsc - 0,1172 * Bryda + 0,3448 * E1/D + 1,3456
4	WQ=	0,2708 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0948 * Parsc - 0,0947 * Bryda + 0,2511 * E1/D + 1,3171
5	WQ=	0,2708 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,1258 * Parsc - 0,0947 * Bryda + 0,2511 * E1/D + 1,4372
6	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0388 * Parsc - 0,1414 * Bryda + 0,6143 * E1/D + 1,7978
7	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0388 * Parsc - 0,1414 * Bryda - 0,011 * Diacl + 0,5146 * E1/D + 1,8555
8	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0388 * Parsc - 0,1414 * Bryda - 0,011 * Diacl + 0,5146 * E1/D + 1,8477
9	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,0129 * type=10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0388 * Parsc - 0,1414 * Bryda - 0,011 * Diacl + 0,5146 * E1/D + 1,8645
10	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,0166 * type=10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0388 * Parsc - 0,1414 * Bryda - 0,011 * Diacl + 0,5146 * E1/D + 1,8667
11	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0592 * Parsc - 0,1469 * Bryda + 0,3747 * E1/D + 1,6413
12	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0592 * Parsc - 0,1469 * Bryda + 0,3747 * E1/D + 1,6261
13	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,0592 * Parsc - 0,1469 * Bryda + 0,3747 * E1/D + 1,6191
14	WQ=	0,2208 * type=1,12,6,7,10,9,11,13,8 + 0,1296 * type=9,11,13,8 + 0,067 * Parsc - 0,1469 * Bryda + 0,3747 * E1/D + 1,6658
15	WQ=	0,2216 * type=1,12,6,7,10,9,11,13,8 + 0,2865 * type=9,11,13,8 + 0,0348 * Parsc - 0,1745 * Bryda + 0,188 * E1/D + 2,4828

16	WQ=	0,2216 * type=1,12,6,7,10,9,11,13,8 + 0,2865 * type=9,11,13,8 + 0,0348 * Parsc - 0,1745 * Bryda + 0,188 * E1/D + 2,4889
17	WQ=	0,2216 * type=1,12,6,7,10,9,11,13,8 + 0,2865 * type=9,11,13,8 - 0,0183 * type=11,13,8 + 0,0348 * Parsc - 0,1745 * Bryda + 0,188 * E1/D + 2,4771
18	WQ=	0,2216 * type=1,12,6,7,10,9,11,13,8 + 0,2865 * type=9,11,13,8 - 0,015 * type=11,13,8 + 0,0348 * Parsc - 0,1745 * Bryda + 0,188 * E1/D + 2,4701
19	WQ=	0,2216 * type=1,12,6,7,10,9,11,13,8 + 0,2865 * type=9,11,13,8 + 0,0348 * Parsc - 0,2358 * Bryda + 0,188 * E1/D + 2,4113
20	WQ=	0, 2216 * type=5,1,12,6,7,10,11,13,9,8 + 0,2038 * type=11,13,9,8 + 0,5471 * type=9,8 - 0,1031 * Bryda - 0,1244 * Brypy + 0,188 * E1/D + 2,4262

The validity of the model was further tested by exposing it to new data. The model performed fairly well at locations with WFD classes 1 and 2, but encountered serious problems at higher classes (Table 20). Overall performance was 50 %. Partitioned into classes, performance decreased drastically, from class 1 to class 4. In class 1, both samples were classified correctly, in class 2 a third of the samples was misclassified. In classes 3 and 4, none of the samples were classified correctly.

Table 20*: Data used for testing the performance of the model on new, unseen data. Type: river type; (1-D): Simpson's diversity index; $E_{1/D}$: Simpson's evenness; Equation: the equation number from the model; model solution: calculated solution for the given data; WFD: water quality according to WFD guidelines. Numbers next to species are absolute abundances.*

	t_KBis1	t_KBis2	t_KBis3	t_KBis4	t_KBis5	t_KBis6	t_KBis7	t_KBis8
type	2	2	6	6	6	6	6	6
(1-D)	0,72	0,69	0,84	0,85	0,85	0,80	0,60	0,60
$E_{1/D}$	0,89	0,82	0,78	0,65	0,73	0,82	0,82	0,82
Elael			3	3	1			
Epari			1	2				
Parsc					3	3		
Bryda	2	3	2	4	2	1		
Brypy			2	1				
Bryzs			2	3	3	4		
Bryty					2			
Morra	3	1						
Canst							2	3
Diacl			5	4	5	6		
Diala			3	4				
Diacr					4	3		
Diabi							3	2
Spein	1	1						
equation	4	4	12	13	12	10	10	10
model solution	1,35	1,24	1,85	1,50	1,86	2,30	2,51	2,51
WFD	1	1	2	2	2	3	4	4

4 DISCUSSION

4.1 Environmental conditions

Most of the measured environmental variables exhibit an upstream-downstream gradient. The temperature of running waters depends on elevation, climate, extent of riparian vegetation, groundwater inputs, water depth and flow velocity (Allan 1995). Even though temperature in gravel beds shows a pronounced gradient, from low temperatures at high altitudes, to high temperatures at lower, fluctuations in interstitial habitats, adjacent to the main river channel, are greater than in the main channel and underlying sediments. This may be due to direct exposure to air temperatures and sunlight, or intensity of groundwater – surface water exchange. Low hydraulic conductivity enhances water residence time and can thus influence temperatures. On the other hand, high hydraulic conductivity can augment water exchange with the main channel, thus lowering temperatures in gravel beds (Stanford and Ward 1993; Brunke and Gonser 1997).

The largest part of gravel bed sediments had a high oxygen content and saturation (> 3 O_2 mgl^{-1} and > 60% saturation) thus indicating good permeability. This is further demonstrated by negative correlations between oxygen content and saturation, with hydraulic conductivity. In coarse sediments, oxygen concentrations are usually high, and hypoxia (< 3 O_2 mgl^{-1}) is in most cases a comparatively rare occurrence (Bretschko 1991; Stanford *et al.* 1994). The distribution of oxygen in the hyporheic is heterogenous and influenced by many factors. Its interpretation can therefore become rather complex (Ward *et al.* 1998). In order to reduce the complexity of this particular interpretation problem, we sampled only the upper part of the hyporheic. This way surface water was the prevailing factor influencing oxygen content and saturation, as well as general chemistry. Greater differences in oxygen content and saturation between the river and hyporheic could be attributed to pollutants, biological factors (i.e. microbial communities), or a combination of both. More impacted sited could thus be identified (Younger 2007).

Approximately two thirds of all sampling sites in our study are arranged along the upstream-downstream gradient. Locations close to springs have higher oxygen content and lower temperatures, conductivity and total phosphorous, compared to those in downstream reaches. Hydraulic conductivity does not completely conform to this gradient, even though a positive relation with the distance from the spring is evident. Many of the studied rivers have their springs in glacial valleys where glacial chalk deposits can clog interstitial spaces and in extreme cases form impermeable sections within gravel beds. Along the water course, the amount of fine and organic sediment increases. The effect can be similar to glacial chalk deposits, in that it impedes water transfer between the main channel and hyporheic. In addition, fine and organic sediments contain nutrients, thus enhancing microbial activity. This in turn leads to higher oxygen consumption (Wood *et al.* 2007). Sulfate and chloride concentrations also increase with the distance from the spring. Their presence in water is naturally associated with leaching from sedimentary rocks and minerals. Another source of sulfate is decomposing organic matter. Both can also indicate the presence of industrial and municipal waste. Slovene national legislation sets the limit (between good and bad status) for sulfate in surface waters, at 150 mgl^{-1}. Nitrate is set at 25 mgl^{-1}. Limit values for chloride, potassium, magnesium, calcium, and sodium are not defined (UL RS 2002). None of the studied locations exceed the limit for sulfate. Nevertheless, six locations have concentrations higher than 40 mgl^{-1}, and can be, in regard to the rest, considered as outliers. Five of them are in lower reaches (Drago1, Savi9, Drava3, Paka2 and Sava11) and pn in upper (Kolpa1). Sulfate concentrations in these locations are at least twice higher than in adjoining locations (Paka2 has almost seven times as much sulfate than Paka1), thus indicating a pollution event. The limit for nitrate is exceeded only by one location (Savi7), while two locations reach above 20 mgl^{-1} (KBis4 and Drava2). All three locations are in lower reaches and belong among those with bad water quality status.

Deviations from this main gradient are signified by longer retention times, higher biological oxygen demand, and nitrate and ammonia concentrations. Nitrates themselves are not necessarily an indicator of pollution. Groundwater is devoid of plant life and thus primary production (although cases of chemoautotrophy were reported) (Griebler *et al.* 2001). Furthermore, mineralization of organic compounds (either natural or human-derived) acts as the main source of nitrates in the hyporheic (Younger 2007). As a result, nitrates can accumulate. This becomes apparent at springs and upwelling zones, where nitrates from groundwater are transported into surface waters. The upper parts of the Soča, Sava Bohinjka and Sava Dolinka rivers may be such a case. Generally, nitrate concentrations were higher in the hyporheic, than in surface waters. This is in part a consequence of our sampling regime. We sampled during summer, at low discharge. Under such conditions, the water table surrounding the main channel can reach above it, and the river is recharged with groundwater (Younger 2007).

Slovene guidelines for assessing chemical water status are based on nitrate concentrations and biological oxygen demand. They fulfill the requirements of the WFD in that they are systematically and easily measured, and associated to well-known sources, inter-relationships, and processes. The correct definition of these reference states is a key element in the implementation of the Water framework directive. We applied a combination of approaches (reference-site approach, ambient distributions, predictive modeling and expert judgment), which we deem to be the best approach in the absence of valid historical data. The WFD also specifies that all benchmarks have to be type-specific. River types for Slovenia have been classified using several WFD-approved classifiers. These included ecoregions, bioregions, and geological substrate. This typology is fairly accurate and was shown to be applicable to the hyporheic as well. This way one hurdle, for including hyporheic communities in water quality assessment schemes, is avoided.

Water quality in an ecotone (e.g. the hyporheic zone) largely depends on the quality of influx water, flow velocity, and content of organic matter within the ecotone. Under ideal conditions, if the input consists of nitrate rich groundwater, denitrification will be the main biogeochemical process. The final products are gaseous nitrogen and carbon dioxide. However, oxidation is rarely a 100 %, and dissolved organic carbon and ammonia are released (Vanek 2009). Yet enrichment is a natural process, where organic material derived from the catchment area is decomposed. Some form of flow hindrance through the hyporheic, such as siltation, can lead to strong water quality deterioration. Hypoxia and anoxia, increased nutrient, metal and metabolites content are all natural indicators of lower water quality. Pollutants of strictly human origin would be, for example, pesticides and certain toxins.

4.2 Species richness

Our sampling effort sufficed to achieve a good overview over copepod species richness in the hyporheic of Slovene rivers. Approximately 100 samples would be the minimum requirement for a species richness estimate, at river basin scales. For two basins, Drava and Primorje, a larger sampling effort would be needed to attain better estimates. In the third undersampled river basin, Mura, only one species was found in all samples. This indicates a strong dominance of *E. elaphoides* and a much larger sample size would be needed to find any additional species.

Species richness estimators, which are still rising, or are unstable when all samples have been included, do not provide a reliable estimate (Magurran 2004). In such cases a valid minimum estimate of richness can be obtained with Chao estimators (Longino *et al.* 2002). Chao1 stops increasing once at least two individuals represent every species, and the inventory can be considered as complete (Coddington *et al.*1996.) Colwell and Codington (1994) are of the opinion that a census can be regarded as complete if all species have an abundance of at least two (when abundance data are collected) or occur in at least two samples (when occurrence data are used). With our sampling effort all species were represented with at least two individuals, which were present in at least two samples. According to both sets of guidelines, the inventory of hyporheic copepods in Slovenia can be considered complete. Estimated richness would therefore be around 50 copepod species in the hyporheic zone of Slovene rivers. However, when we calculated species accumulation curves for each river basin individually, the results are not as clear-cut. The Sava and Soča river basins were sampled the most thoroughly, and estimates reach 40 and between 30 and 40 species, respectively.

Species accumulation curves did not approach an asymptote in the Drava and Primorje basins, thus indicating that species richness in these two basins is underestimated. But while sampling effort influences species accumulation curves, species richness is influenced by several other factors. Tributaries of the river Drava (rivers Oplotnica, Mislinja and Dravinja) all have their springs on the mountain range Pohorje, which is silicate. In addition, there are nine hydropower plants on the main channel of the Drava. Each of these profoundly changes river hydrology and gravel carrying power of the river is significantly reduced (Bonacci 1987, Wood *et al.* 2007). Furthermore, all riverbeds have been regulated in the past to a certain degree. River regulation has a great impact on groundwater – surface water ecotone assemblages by modifying the dynamics of water exchange between the river and alluvial aquifer (Marmonier *et al.* 2009). The number of available and acceptable gravel beds is therefore low. Consequently, we could not sample along the entire environmental gradient, as we could in the Sava and Soča rivers. The tributaries could be sampled only in the upper parts, with the exception of Dravi2, a location in the middle part of the river Dravinja, where it's regulated. The river Drava's spring is in Germany, by the time it reaches Slovenia the river has been flowing for approximately 300 kilometers, past several large cities and countless farms. This leads, together with river bed regulations and hydropower plants, to the river Drava being one of the most impacted rivers in Slovenia. Under such environmental conditions, species assemblages would be heterogeneous, dominance levels would be high. The Primorje basin, on the far west of Slovenia, has similar problems. The rivers Rižana and Reka are regulated in almost their entire length. The geological composition of sediments of the river Reka is carbonates; the river itself has the capacity of gravel transport. But water quality deteriorates rather quickly, thus leading to heterogeneous assemblages. The river Dragonja is intermittent, able to carry gravel only at stronger discharges. The regulation is not as problematic as with the other rivers, but its intermittent nature leads to problems with oxygen supply.

The Mura basin presents a deviation from the already mentioned patterns. All tributaries there are regulated and too small to transport gravel. In addition, the main channel is heavily regulated with concrete slabs along its entire stretch in the Slovenian part of its channel. Even though the river has a significant carrying capacity, gravel beds are at least partly separated from alluvial aquifers, due to riverbank regulations. Surface water is thus most often the prevailing input into the interstitial. In all samples, we found only one species in comparatively low abundances. The most probable explanation is a too low sampling effort (15 samples) combined with an assemblage strongly dominated by just one species. The probability of finding less abundant and rare species is therefore low. In overall, the rivers Mura and Drava are quite similar and species richness in both can be expected to be close to equal.

4.3 Copepod biogeography and biodiversity

The Simpson diversity index was shown to be the most appropriate diversity measure we used. Out of four tested indices (Fisher's α, H', $e^{H'}$ and 1-D) it was the only one that showed the prevalence of uneven samples. Since Simpson's index emphasizes dominance of a community, when the reciprocal (1-D) is used, higher values indicate a more even assemblage. Even though a log series model indicated that Fisher's α is a good descriptor of diversity, it failed to show the patterns Simpson's index revealed.

Hyporheic copepods appear to have two biodiversity hotspots, in the Soča/Idrijca and Sava/Savinja confluence areas. But Simpson's evenness revealed the second hotspot has an uneven species abundance distribution. Equilibrium communities consist of several species in similar abundances, and dominance is low. After disturbances the equilibrium shifts, relative abundances change and previously abundant species can become comparatively rare. Thus new species can colonize the habitat, even though they are less competitive than autochthonus species. In this way, biodiversity can increase, even though evenness decreases. After the disturbance passes and equilibrium is re-established, biodiversity decreases. If such ecological disturbances are not too strong, and neither too rare nor frequent, local species diversity can be maximized. This phenomenon is referred to as the intermediate disturbance hypothesis (Connel 1978). According to this theory, at low to intermediate levels of disturbance, diversity is maximized because K-selected and r-selected species can coexist. K-selected species tend to be more competitive, they invest a larger amount of resources into individual growth and competition. Therefore, they generally dominate stable ecosystems. On the other hand, r-selected species have the ability to colonize areas rather quickly and therefore dominate habitats recently impacted by disturbances. If these disturbances appear occasionally, both K- and r-species can coexist. For that reason, the second hotspot (Sava/Savinja) cannot be regarded as a true biodiversity hotspot, as species distribution is uneven and indicates the presence of a disturbance.

Species diversity in a river follows a distinct gradient. In the upper reaches, diversity is low, only few species are present. Diversity then increases rapidly as the river widens, temperatures increase, and resources become more abundant. Such conditions allow the coexistence of several species, while in the upper reaches only certain species can survive (for example frigostenotherm species in the Alps) (Krebs 2001). At a certain river size (depending on the river) diversity reaches a plateau (set by species competition, predation, disease and niche overlap), establishing a dynamic equilibrium (Odum 1969). Debeljak (2002) and Jørgensen (2007) describe this process in terms of thermodynamics, as a relation between the amount of exergy stored in the ecosystem and the amount of exergy utilization or destruction. At climax, the number of species is in a dynamic equilibrium, but the amount of information stored in the system increases, while utilization remains the same (Jørgensen and Svirezhev 2004). Here we have to acknowledge that climax conditions differ from region to region and have to be evaluated for each area. In lower reaches of rivers a higher diversity than in the upper reaches would be expected. However, as we have seen with lower reaches of certain rivers in Slovenia (for example rivers Kamniška Bistrica, Mura, Drava), diversity recedes, with only one or two copepod species present in the hyporheic. This is probably a consequence of human activity (e.g. pollution, hydromelioration). These locations deviate from the expected diversity of lower reaches.

Culver *et al.* (2006) speculate that the last ice age is the main reason for the existence of the Idrijca hotspot. This region was not covered in ice; food and water were available at most times. Further south, in the Dinaric Mountains, the environmental conditions were similar. But, in the last 3000 years the karst regions of Slovenia were clear cut, today virtually no forests remain. Surface river flows are few, most of them small, with a low gravel carrying capacity (Gams 2003). Even though this region might have been a hyporheic biodiversity hotspot, historical data does not exist and we have no way of knowing, whether it was. The river Idrijca region therefore remains the only true hyporheic copepod biodiversity hotspot.

Groundwater fauna is characterized by a high degree of endemism. In most cases, endemic species are limited to only one river basin or karst system (Marmonier *et al.* 2009). Of the 49 species we found in our study, one (*M. radovnae*) can be considered as an endemic species. So far this species has been found only in the Eastern Julian Alps and Karavanke mountain range (we also found it in the Kamniško-Savinjske Alpe) in Slovenia (Brancelj 2001, pers. comm.). Endemism makes freshwater species highly vulnerable to disturbance, since due to its constrained geographical distribution, immigration from other populations (if they exist) is at least very difficult if not outright impossible.

Most of the rivers included in our study were not glaciated, except for their northernmost parts. A large-scale climatic shift in the Pleistocene had a significant role in structuring the composition of copepod assemblages in shallow hyporheic zones. Yet this habitat offers comparatively little protection during glaciations events and karstic fissures may have served as refugia (Stoch 2000, Galassi 2001). The upper reaches of rivers, which were under the glaciers or close to them, have therefore similar species assemblies. Lower parts of rivers tend to have similar environmental problems (e.g. riverbed regulation and hydropower plants) and only species adapted to such environments can persist. When the assemblages of low-lying and upper reaches are compared, a pattern emerges. In lower, more impacted, reaches the most common species prevailed, but average abundances were lower than in upper reaches. This indicates that these species were at the extreme end of their tolerance to certain environmental conditions. *C. staphylinus* was the only species found exclusively in lower reaches. This species was shown to be tolerant of less favorable environmental conditions of impacted sites (Kurashov 1996).

One of the primary goals of field studies in ecology is to estimate how many species of a given taxon occur in an area. The rate at which new species are added to the inventory gives important clues about the species richness. The order in which samples are added to the curve influences its overall shape. For example, a species-rich sample will have a much greater influence on the accumulation curve, if it is encountered early in the process. To avoid this effect and obtain a smooth curve, a randomizing procedure is performed (usually with 50 iterations) (Magurran 2004). As the number of samples increases and more and more rare species are added, the curve reaches an asymptote. Beyond this point, increased sampling effort will yield very few or no new species at all. One of the benefits of species accumulation curves is that they can be extrapolated to larger sample sizes. That way an estimate of species richness in a certain area is possible.

In our study we focused on the longitudinal gradient, ignoring vertical and lateral (perpendicular to the main channel) gradients of species diversity. This way we minimized variance due to sampling site characteristics, and could focus on overall biodiversity patterns. These patterns are influenced by a variety of interacting factors. Ward and Palmer (1994) divided these into eight categories: (1) characteristics of the alluvium; (2) exchange properties; (3) disturbance; (4) hypogean affinity; (5) food resources; (6) biotic interactions; (7) reproductive patterns, and (8) age distribution. Yet few definitive data are available to assess their specific roles in field studies. To the best of our knowledge, no study up to date has encompassed all these factors. Categories (1), (2) and (3) and (5) entail extensive modeling approaches, while the rest require ecological and natural history approaches, including experimental studies.

Biodiversity is divided into two components, richness and evenness. Indices dealing with biological diversity must emphasize one of the two constituents; a perfectly unified diversity index is not possible (Magurran 2004). One possible solution is to calculate several diversity indices for each sampling site and use the results in a principal component analysis. The first two axes represent richness and evenness, and will account for most of the variation (Clarke and Warwick 2001). Each biodiversity index measures either species richness or evenness in its own way. The Shannon-Wiener index has its origins in information theory and tries to encapsulate both aspects of biodiversity. Even though this index is in wide use, there are an overwhelming number of papers advising against it (Magurran 2004). Some of them go out of their way to underline its disadvantages (Lande 1996, Martín and Rey 2000, Southwood and Henderson 2000). One of the main drawbacks is its narrow span of values, as it rarely reaches above 4. Hence interpretation becomes difficult. This problem can be circumvented by using $e^{H'}$. This gives the number of species at a location, if all species were equally common. Additionally, H' is not a robust measure of diversity, as it is sensitive to sample size. Yet it is still being used, even though better methods are available (Magurran 2004).

4.4 Species-environment relations

None of the found species showed any real habitat preference. *M. radovnae* and *C. staphylinus* could be regarded as exceptions, but both were found in a very limited number of samples, where environmental variables have limited ranges. These two species could be placed at opposing ends of an upper reaches - lower reaches environmental gradient. *M. radovnae* appears to prefer low temperatures, high dissolved oxygen content and low food availability, while *C. staphylinus* favors the opposite. The observed distributions of these two species in our study confirm the observations by Brancelj (2001) and Kurashov (1996).

Even though *B. dacicus* has a widespread distribution, it appears to prefer locations with higher oxygen content and lower temperatures. This species' preferred habitat is groundwater, but it has been found springs and benthos of alpine lakes (Jersabek *et al.* 2001). *A. wierzejskii* seems to prefer similar environmental conditions. Three more species appear to favor such habitats. Whether *M. radovnae* is a stygobiotic species is as yet unknown, although the amount of data confirming its stygobiotic nature is increasing (Brancelj, pers. comm.). On the other hand, *S.infernus* and *P. gertrudae* are both stygobiotic (Einsle 1993, Janetzky *et al.* 1996). The latter was found only once and is probably a chance find. *Speocyclops* on the other hand was found on three locations, two of them (KBis1 and Bela1) comparatively close to springs, while the third (Idri2) is downstream of the karst siphon lake Divje jezero. This species was therefore probably sampled in upwelling zones, where they were pushed out of the aquifer into the main channel.

On the other hand of this spectrum, *E. elaphoides, C. staphylinus* and *N. psammophila* show a preference of warmer habitat, with better food availability and higher amounts of sediment. Of these three only *E. elaphoides* was found in a large enough proportion of samples that any judgments on its habitat preferences can be given. Even though this was the only species found in the river Mura, *E. elaphoides* shows an increase away from the Mura cluster. In this river, abundances of *Elaphoidella* were among the smallest. In addition, most of the locations where this species was found are located close to the zero-point of both axes, or on the positive side of the 2nd one. Data from these sampling locations prevails over the Mura data and *E. elaphoides* shows a decrease towards the Mura cluster.

The remaining species are largely distributed in the medium ranges of the main (upstream – downstream) gradient of the distance-based Redundancy analysis, or their correlations with this gradient are low (between -0,2 and 0,2). Increasing conductivity and amount of sediment, shorter water retention times, and lower nitrate concentrations appear to be the only factors influencing these species. These variables indicate that the water in the hyporheic of these sampling sites was mostly from the river, that these could have been downwelling zones. Epigean species therefore prevail.

Water chemistry and characteristics of the physical habitat explained approximately 40 % of the variability. Geological composition of sediments appeared to have little impact, explaining only 3 %. Water chemistry and physical habitat characteristics are both subject to the same upstream-downstream gradient. Deviations from this pattern are most often a consequence of human interference. The comparative unimportance of geological characteristics is probably the result of Slovenia's great geological diversity. Most rivers have tributaries originating from several different geological compositions and the water in the main channels represents a mixture of all influences, with none prevailing (Bonacci 1987, Gams 2003). Geographical location (river membership) accounted for 38 % of the variability, thus being the single most important factor. Water chemistry and physical habitat, both of which include human impacts, come second in importance. This suggests that historical factors (i.e. since the last ice age) are the main determinant in shaping copepod species assemblies.

In ecological datasets (e.g. sampling location by species abundance data matrix), an inflation of zero values is quite common. Therefore, using mathematical transformations, a normal distribution is impossible to obtain. A negative binomial distribution usually works best with such datasets, but is not appropriate for all statistical methods. In unconstrained ordination analyses (such as principal component analysis and correspondence analysis), such data can lead to prolonged gradients (i.e. larger eigenvectors). These, in turn, do not realistically represent the gradients in species data. In addition, relationships between the derived components are often quadratic, instead of linear. Correspondence analysis is based on the χ^2 distribution and is less prone to these problems, than principal component analysis (which is based on Euclidean distances). Constrained ordination methods (e.g. redundancy analysis and canonical correspondence analysis) suffer from the same problems, but to a lesser extent (Legendre and Legendre 1998). One option in avoiding the effects of zero inflation is the use of distance or similarity matrices. These matrices can then be used in constrained ordination methods.

The zero-adjusted Bray-Curtis dissimilarity was developed specifically to deal with ecological datasets (Clarke *et al.* 2006). The za-BC is a semi-metric measure and is therefore inappropriate to be used in redundancy analysis, which is in Euclidean space (RDA is basically a constrained version of principal component analysis). For that reason, a principal correspondence analysis on a za-BC dissimilarity matrix has to be performed prior to the db-RDA. Since the Bray-Curtis dissimilarity does not obey the triangle inequality axiom, principal coordinate analysis is likely to produce negative eigenvalues, when numbers of individuals are quite different from site to site. This effect can be corrected for by using a square-root transformed BC dissimilarity matrix (Legendre and Legendre 1998).

Interstitial communities are determined by hydrology, geomorphology, disturbance history, biological interactions, temperature, and seasonal and interannual variation (Strayer *et al.* 1997). We minimized the effect of the last two by sampling only in summer. Furthermore, variation due to hydrological conditions was minimized by sampling during low river discharge. Stream sediments are highly heterogenous in their sediment size distribution, and species distributions mostly follow these patterns. By obtaining three samples from each gravel bed we encompassed at least a small part of this variability.

4.5 Model development

Our results indicate that copepod community, combined with an accurate typology and diversity measures, can be used for water quality assessment. The obtained final model is most accurate in mid-range classes, while being inaccurate at extreme values. The most likely reason is the low number of locations of very low or very good WFD status (classes 1, 4 and 5). We analyzed only two locations with WFD class four; these two locations represent only 2 % of all sampling stations. With the chosen validation method (10-fold cross validation) the chances of these extreme cases to be included in the learning process, are comparatively low. If these cases were not included in the learning process, the algorithm could not learn how to classify them. Furthermore, none of the locations were of class 5 ("Bad"). Wrong classifications can therefore be, in cases such as this, expected. In general, a 1:5 ratio between the number of attributes and instances, is recommended for data mining purposes (Witten and Frank 2005).

We have achieved this requirement when all data is pooled, but considering each river type, the amount of quality data is lacking.

The developed model can be counted among the group of Diversity indices. Changes in diversity (including evenness) and community composition are expected to reflect disturbances. The main benefit of this index is the inclusion of a hitherto unconsidered part of riverine ecosystems in surface water quality assessment. Animals in the hyporheic are exposed to potential pollutants for longer periods, and a stronger effect would be expected (Newman 2009). This new method can be used in all Slovene river basins, but is not defined for anywhere else. If species compositions of hyporheic copepods are similar as in Slovenia (especially if the same species prevail) this index can be used, but with caution and not as a stand-alone measure of surface water quality.

Of the five guidelines, described by O'Connor and Dewling (1986), this index fulfills four: It's quantifiable, easily understood by laymen, scientifically justifiable and acceptable in terms of costs. The fifth, relevance, can only be shown experimentally and with prolonged use. Furthermore, this new index also has two characteristics of good ecological indicators (Jørgensen *et al.* 2005): ease of handling and possible quantification. The others are sensibility to small variations in the environment, applicability in extensive geographical areas and in the greatest possible number of communities or ecological environments, and independence of reference states). In our case, reference states are not independent, but are based on reference states for surface water. There are not enough data on hyporheic waters for establishing independent benchmarks. While this new method is applicable for all of Slovenia, it is limited to rivers in this country only. In addition, it considers only one community and one environment (copepods in the hyporheic). Whether this index is capable of detecting low concentration and short term pollution (i.e. is sensible to small variations in environmental stress) remains to be tested. None of the currently used indices fulfill all of these criteria. Deficiencies of individual indices (e.g. our new index, Saprobia, Trophic index) can be circumvented by using several indices, thus encompassing as much of the ecosystem under study as possible.

Geographical data are of utmost importance in the model. Just by including two geographical attributes, altitude and distance from the spring (river kilometer), the model's precision could be vastly improved. These two attributes can help to pinpoint the exact location and are therefore usually most useful in connection with site-specific reference sites. However, this approach to defining reference sites requires an independent definition of benchmark conditions for each sampling site under study. For large scale or long-term investigations, such requirements can become burdensome and require comparatively large amounts of time and resources. For such studies, definitions of reference conditions on a larger scale are more appropriate. The site-specific approach requires reference conditions to be defined for each sampling session, unless long-term data for each location are already present. In that case, an overall status for such a location could be established and also used for several years. But in large scale studies this problem persists. Therefore, developing regional or river basin reference conditions is more appropriate. For developing these, long-term data are welcome, but not a necessity. The main requirements are several sampling sites per region. These sites will then be assessed based on the same reference. Another benefit of larger scale definitions over site-specific ones is their invariability over shorter timescales (Hawkins *et al.* 2010). Since Slovenia has a very diverse geology and most rivers flow through several regions, a regional approach, based on surface water or, if applicable, hyporheic typology is the best choice.

The chosen reference conditions could also account for the observed results. If the defined reference conditions describe an already heavily polluted system, the difference between the system under study and the reference will be comparatively small. This leads to this system being wrongly classified into "good" categories. Historical data on hyporheic communities does not exist and we simply do not know what a "natural" state would be like. Probably the most adequate method for determining reference conditions is extrapolation from empirical models, based on current attributes. But the main problem with these is calibration, since the values we are interested in are outside the models calibration range. If these models are joined with the other methods, a comparatively accurate description of reference conditions can be achieved. However, community structure cannot be included in these predictions, as knowledge of species ecology and physiology is still lacking. Such models could therefore predict only physical and chemical attributes of "natural" or "near-natural" systems, ignoring community compositions. What these communities would look like under unimpacted conditions, we can only guess.

Machine learning (also referred to as data mining) is the extraction of previously unknown and potentially useful information from data. This is achieved with the use of automated computer programs that examine databases, searching for patterns. If any strong patterns are found, they will likely generalize to make predictions on future data. Nevertheless, as with any generalization, there will be exceptions. Therefore, anything discovered will be inexact. Algorithms need to be robust enough to cope with imperfect data and to extract patterns that are inexact, but useful.

In statistical models over-fitting is a common problem. Over-fitting occurs when the models describes random noise, instead of actual patterns. This generally occurs when the ratio between the number of variables and observations is too high (there are too many attributes relative to the number of observations). An extreme example is when there are as many attributes as there are observations. In such a case, a simple model can learn to perfectly predict the training data by memorizing the data in its entirety. Such a model will completely fail when used on unseen data, as it has not learned to generalize. Robust algorithms can reduce the chance of fitting noise. A generally used method to avoid overfitting in machine learning is pruning of the tree.

5 CONCLUSIONS

Environmental conditions in the interstitial of gravel beds conform to a strong, upriver – downriver gradient. Deviations from this gradient are signified by longer retention times, higher biological oxygen demand, and higher nitrate and ammonia concentrations. Slovene guidelines for chemical water quality determination are based on nitrate concentrations and biological oxygen demand. In upwelling zones nitrate rich groundwater enters the stream bed and concentrations in the hyporheic increase. This is a natural phenomenon and does not indicate pollution.

Our sampling effort sufficed to achieve a good overview over copepod species richness in the hyporheic of Slovene rivers. Approximately 100 samples would be the minimum requirement for a species richness estimate, at river basin scales. For two basins, Drava and Primorje, a larger sampling effort would be needed to attain better estimates. In the third undersampled river basin, Mura, only one species was found in all samples. This indicates a strong dominance of *E. elaphoides* and a much larger sample size would be needed to find any additional species.

Hyporheic copepods in Slovenia have a distinct biodiversity hotspot in the region of the river Idrijca. This coincides with data available from other authors on invertebrate biodiversity in this region. Simpson’s diversity measure, when coupled with Simpson’s evenness, gives a good insight into diversity patterns of hyporheic copepods.

Geographical location (river membership), water chemistry and physical habitat account for most of the variability in species data. A few species (e.g. *Moraria radovnae* and *Canthocamptus staphylinus*) indicate certain habitat preferences, but data are insufficient for any definite conclusions as to their value as bioindicators.

The hyporheic copepod community, combined with the WFD river typology and diversity measures (Simpson’s index and Simpson’s evenness), can be used for water quality assessment. But, the best obtained model, even though it reaches acceptable correlations, is not dependable enough in extreme classes. While the applied river typology is in order, more data are needed from each river type. Furthermore, more sampling sites with low WFD status (classes 4 and 5) have to be included in the data mining process, for the model to reach a truly practical value.

6 SUMMARY

Water quality is affected by several complex factors, and we use numerous variables to describe the status of water bodies and their surrounding area. The term “water quality” is therefore comparatively hard to define. “Water quality” can be regarded as a neutral term that relates to the composition of water as affected by natural processes and human activities. It depends not only on water's chemical, but also its biological, physical and radiological condition. The quality of water is also related to specific use, and is usually measured in terms of constituent concentrations. The level of water quality is based upon the evaluation of measured quantities and parameters, which then are compared to water quality standards, objectives, or criteria. Most often we use the term in relation to the water’s suitability for drinking.

The hyporheic extends vertically and laterally from the river channel. Water in streams and rivers is continuously exchanged between the active channel and subsurface (hyporheic) flowpaths. This interaction can be fast enough that, within several kilometers, water in streams is completely exchanged with interstitial water of the hyporheic zone. Dissolved material and organisms are also continuously exchanged between surface and groundwater. Thus surface water is in close contact with chemically reactive mineral coatings and microbial communities in interstitial (i.e. below surface) waters. This process has the effect of enhancing biogeochemical reactions and downstream water quality.

Up to date, few ecoremediation measures took the surface water – groundwater ecotone into consideration. By developing this index and promoting its use in water quality assessment and environmental impact studies, we intend to expand currently applied methods. Evaluation of the hyporheic zone could be implemented into national water quality assessment methods. An index should be relevant, simple and easily understood by laymen, scientifically justifiable, quantitative, and acceptable in terms of cost.

Most of the measured variables were significantly correlated with distance from the spring. Only nitrate and ammonia concentrations, hydraulic conductivity, and biological oxygen demand were correlated with less than five variables. The presence of a strong upriver – downriver gradient was confirmed by both constrained and unconstrained multivariate ordination methods. The four less often correlated variables (NO_3^-, NH_4^+, biological oxygen demand and hydraulic conductivity) can influence the deviation of sampling sites from this gradient and indicate impacted locations.

We found 49 copepod species, belonging to 23 genera of 6 families. Of these, 11 species were found in only one location (five Harpacticoida and six Cyclopoida species). Three species were found in more than 40 % of locations (*Bryocamptus dacicus, B. zschokkei* and *Elaphoidella elaphoides*) and are believed to be ubiquitous. Three more species (*Diacyclops clandestinus*, *Echinocamptus pilosus* and *Paracamptus schmeili*) were found in approximately a third of all locations. Estimated species richness in overall is around 50 copepod species in the hyporheic. Several methods confirmed this estimate. These results suggest that a sampling effort of minimum 100 samples is necessary to obtain reliable results.

Three indices (Fisher’s alpha, modified Shannon and Simpson’s) indicate two biodiversity hotspots; the first in the central-western, the second in the central-eastern part of Slovenia. The eastern part belongs to the Soča basin (the hotsport is centered on the middle part of the river Idrijca), the western to the Sava basin (the center of the hotspot is around the confluence of the rivers Sava and Savinja). A third, smaller, hotspot is indicated in the southern part of Slovenia, in the middle part of the river Kolpa. The Simpson’s evenness index indicates that even though diversity in the second hotspot is comparatively high, the abundance distribution of species is uneven.

The Simpson diversity index was shown to be the most appropriate diversity measure we used. Out of four tested indices, it was the only one that showed the prevalence of uneven samples. Since Simpson’s index emphasizes dominance of a community; when the reciprocal (1-D) is used, higher values indicate a more even assemblage. Even though a log series model indicated that Fisher’s α is a good descriptor of diversity, it failed to show the patterns Simpson’s index revealed.

Geographical location (river membership) accounted for 38 % of the variability in species data, thus being the single most important factor. Water chemistry and characteristics of the physical habitat explained approximately 40 %, while geological substrate was the least important, explaining 3 %. This suggests that historical factors are the main determinant in

shaping copepod species assemblies. Water chemistry and physical habitat, both of which include human impacts, come second in importance

The hyporheic copepod community, combined with an accurate typology and diversity measures, can be used for water quality assessment. However, the best obtained model, even though it reaches acceptable correlations, is not dependable enough in extreme classes. While the applied river typology is in order, more data are needed from each river type. Furthermore, more sampling sites with low Water Framework Directive (WFD) status (classes 4 and 5) have to be included in the data mining process, for the model to reach a truly practical value. The obtained final model is most accurate in mid-range classes, while being inaccurate at extreme values. The most likely reason is the low number of locations of very low or very good WFD status (classes 1, 4, and 5). We analyzed only two locations with WFD class four; these two locations represent only 2 % of all sampling stations. With the chosen validation method (10-fold cross validation), the chances of these extreme cases to be included in the learning process are comparatively low. If these cases were not included in the learning process, the algorithm could not learn how to classify them. Furthermore, none of the locations were of class 5 ("Bad"). Wrong classifications can therefore be, in cases such as this, expected. In general, a 1:5 ratio between the number of attributes and instances, is recommended for data mining purposes. We have achieved this requirement when all data is pooled, but considering each river type, the amount of quality data is lacking.

7 REFERENCES

Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on automatic control, 19(6): 716 – 723.

Alden A.S., Munster C.L. 1997. Assessment of river-floodplain aquifer interactions. Environmental and engineering geoscience, 3: 537 – 548.

Allan J.D. 1995. Stream ecology. Structure and function of running waters. Kluwer Academic Publishers, Dordrecht: 388 pp.

Aller C, Aller J.Y. 1992. Meiofauna and solute transport in marine muds. Limnology and oceanography, 37: 1018 – 1033.

Alley W.M., Healy R.W., LaBaugh J.W., Reilly T.E. 2002. Flow and storage in groundwater systems. Science, 296: 1985 – 1990.

Anderson J.E. 1991. A conceptual framework for evaluating and quantifying naturalness. Conservation biology, 5(3): 347 – 352.

Anderson M.J. 2001. A new method for non-parametric multivariate analysis of variance. Austral ecology, 26(1): 32 – 46.

Angradi T., Hood R. 1998. An application of the plaster dissolution method for quantifying water velocity in the shallow hyporheic zone of an Appalachian stream system. Freshwater biology, 39: 301 – 315.

ARSO 2008. Podatki o kakovosti voda – 2008. Reke. ARSO, MOP RS, Ljubljana: 449 pp.

Bailey R.C., Norris H.C., Reynoldson T.B. 2003. Bioassessment of freshwater ecosystems: Using the reference condition approach. Springer-Verlag, Berlin: 184 pp.

Bengtsson B.E. 1978. Use of a harpacticoid copepod in toxicity tests. Marine pollution bulletin, 16: 238 – 241.

Bonacci O. 1987. Karst hydrology. Springer-Verlag, Berlin: 184 pp.

Bork J., Berkhoff S.E., Bork S., Hahn H.J. 2009. Using subsurface metazoan fauna to indicate groundwater–surface water interactions in the Nakdong River floodplain, South Korea. Hydrogeology Journal, 17: 61 - 75.

Boulton A. 2000. The subsurface macrofauna. In: Jones J.B., Mulholland P.J. (eds.). Streams and groundwaters. Academic Press, San Diego: 337 – 361.

Boulton A.J., Stanley E.H. 1995. Hyporheic processes during flooding and drying in a Sonoran Desert stream. II: Faunal dynamics. Archiv für Hydrobiologie, 134: 27 - 52.

Boxshall G.A., Defaye D. 2008. Global diversity of copepod (Crustacea: Copepoda) in freshwater. Hydrobiologia, 595: 195 – 207.

Brancelj A. 2001. Male of *Moraria radovnae* Brancelj, 1988 (Copepoda: Crustacea), and notes on endemic and rare copepod species from Slovenia and neighboring countries. Hydrobiologia, 453/454: 513 – 524.

Brancelj A., Šiško M., Rejec Brancelj I., Jeran Z., Jaćimović R. 2000a. Effect of land use and fish stocking on a mountain lake - evidence from the sediment. Periodicum Biologorum, 102: 259 – 268.

Brancelj A., Šiško M., Lami A., Appleby P., Livingstone D.M., Rejec Brancelj I., Ogrin D. 2000b. Changes in the trophic level of an Alpine lake, Jezero v Ledvici (NW Slovenia), induced by earthquakes and climate change. Journal of Limnology (Supplement 1), 1: 29 – 42.

Brancelj A., Šiško M., Muri G., Appleby P., Lami A., Shilland E., Rose N. L., Kamenik C., Brooks S.J., Dearing J.A. 2002. Lake Jezero v Ledvici (NW Slovenia) - changes in sediment records over the last two centuries. Journal of Paleolimnology, 28: 47 – 58.

Bretschko G. 1991. Bedsediments, groundwater and stream limnology. Verhandlungen der internationalen Vereinigung für theoretische und angewandte Limnologie, 24: 1957 – 1960.

Brown R.J., Rundle S.D., Hutchinson T.H., Williams T.D., Jones M.B. 2005. A microplate freshwater copepod bioassay for evaluating acute and chronic effects of chemicals. Environmental toxicological chemistry, 24(6): 1528 – 1531.

Brunke M., Gonser T. 1997. The ecological significance of exchange processes between rivers and groundwater. Freshwater biology, 37: 1 – 33.

Burton S.M., Rundle S.D., Jones M.B. 2002. Evaluation of the meiobenthic copepod *Bryocamptus zschokkei* (Schmeil) as an ecologically relevant test organism in lotic freshwaters. Journal of aquatic ecosystem stress and recovery, 9: 185 – 191.

Chao A. 1984. Non-parametric estimation of the number of classes in a population. Scandinavian Journal of Statistics, 11: 265 – 270.

Chao A. 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics, 43: 783 – 791.

Chao A., Hwang W.-H., Chen Y.-C., Kuo C.-Y. 2000. Estimating the number of shared species in two communities. Statistica Sinica, 10: 227 – 246.

Chazdon R.L., Colwell R.K., Denslow J.S., Guariguata M.R. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. In: Dallmeier F., Comiskey J. A. (eds.). Forest biodiversity research, monitoring and modeling: Conceptual background and Old World case studies. Parthenon Publishing, Paris: 671 pp.

Clarke K.R. 1993. Non-parametric multivariate analysis of changes in community structure. Australian journal of ecology, 18: 117 – 143.

Clarke K.R., Warwick R.M. 2001. Change in marine communities: An approach to statistical analysis and interpretation. Plymouth marine laboratory, Plymouth: 154 pp.

Clarke K.R., Somerfield P.J., Chapman M.G. 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages. Journal of Experimental Marine Biology and Ecology, 330: 55 – 80.

Coddington J.A., Young L.H., Coyle F.A. 1996. Estimating spider species richness in a southern Appalachian cove hardwood forest. Journal of arachnology, 24: 111 – 128.

Colwell R.K., Codington J.A. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical transactions of the Royal Society B: Biological sciences, 345: 101 – 118.

Colwell R.K., Mao C.X., Chang J. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. Ecology, 85: 2717 – 2727.

Connell J.H. 1978. Diversity in tropical rain forests and coral reefs. Science, 199: 1302 – 1310.

Costanza R., Norton B.G., Haskell B.D. (eds.) 1992. Ecosystem health: New goals for environmental management. Island Press, Washington D.C.: 279 pp.

Craft J.A., Stanford J.A., Pusch M. 2002. Microbial respiration within a floodplain aquifer of a large gravel-bed river. Freshwater Biology, 47: 251 - 261.

Culver D.C., Deharveng L., Bedos A., Lewis J.J., Madden M., Reddell J.R., Sket B., Trontelj P., White D. 2006. The mid-latitude biodiversity ridge in terrestrial cave fauna. Ecography, 29: 120 – 128.

Danielopol D.L., Rouch R., Pospisil P., Torreiter P., Mösslacher F. 2009. Ecotonal animal assemblages; their interest for groundwater studies. In: Gibert J., Mathieu J., Fournier F. (eds.). Groundwater/surface water ecotones: Biological and hydrological interactions and management options: 246 pp.

Davis J.C. 2002. Statistics and data analysis in geology. John Wiley & Sons, New York: 656 pp.

Davies S.P., Jackson S.D. 2006. The biological-condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications, 16(4): 1251 – 1266.

Debeljak M. 2002. Characteristics of ecological energetic in virgin and managed forest. Doctoral dissertation. University of Ljubljana, Ljubljana: 156 pp.

Delamare Debouttville C. 1954. Recherches sur l'écologie et la réparation du mystacocaride *Derocheilocaris remanei* Delamare et Chappuis, en Méditerranée. Vie et Milieu, 4: 321 – 380.

Di Lorenzo T., Stoch F., Fiasca B., Gattone E., De Laurentiis P., Ranalli F., Galassi D.M.P. 2005. Environmental quality of deep groundwater in the Lessinian Massif (Italy): signposts for sustainability. In: Gibert J. (ed.). World subterranean biodiversity, Proceedings of an international symposium, University Claude Bernard, Lyon: 115 – 125.

Directive 2000/60/EC. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (Water framework directive). Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0060:EN:NOT

Einsle U. 1993. Crustacea: Copepoda: Calanoida und Cyclopoida. Süßwasserfauna von Mitteleuropa 8/4-1. Gustav Fischer Verlag, Stuttgart: 209 pp.

Fath B.D., Patten B.C., Choi J.S. 2001. Complementarity of ecological goal functions. Journal of theoretical biology, 208: 493 – 506.

Fath B.D., Jørgensen S.E., Patten B.C., Straškraba M. 2004. Ecosystem growth and development. Biosystems, 77: 213 – 228.

Findlay S., Sobczak W.V. 2000. Microbial communities in hyporheic sediments. In: Jones J.B., Mulholland P.J. (eds.). Streams and groundwaters. Academic Press, San Diego: 287 – 306.

Freckman D.W., Blackburn T.H., Bussaard L., Hutchings P., Palmer M.A., Snelgrove P.V.R. 1997. Linking biodiversity and ecosystem functioning of soils and sediments. Ambio, 26: 556 – 562.

Galassi D.M.P 2001. Groundwater copepods: diversity patterns over ecological and evolutionary scales. Hydrobiologa, 453/454: 227 – 253.

Galassi D.M.P., Huys R., Reid J.W. 2009. Diversity, ecology and evolution of groundwater copepods. Freshwater Biology, 54: 691 - 708.

Gams I. 2003. Kras v Sloveniji v prostoru in času. ZRC SAZU, Ljubljana: 516 pp.

Gibert J., Deharveng 2002. Subterranean ecosystems: a truncated functional biodiversity. Bioscience, 52: 473 – 481.

Gibert J., Dole-Olivier M.-J., Marmonier P., Vervier P. 1990. Surface water/groundwater ecotones. In: Naiman R. J. & H. Decamps (eds), Ecology and Management of Aquatic-Terrestial Ecotones. The Parthenon Publishing Group, Carnforth**:** 199 - 225.

Gibert J., Stanford J.A., Dole-Olivier M.-J., Ward J.V. 1994. Basic attributes of groundwater ecosystems and prospects for research. In: Gibert J., D. L. Danielopol & J. A. Stanford (eds), Groundwater Ecology. Academic Press, New York**:** 7 - 40.

Giere O. 1993. Meiobenthology: the microscopic fauna in aquatic sediments. Springer-Verlag, Berlin: 328 pp.

Griebler C., Danielopol D.L., Gibert J., Nachtnebel H.P., Notenboom J. (eds.) 2001. Groundwater ecology. A tool for management of water resources. Office for official publications in the European communities, Luxembourg: 415 pp.

Hahn H.J. 2006. The GW-Fauna-Index: a first approach to a quantitative ecological assessment of groundwater habitats. Limnologica, 36: 119 – 137

Hakenkamp C.C., Palmer M.A. 2000. The ecology of hyporheic meiofauna fauna. In: Jones J. B. & P. J. Mulholland (eds.), Streams and ground waters. Academic Press, London**:** 307 - 333.

Hammer O., Harper D.A.T., Ryan P.D. 2001. PAST: Paleontological Statistics Software package for education and data analysis. Paleontologica Electronica, 4(1): 9 pp.

Harvey J.W., Wagner B.J., Bencala K.E. 1996. Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange. Water resources research 32: 2441 – 2451.

Haunschmid R., Jagsch A. (eds.) 2006. Erstellung einer Fischbasierten Typologie Österreichischer Fließgewässer sowie einer Bewertungsmethode des fischökologischen Zustandes gemäß EU-Wasserrahmenrichtlinie. Bundesamt für Wasserwirtschaft, Mondsee: 105 pp.

Hawkins C.P., Olson J.R., Hill R.A. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society, 29(1): 312 – 343.

Hoehn E. 2001. Exchange processes between rivers and groundwaters – the hydrological and geochemical approach. In: Griebler C., Danielopol D.L., Gibert J., Nachtnebel H.P., Notenboom J. (eds.). Groundwater ecology. A tool for management of water resources. Office for official publications in the European communities, Luxembourg: 415 pp.

Hughes R.M., Larsen D.P., Omernik J.M. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management, 10: 629 – 635.

Hughes R.M., Kaufmann P.R., Herlihy A.T., Kincaid T.M., Reynolds L., Larsen D.P. 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences, 55: 1618 – 1631.

Jackson S., Davis W. 1995. Meeting the goal of biological integrity in water-resource programs of the U.S. Environmental Protection Agency. Journal of the North American Benthological Society, 13: 592 – 597.

Janetzky W., Enderle R., Noodt W. 1996. Crustacea: Copepoda: Gelyelloida und Harpacticoida. Süßwasserfauna von Mitteleuropa 8/4-2. Gustav Fischer Verlag, Stuttgart: 227 pp.

Jersabek C.D., Brancelj A., Stoch F., Schabetsberger R. 2001. Distribution and ecology of copepods in mountainous regions of the Eastern Alps. Hydrobiologia, 453/454: 309 – 324.

Jiang J. 2005. Development of a new biotic index to assess freshwater pollution. Environmental pollution, 139: 306 – 317.

Jiang J., Shen Y. 2005. Use of the aquatic protozoa to formulate a community biotic index for an urban water system. Science of the total environment, 346: 99 – 111.

Jørgensen S.E. 2007. Description of aquatic ecosystem's development by eco-exergy and exergy destruction. Ecological modeling, 204: 22 – 28.

Jørgensen S.E., Svirezhev Y.M. 2004. Towards a thermodynamic theory for ecological systems. Elsevier: 366 pp.

Jørgensen S.E., Costanza R., Xu F. (eds.) 2005. Handbook of ecological indicators for assessment of ecosystem health. CRC Press, Florida: 439 pp.

Kaplan L. A., Newbold J.D. 2000. Surface and subsurface dissolved organic carbon. In: Jones J.B., Mulholland P.J. (eds.). Streams and Ground Waters. Academic Press, San Diego: 237 – 258.

Karr J.R. 1981. Assessment of biotic integrity using fish communities. Fisheries, 6(6): 21 – 27.

Karr J.R., Chu E.W. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington: 174 pp.

Krebs C. J. 2001. Ecology: The experimental analysis of distribution and abundance. Benjamin Cummings, San Francisco: 695 pp.

Kurashov E.A. 1996. Distribution and summer diapause of *Canthocamptus staphylinus* (Jurine) (Copepoda: Crustacea) in Lake Ladoga. Hydrobiologia, 320: 191 – 196.

Lande R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communitites. Oikos, 76: 5 – 13.

Lawrence J. 1994. Introduction to Neural Networks: Design, theory and application. California Scientific Software Press: 348 pp.

Legendre P., Anderson M.J. 1999. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. Ecological Monographs, 69(1): 1 – 24.

Legendre P., Legendre L. 1998. Numerical ecology. Developments in environmental modelling 20, Elsevier Science, Amsterdam: 853 pp.

Longino J.T., Coddington J., Colwell R.K. 2002. The ant fauna of a tropical rain forest: estimating species richness three different ways. Ecology, 83: 689 – 702.

Maddock I.P., Petts G.E., Evans E.C., Greenwood M.T. 1995. Assessing river – aquifer interactions within the hyporheic zone. In: Brown A.G. (ed.).Geomorphology and groundwater, Wiley, Chichester: 53 – 74.

Magurran A.E. 2004. Measuring biological diversity. Blackwell Publishing, Malden: 256 pp.

Malard F., Reygrobellet J.L., Mathieu J., Lafont M. 1994. The use of invertebrate communities to describe groundwater flow and contaminant transport in a fractured rock aquifer. Archiv für Hydrobiologie, 131: 93 – 110.

Malcolm I.A., Youngson A.F., Soulsby C. 2003. Survival of salmonid eggs in a degraded gravel-bed stream: effects of groundwater – surface water interactions. River research and applications, 19: 303 – 316.

Marmonier P., Creuzé Des Châtelliers M. 1991. Effects of spates on interstitial assemblages of the Upper Rhone River. Importance of spatial heterogeneity. Hydrobiologia, 210: 243 - 251.

Marmonier P., Creuzé Des Châtelliers M. 1992. Biogeography of the benthic and interstitial living ostracods (Crustacea) of the Rhone River (France). Joumal of Biogeography, 19: 693 - 704.

Marmonier P., Ward J.V., Danielopol D.L. 2009. Round table 2: Biodiversity in groundwater/surface water ecotones: central questions. In: Gibert J., Mathieu J., Fournier F. (eds.). Groundwater-surface water ecotones: Biological and hydrological interactions and management options. Cambridge university press, Cambridge: 231 – 235.

Marshall M.C., Hall Jr R.O. 2004. Hyporheic invertebrates affect N cycling and respiration in stream sediment microcosms. Journal of the North American Benthological Society, 23: 416 - 428.

Martín M.A., Rey J.M. 2000. On the role of Shannon's entropy as a measure of heterogeneity. Geoderma, 98: 1 – 3.

Matzke D., Hahn H.J., Ramstöck A., Rother K. 2005. Bewertung von Altlasten im Grundwasser anhand der Meiofaunageselschaften: erste Ergebnisse. Grundwasser, 1: 25 – 34.

McArdle B.H., Anderson M.J. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology, 82(1): 290 – 297.

Meyer J.L.1997. Stream health: incorporating the human dimension to advance stream ecology. Journal of the North American benthological society, 16(2): 439 – 447.

MOP 2008. Tipi površinskih voda za vrednotenje ekološkega stanja (ekološki tipi površinskih voda). MOP RS, Ljubljana: 10 pp.

MOP 2009a. Metodologija vrednotenja ekološkega stanja rek z bentoškimi nevretenčarji. MOP RS, Ljubljana: 74 pp.

MOP 2009b. Metodologija vzorčenja in laboratorijske obdelave vzorcev za vrednotenje ekološkega stanja rek z bentoškimi nevretenčarji. MOP RS, Ljubljana: 35 pp.

MOP 2009c. Vrednotenje ekološkega stanja površinskih voda s splošnimi fizikalno-kemijskimi elementi. MOP RS, Ljubljana: 11 pp.

Mösslacher F. 1998. Subsurface dwelling crustaceans as indicators of hydrological conditions, oxygen concentrations and sediment structure in an alluvial aquifer. Internationale Revue der Hydrobiologie, 83: 349 – 364.

Mösslacher F., Griebler C., Notenboom J. 2001. Biomonitoring of groundwater systems: methods, applications and possible indicators among groundwater biota. In: Griebler C., Danielopol D.D., Gibert J., Nachtnebel H.P., Notenboom J. (eds). Groundwater ecology. A tool for management of water resources. European commission, UK: 173 – 182.

Mullholland P.J., Marzolf E.R., Webster J.R., Hart D.R., Hendricks P.J. 1997. Evidence that hyporheic zones increase heterotrophic metabolism and phosphorous uptake in forest streams. Limnology and oceanography, 42: 443 – 451.

Newman M.C. 2009. Fundamentals of ecotoxicology. CRC Press: 571 pp.

O'Connor J.S., Dewling R.T. 1986. Indices of marine degradation: their utility. Environmental management, 10: 335 – 343.

Odum E.P. 1969. Fundamentals of ecology. 3rd edition. W.B. Saunders company, London: 575 pp.

Orghidan T. 2010. A new habitat of subsurface waters: the hyporheic biotope. Fundamental and Applied Limnology/Archive fur Hydrobiologie, 176: 291 - 302.

Orlando-Bonaca M., Lipej L., Orfanidis S. 2008. Benthic macrophytes as a tool for delineating, monitoring and assessing ecological status: The case of Slovenian coastal waters. Marine pollution bulletin, 56: 666 – 676.

Osborne J.W. 2010. Improving your data transformations: Applying the Box-Cox transformation. Practical Assessment, research & evaluation, 15(12): 1 – 9.

Paran F., Malard F., Mathieu J., Lafont M., Galassi D.M.P., Marmonier P. 2005. Distribution of groundwater invertebrates along an environmental gradient in a shallow water-table aquifer. In: Gibert J. (ed.). World subterranean biodiversity, Proceedings of an international symposium, University Claude Bernard, Lyon: 99 – 105.

Pospisil P. 1994. The groundwater fauna of a Danube Aquifer in the "Lobau" Wetland in Vienna, Austria. In: Gibert J., Danielopol D.L., Stanford J.A. (eds). Groundwater Ecology. Academic Press, New York**:** 347 - 366.

Pusch M. 1996. The metabolism of organic matter in the hyporheic zone of a mountain stream, and its spatial distribution. Hydrobiologia, 323: 107 – 118.

Quinlan R. 1992. C4.5: Programs for machine learning. Morgan Kaufmann: 302 pp.

REFCOND 2003. Guidance on establishing reference conditions and ecological status class boundaries for inland surface waters. Produced by Working group 2.31 Reference conditions for inland waters (REFCOND). Common implementation strategy of the water framework directive, European Commission: 86 pp.

Rouch R., Danielopol D.L. 1997. Species richness of microcrustacea in subterranean freshwater habitats. Comparative analysis and aproximate evaluation. Internationale Revue der gesamten Hydrobiologie, 82: 121 - 145.

Rundle S.D., Ramsay P. 1997. Microcrustacean communities from two physiographically contrating regions of Britain. Journal of Biogeography, 24: 101 – 111.

Runkel R.L., McKnight D.M., Bencala K.E., Chapra S.C.1996. Reactive sollute transport in streams. 2. Simulation of a pH modification experiment. Water resources research, 32: 419 – 430.

Ryder R.A. 1990. Ecosystem health, a human perception: definition, detection and the dichotomous key. Journal of Great Lakes Research, 16(4): 619 – 624.
Sampat P. 2000. Deep trouble: The hidden threat of groundwater pollution. Worldwatch paper 154, Worldwatch institute, Washington D.C.: 55 pp.
Schmid-Araya J.M. 2009. Temporal and spatial dynamics of meiofaunal assemblages in the hyporheic interstitial of a gravel stream. In: Gibert J., Mathieu J., Fournier F. (eds.). Groundwater/surface water ecotones: Biological and hydrological interactions and management options: 246 pp.
Schmidt S.I., Hahn H.J., Hatton T., Humphreys W.F. 2007. Do faunal assemblages reflect the exchange intensity in groundwater zones? Hydrobiologia, 583: 1 – 19.
Sear D.A., Armitage P.D., Dawson F.H. 1999. Groundwater dominated rivers. Hydrological processes, 13: 255 – 276.
Siligardi M., Bernabi S., Cappelletti C., Ciutti S., Dallafior V., Dalmiglio A., Fabiani C., Mancini L., Monauni C., Pozzi S., Scardi M., Tancioni L., Zennaro B. 2010. Lake shorezone functionality index (SFI): A tool for the definition of ecological quality as indicated by Directive 2000/60/CE. Agenzia Provinziale Protezione Ambiente (APPA): 73 pp.
Soberón J., Llorente J. 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology, 7**:** 480 – 488.
Southwood T.R.E., Henderson P.A. 2000. Ecological methods. Blackwell science, Oxford: 592 pp.
Stanford J.A., Ward J.V. 1993. An ecosystem perspective of alluvial rivers: connectivity and the hyporheic corridor. Journal of the North American Benthological Society, 12: 48 – 60.
Stanford J.A., Ward J.V., Ellis B.K. 1994. Ecology of the alluvial aquifers of the Flathead River, Montana. In: Gibert J., Danielopol D.L., Stanford J.A. (eds). Groundwater Ecology. Academic Press, New York**:** 367 – 390.
Steube C., Richter S., Griebler C. 2009. First attempts towards an integrative concept for the ecological assessment of groundwater ecosystems. Hydrogeology Journal, 17: 23 – 35.
Stoch F. 2000. New and little known Parastenocaris (Copepoda, Harpacticoida, Parastenocariidae) from cave waters in Northeastern Italy. Bolletino del Museo Civico di Storia Naturalle di Verona, Botanica Zoologia, 24: 223 – 234.
Stoch F., Artheau M., Brancelj A., Galassi D.M.P., Malard P. 2009. Biodiversity indicators in European ground waters: towards a predictive model of stygobiotic species richness. Freshwater Biology, 54: 745 - 755.
Stoddard J.L., Larsen D.P., Hawkins C.P., Johnson R.K., Norris R.H. 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. Ecological Applications, 16(4): 1267 – 1276.
Strayer D.L., May S.E., Nielsen P., Wolfheim W., Hausam S. 1997. Oxygen, organic matter and sediment granulometry as controls on hyporheic animal communities. Archiv für Hydrobiologie, 140: 131 – 144.
UL RS 2002. Uredba o kemijskem stanju površinskih voda. Ur.l. RS 11/2002: 811 – 822.
Ugland, K. I., J. S. Gray, & K. E. Ellingsen. 2003. The species-accumulation curve and estimation of species richness. Journal of Animal Ecology, 72**:** 888 – 897.
UNEP 1996. Groundwater: A Threatened Resource. Nairobi: 36 pp.
Urbanič G., Toman M.J. 2003. Varstvo celinskih voda. Študentska založba, Ljubljana: 94 pp.
USEPA 1972. Clean water act. 33 U.S.C. § 1251 et seq. (2002). Retrieved from http://epw.senate.gov/water.pdf
USEPA 2000a. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. EPA-600/R-99/064. U.S. EPA Office of Water, Washington D.C.: 192 pp.
USEPA 2000b. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00-002. U.S. EPA Office of water, Washington D.C.: 253 pp.
USEPA 2000c. Mid-Atlantic highlands streams assessment. EPA/903/R-00/015. U.S. EPA Office of water, Washington D.C.: 74 pp.
Van Damme D., Heip C., Willems K.A. 1984. Influence of pollution on the harpacticoid copepods of two North Sea estuaries. Hydrobiologia, 112: 143 – 160.
Vanek V. 2009. Heterogeneity of groundwater-surface water ecotones. In: Gibert J., Mathieu J., Fournier F. (eds.). Groundwater-surface water ecotones: Biological and hydrological

interactions and management options. Cambridge university press, Cambridge: 151 – 161.

Vervier P., Gibert J., Marmonier P., Dole-Olivier M.-J. 1992. A perspective on the permeability of the surface freshwater-groundwater ecotone. Journal of the North American Benthological Society, 11: 93 - 102.

Wallin M., Wiederholm T., Johnson R.K. 2003. Final guidance on establishing reference conditions and ecological status class boundaries for inland surface waters. EU Common Implementation Strategy (CIS) for the Water Framework Directive: 93 pp.

Wang Y., Witten I.H. 1997. Induction of model trees for predicting continuous classes. Proceedings of the poster papers of the European Conference on machine learning, Prague: 128 – 137.

Ward J.V., Bretschko G., Brunke M., Danielopol D., Gibert J., Gonser T., Hildrew A.G. 1998. The boundaries of river systems: the metazoan perspective. Freshwater biology, 40: 531 – 569.

Ward J.V., Palmer M.A. 1994. Distribution patterns of interstitial freshwater meiofauna over a range of spatial scales, with emphasis on alluvial river-aquifer systems. Hydrobiologia, 287(1): 147 – 156.

Williams D.D., Fulthorpe R.R. 2003. Using invertebrate and microbial communities to assess the condition of the hyporheic zone of a river subject to 80 years of contamination by chlorobenzenes. Canadian journal of zoology, 81: 789 – 802.

Williams D.D., Hynes H.B.N. 1974. The occurrence of benthos deep in the substratum of a stream. Freshwater Biology, 4: 233 - 256.

Witten I.H., Frank E. 2005. Data mining: Practical machine learning tools and techniques. Second edition. Morgan Kaufmann: 560 pp.

Wood P.J., Hannah D.M., Sadler J.P. (eds.) 2007. Hydroecology and ecohydrology: past, present and future. Wiley & Sons, West Sussex: 436 pp.

Wondzell S.M., Swanson F.J. 1996. Seasonal and storm dynamics of the hyporheic zone of a 4th-order mountain stream. I: Hydrologic processes. Journal of the North American benthological society, 15: 3 – 19.

Younger P.L. 2007. Groundwater in the environment: An introduction. Blackwell publishing, Malden: 317 pp.

Zucker L.A., Brown L.C., Volgstadt C.E., Antosch L.M. 1998. Nonpoint source assessment: User’s guide to Ohio State waters. Bulletin 873, The Ohio State University: 21 pp.

ACKNOWLEDGEMENTS

I would like to address special thanks to my supervisor prof. dr. Anton Brancelj, who introduced me to this research topic, taught and helped me with copepod species determination, and guided me to the final version of this thesis.

Several colleagues assisted with field sampling and laboratory work: Marcus from Finland, dr. Nataša Mori, Simona Prust, Anja Mavrič, Luka Mrzelj, Barbara Bric, dr. Simon Lukančič, dr. Tadej Mezek, Špela Ambrožič and Tina Jaklič. Without their help, this thesis would not have been possible.

Dr. Maarten de Groot, doc. dr. Al Vrezec, Špela Ambrožič, dr. Irena Bertoncelj and prof. dr. Marko Debeljak assisted me with in-depth discussions about my work and specially statistical analyses and data mining procedures.

I am also grateful to my colleagues from the Department of freshwater and terrestrial ecosystems research at the National institute of biology in Ljubljana, Slovenia.

I would also like to thank the members of my thesis committee, prof. dr. Diana Maria Paola Galassi, doc dr. Barbara Čenčur Curk and doc. dr. Tatjana Tišler, for giving me new ideas and substantial arguments when I presented my thesis theme.

This thesis was funded by the Slovenian research agency, through the Young researchers program.

Last but not least, I would like to express my deepest gratitude to my family. Without their support and endless patience, my student years, especially as a PhD student, would have been much more difficult.

I would like to express s sincere “Thank you” to all those mentioned above. If I forgot anyone, it was not on purpose.

APPENDICES

Appendix A

List of all environmental variables with their units of measurement. BOD: biological oxygen demand; COD: chemical oxygen demand; TOC: total organic carbon; Nr.: number of samples per gravelbed.

Group		Variable	Unit	Nr.
Water chemistry (WC)		Cl^-	mgl^{-1}	1
		NO_2^-	mgl^{-1}	1
		NO_3^-	mgl^{-1}	1
		SO_4^{2-}	mgl^{-1}	1
		NH_4^+	mgl^{-1}	1
		Ca^{2+}	mgl^{-1}	1
		Mg^{2+}	mgl^{-1}	1
		Na^+	mgl^{-1}	1
		K^+	mgl^{-1}	1
		TN	mgl^{-1}	1
		TP	μgl^{-1}	1
		Alkalinity	$mMoll^{-1}$	1
		Ca/Mg	/	1
		T	°C	3
		O_2	mgl^{-1}	3
		O_2 saturation	%	3
		Conductivity	μScm^{-1}	3
		BOD	mgl^{-1}	1
		COD	mgl^{-1}	1
		TOC	mgl^{-1}	1
Physical habitat (PH)	Gravel	1 cm – 3 cm	fuzzy coded	3
	Gravel	3 cm – 6 cm	fuzzy coded	3
	Gravel	6 cm – 10 cm	fuzzy coded	3
	Gravel	>10cm	fuzzy coded	3
		Sediment	ml	3
		Hcond	s	3
		Depth	cm	3
		Width	cm	3
Geology and Geography (GG)		Clastites	fuzzy coded	1
		Gravel deposits	fuzzy coded	1
		Clay-gravel deposits	fuzzy coded	1
		Tertiary sediments	fuzzy coded	1
		Tonalite	fuzzy coded	1
		Carbonates	fuzzy coded	1
		River memberhip	dummy coded	1
		River_km	km	1
		masl	m	1

Appendix B

Gravel size distribution and geological characteristics of all sampling sites. Class membership is fuzzy coded. CL: clastites; GD: gravel deposits; CGD: clay-gravel deposits; TS: tertiary sediments; TL: tonalite; CA: carbonates.

Location	1 cm - 3 cm	3 cm - 6 cm	6 cm - 10 cm	>10 cm	CL	GD	CGD	TS	TL	CA
Sava1	0,25	0,25	0,25	0,25	0	1	0	0	0	0
Sava2	0,33	0,33	0,33	0	0	1	0	0	0	0
Sava3	1	0	0	0	0	1	0	0	0	0
Sava4	0,25	0,25	0,25	0,25	0	1	0	0	0	0
Sava5	0,33	0,33	0,33	0	0	1	0	0	0	0
Sava7	0	0,33	0,33	0,33	0	1	0	0	0	0
Sava8	0,33	0,33	0,33	0	0	1	0	0	0	0
Sava9	0,25	0,25	0,25	0,25	0	1	0	0	0	0
Sava10	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Sava11	0	0,33	0,33	0,33	0	0	0	0	0	1
Sava12	0,25	0,25	0,25	0,25	0	1	0	0	0	0
Sava13	0	0,33	0,33	0,33	0	1	0	0	0	0
SDol1	0	0,33	0,33	0,33	0	1	0	0	0	0
SDol2	0	0,33	0,33	0,33	0	1	0	0	0	0
SDol3	0,5	0,5	0	0	0	1	0	0	0	0
SDol4	0,5	0,5	0	0	0	1	0	0	0	0
SBoh1	0	0,33	0,33	0,33	0	1	0	0	0	0
SBoh2	0	0,33	0,33	0,33	0	0	0	0	0	1
SBoh3	0	0,33	0,33	0,33	0	1	0	0	0	0
KBis1	0,25	0,25	0,25	0,25	0	0	0	0	0	1
KBis2	0,5	0,5	0	0	0	0	0	1	0	0
KBis3	0,25	0,25	0,25	0,25	0	1	0	0	0	0
KBis4	0,5	0,5	0	0	0	1	0	0	0	0
Paka1	0,33	0,33	0,33	0	0	0	0	0	0	1
Paka2	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Krka1	0,33	0,33	0,33	0	0	1	0	0	0	0
Sora1	0,25	0,25	0,25	0,25	0	1	0	0	0	0
SSora1	0,5	0,5	0	0	0	0	0	0	0	1
SSora2	0,33	0,33	0,33	0	0	0	0	0	0	1
PSora1	0,25	0,25	0,25	0,25	0	0	0	0	0	1
PSora2	0,33	0,33	0,33	0	0	0,5	0	0	0	0,5
PSora3	1	0	0	0	0	0,5	0	0	0	0,5
Savi1	0,33	0,33	0,33	0	0	0	0	0	0	1
Savi2	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Savi3	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Savi4	0,25	0,25	0,25	0,25	0,5	0	0,5	0	0	0
Savi5	0,33	0,33	0,33	0	0	0	1	0	0	0

Savi6	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Savi7	0,25	0,25	0,25	0,25	0	1	0	0	0	0
Savi8	0,25	0,25	0,25	0,25	0,5	0	0,5	0	0	0
Savi9	0,25	0,25	0,25	0,25	0	0,5	0	0,5	0	0
Bela1	0,33	0,33	0,33	0	0	0	0	0	0	1
Iska1	0,33	0,33	0,33	0	0	0	0	0	0	1
Iska2	0,25	0,25	0,25	0,25	0	0,5	0	0	0	0,5
Kolpa1	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Kolpa2	0,33	0,33	0,33	0	1	0	0	0	0	0
Kolpa3	0,33	0,33	0,33	0	0	0	0	0	0	1
Kolpa4	0,5	0,5	0	0	0	0	0	0	0	1
Kolpa5	0,5	0,5	0	0	0	0	0	0	0	1
Mura1	0,5	0,5	0	0	0	1	0	0	0	0
Mura2	0,33	0,33	0,33	0	0	0	1	0	0	0
Mura3	0,33	0,33	0,33	0	0	0,5	0,5	0	0	0
Mura4	0,33	0,33	0,33	0	0	1	0	0	0	0
Mura5	0,33	0,33	0,33	0	0	1	0	0	0	0
Drava1	0,25	0,25	0,25	0,25	0	0	0	1	0	0
Drava2	0,25	0,25	0,25	0,25	0	0	1	0	0	0
Drava3	0,33	0,33	0,33	0	0	1	0	0	0	0
Dravi1	0,5	0,5	0	0	0	0	0	1	0	0
Dravi2	0,25	0,25	0,25	0,25	0	0	1	0	0	0
Misl1	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Oplo1	0,25	0,25	0,25	0,25	0	0	0	0	1	0
Soca1	0,25	0,25	0,25	0,25	0	0,5	0	0	0	0,5
Soca2	0,25	0,25	0,25	0,25	0	0,5	0	0	0	0,5
Soca3	0,25	0,25	0,25	0,25	0	0,5	0	0	0	0,5
Soca4	0,25	0,25	0,25	0,25	0,5	0	0	0	0	0,5
Soca5	0,33	0,33	0,33	0	0,33	0,33	0	0	0	0,33
Soca6	0,33	0,33	0,33	0	0	0,5	0	0	0	0,5
Soca7	0,33	0,33	0,33	0	0,5	0,5	0	0	0	0
Soca8	0,33	0,33	0,33	0	1	0	0	0	0	0
Soca9	0,5	0,5	0	0	0,33	0,33	0	0	0	0,33
Idri1	0,5	0,5	0	0	0	0	0	0	0	1
Idri2	0,33	0,33	0,33	0	0	0	0	0	0	1
Idri3	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Idri4	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Idri5	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Idri6	0,33	0,33	0,33	0	0	0	0	0	0	1
Idri7	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Idri8	0,25	0,25	0,25	0,25	0,5	0	0	0	0	0,5
Belca1	0,5	0,5	0	0	0	0	0	0	0	1
Vipa1	0,33	0,33	0,33	0	0,5	0	0	0	0	0,5
Vipa2	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Vipa3	0,33	0,33	0,33	0	1	0	0	0	0	0
Vipa4	0,5	0,5	0	0	1	0	0	0	0	0

Vipa5	0,5	0,5	0	0	0	0	0	0	0	1
Nadi1	0,33	0,33	0,33	0	1	0	0	0	0	0
Nadi2	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Reka1	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Reka2	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Reka3	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Reka4	0,25	0,25	0,25	0,25	0	0	0	0	0	1
Riza1	0,25	0,25	0,25	0,25	1	0	0	0	0	0
Drago1	0,33	0,33	0,33	0	1	0	0	0	0	0

Appendix C

Calculated reference values for all river types. RV: reference values for each river type (median of all samples in a river type); Good: values for good status; Bad: values for bad status.

River type	Type number	NO_3^- (mgl^{-1})			BOD (mgl^{-1})		
		RV	Good	Bad	RV	Good	Bad
SI-Al	1	2.4	2.7	6.5	1.2	1.9	5.4
KB_AL-D	2	2.1	3.4	8.2	1.0	1.6	4.4
PA_Hrib_D	6	3.3	3.6	8.6	1.2	1.9	5.4
VR1-AL-SA	9	5.8	4.0	8.2	1.2	1.9	4.4
VR3-DN-SA	11	4.2	5.5	7.4	1.4	2.4	5.4
KB_AL-D_u	3	1.9	6.7	8.0	1.9	4.4	5.4
PN_ZalVpliv	8	6.3	15.1	17.9	1.9	4.4	5.4
VR5-Ko	12	3.2	3.6	6.8	1.2	1.9	5.4
VR9-Mu-RavDr	13	3.3	9.2	12.5	1.4	2.4	5.4
KB_AL-J	4	2.4	4.4	10.6	1.2	1.9	5.4
VR2-So	10	3.6	5.5	10.5	1.2	1.9	5.4
PA_Hrib_J	5	3.9	4.9	9.3	1.2	1.9	5.4
SM-Hrib-brez	7	2.6	6.9	13.1	1.2	1.9	5.4

Appendix D

Results of chemical analyses and data on physical habitat from all sampled locations. TN: total nitrogen (mg l^{-1}); TP: total phosphorous (mg l^{-1}); Alka: alkalinity (mMol l^{-1}); T: temperature (°C); Sat: oxygen saturation (%); sediment: amount of sediment in a 10 l sample (ml); Hcond: hydraulic conductivity (sec); COD: chemical oxygen demand (mg l^{-1}); TOC: total organic carbon (mg l^{-1}); BOD: biological oxygen demand (mg l^{-1}).

	NO_3^- (mgl^{-1})	SO_4^{2-} (mgl^{-1})	NH_4^+ (mgl^{-1})	Na^+ (mgl^{-1})	K^+ (mgl^{-1})	TN (mgl^{-1})	TP (μgl^{-1})	Alka ($mMoll^{-1}$)	T (°C)	O_2 (mgl^{-1})	Sat (%)	Cond ($\mu Scml^{-1}$)	Sediment (ml)	Hcond (s)	COD (mgl^{-1})	TOC (mgl^{-1})	BOD (mgl^{-1})
Sava1	2,11	7,712	0,08	1,85	0,83	1,37	21	2,69	15,9	3,7	38,7	423	12,2	152	5	1	1,3
Sava2	2,05	6,943	0,00	1,82	0,43	1,08	18	2,25	15,1	8,6	88,7	286	4,3	126	5	1	1,5
Sava3	1,73	14,129	0,06	2,93	0,67	1,05	9	3,87	15,8	1,8	18,7	451	15,7	171	5	1	1,3
Sava4	6,22	9,716	0,00	2,90	0,66	2,44	10	2,93	18,4	7,2	78,7	386	12,8	326	5	1	1,3
Sava5	5,20	11,761	0,11	4,41	1,14	3,02	127	3,12	17,3	3,9	42,7	396	25,7	151	5	1	1,3
Sava7	5,44	9,053	0,51	5,18	1,12	3,37	110	3,33	19,2	5,3	60,0	415	74	203	6,5	1,7	2,6
Sava8	3,07	10,638	0,00	3,69	1,22	0,70	21	5,06	18,3	2,5	26,0	473	31,3	391	6,5	1,7	2,1
Sava9	3,39	14,272	0,02	4,01	1,17	0,69	13	3,73	22,5	5,3	60,7	395	8,3	148	6,5	1,7	2,4
Sava10	4,80	26,937	0,00	8,54	3,09	2,17	62	4,83	18,1	1,1	12,0	519	36,7	185	7	1,7	2,7
Sava11	12,39	59,281	0,00	32,47	1,25	4,05	24	5,81	19,8	2	23,3	588	22,3	359	7,3	1,8	3
Sava12	5,83	15,332	0,06	5,60	1,20	2,54	27	3,88	21,8	2,9	33,3	412	21,7	186	7,3	2,1	3
Sava13	4,34	18,075	0,00	6,56	1,40	1,21	29	3,73	23,1	1,7	19,0	513	43,3	110	7,5	2,2	3,6
SDol1	2,08	5,736	0,00	1,61	0,34	1,23	1	2,82	12,1	9,4	96,3	277	1,5	566	5	0,6	1,3
SDol2	2,12	6,223	0,03	1,63	0,66	1,66	8	2,74	13,1	9,7	98,0	281	7,7	158	5	0,6	1,3
SDol3	2,37	9,769	0,00	1,96	0,27	1,63	1	2,67	10,3	6,8	70,7	192	4,5	446	5	0,8	1,3
SDol4	3,30	10,012	0,00	2,30	0,47	1,15	9	2,89	14,4	5,5	55,0	358	15,7	160	5	1,1	1,3
SBoh1	2,06	1,777	0,06	0,56	0,42	1,09	1	1,80	16,9	7,1	78,3	223	13,7	201	5	0,9	1,3
SBoh2	2,30	1,91	0,02	0,74	0,40	1,12	1	2,24	16,8	7	77,7	254	8,7	353	5	0,9	1,3
SBoh3	3,31	2,35	0,00	0,95	0,37	1,30	8	2,29	17,6	9,7	107,0	246	15	170	5	1	1,3
KBis1	2,14	1,871	0,00	0,33	0,24	0,70	1	1,74	6,1	15,1	135,3	169	39	373	5	0,5	1,3
KBis2	2,96	4,018	0,00	0,86	0,23	1,10	1	2,44	9	14,6	131,0	241	17,7	99	5	0,5	1,3
KBis3	3,24	5,706	0,00	1,66	0,49	1,04	1	1,39	11,5	12	115,0	303	58,3	195	7	1,5	1,5
KBis4	21,99	12,454	0,16	13,16	3,28	5,47	517	5,07	13,2	1,6	15,3	563	133	142	6	1,4	6
Paka1	2,72	10,358	0,00	3,56	1,28	1,19	8	4,08	9,2	10,9	99,0	379	13,7	150	10	1,4	1,7
Paka2	4,99	73,774	0,00	20,92	7,14	1,82	117	3,25	12,2	2,2	21,3	545	27,7	424	7,1	2,5	7
Krka1	4,12	1,799	0,42	3,88	1,04	0,84	37	4,65	15,2	1,5	15,3	406	29,7	274	5,8	1,9	1,3
Sora1	6,37	11,188	0,00	4,58	1,01	1,59	13	3,35	12,6	8,5	83,7	420	85,5	157	5	0,8	2,3
SSora1	2,73	7,41	0,00	2,95	0,61	0,88	1	2,69	9,5	14,1	134,0	276	9	155	5	0,7	1,3
SSora2	5,91	10,322	0,00	5,63	0,94	0,99	1	3,23	12,6	9,8	95,3	338	48	170	5	0,7	1,3
PSora1	2,97	22,43	0,00	2,20	0,81	0,96	1	4,12	10,3	8,7	83,3	443	66,3	84	5	0,9	1,3
PSora2	3,92	12,036	0,00	2,63	0,98	1,04	1	3,36	12,5	4,9	48,7	380	38,7	125	5	0,9	1,3
PSora3	3,13	10,442	0,00	2,65	0,76	0,95	1	3,36	12,5	5,3	51,0	348	33,7	114	5	0,9	1,3
Savi1	1,81	9,647	1,10	1,00	0,10	1,00	1	2,56	8,4	11,3	104,7	256	14,7	132	5	0,9	1,3
Savi2	2,09	9,663	0,00	1,21	0,39	0,98	1	3,19	8,8	11,1	103,3	303	13,3	231	5	0,9	1,3
Savi3	1,86	11,034	0,01	1,52	0,48	0,64	1	3,47	9,6	10,8	101,3	326	11,3	283	5	1	1,3

Savi4	1,85	9,448	0,00	2,16	0,57	1,14	1	3,11	12	7,4	72,3	288	29,3	186	5	1,1	1,3
Savi5	1,44	11,379	0,00	2,93	0,60	0,68	1	3,84	14	1,9	19,3	369	12,3	139	5	1,1	1,8
Savi6	4,13	9,691	0,00	2,40	0,57	1,47	1	3,36	12,6	7	68,3	318	14	167	5	1,3	1,9
Savi7	51,42	23,169	0,00	4,45	1,65	10,46	9	6,18	14	5,6	55,3	700	16,7	117	5	1,3	6
Savi8	4,72	27	6,13	2,00	1,92	1,37	9	3,93	13,2	1,6	15,3	480	34	233	7	1,6	4
Savi9	7,58	57,418	0,00	20,71	3,44	2,19	17	4,26	13,5	1,8	17,3	532	69,7	231	11	1,9	5
Bela1	2,42	4,346	0,00	0,33	0,10	1,04	1	2,62	8,2	11,3	103,3	235	55	656	5	0,9	1,3
Iska1	3,86	6,871	0,00	1,15	0,50	1,26	13	4,78	9,6	11,2	101,0	403	25	720	5	0,5	0,9
Iska2	2,79	6,956	0,00	0,88	0,58	1,29	7	4,42	12,4	11,5	108,5	388	17	640	5	0,5	1,3
Kolpa1	3,27	42,014	0,00	3,77	0,34	0,99	1	2,80	12,9	7,9	75,7	440	26	186	5	1,2	1,3
Kolpa2	3,23	9,073	0,00	1,83	0,26	1,22	1	2,98	13,2	6,3	61,3	350	14,3	98	5	1,2	1,3
Kolpa3	1,58	7,854	0,00	2,29	0,73	0,80	8	3,41	13,6	4,8	46,3	393	4,7	188	5	1,2	1,3
Kolpa4	3,09	5,911	0,00	2,33	0,34	0,96	85	3,40	13,8	2,2	21,0	325	10	337	5	1,4	2
Kolpa5	3,64	8,023	0,00	2,24	0,62	0,39	13	3,65	14,2	3	31,3	357	62,3	183	5	1,4	3
Mura1	1,27	20,183	0,00	8,07	1,54	1,72	10	2,71	16,7	2,6	28,0	310	72,7	178	23	2,1	2,5
Mura2	3,05	23,534	0,00	7,60	2,66	0,64	30	2,89	18,3	2,1	22,3	365	16,7	673	17	2,1	3,4
Mura3	2,10	25,6	0,00	8,57	2,14	0,82	16	2,25	18,4	1,5	16,3	314	15,7	526	20	2,3	3,5
Mura4	3,38	27,538	0,00	9,46	2,13	1,46	53	2,32	19,3	1,1	12,0	337	22,3	497	21	2,7	3,6
Mura5	3,47	26,51	0,00	9,20	2,65	1,44	54	2,39	18,1	0,8	8,0	337	57,7	415	17	2	4,1
Drava1	2,00	21,05	0,00	4,39	1,25	0,98	8	1,79	17,8	4,6	49,0	286	54,7	84	5	1,1	3,3
Drava2	20,61	29,78	0,00	8,04	1,82	0,98	12	4,73	18,6	3	32,3	605	80,7	142	5	1,1	3,3
Drava3	3,27	41,317	0,07	4,02	1,70	4,49	8	2,78	19,5	7,5	83,0	410	70,7	362	5	1,1	3,6
Dravi1	1,35	12,068	0,00	7,15	1,78	1,14	131	2,34	11	1,8	17,0	282	11,7	198	11	2,9	1,7
Dravi2	5,33	11,645	0,00	7,38	2,68	2,91	59	2,81	11,2	4	37,3	317	31,7	83	11	2,9	1,7
Misl1	2,32	12,351	0,00	2,31	0,74	0,80	8	0,92	7,5	10,7	98,3	119	33,3	264	6	1,4	0,7
Oplo1	0,84	7,417	0,00	2,25	0,63	0,76	35	0,77	8,4	9,4	86,7	86	13	284	0,0	0,0	0,7
Soca1	1,85	1,935	0,00	0,46	0,11	0,45	7	2,14	10,3	11,2	107,0	202	9	85	5	0,7	0,5
Soca2	1,91	1,929	0,00	0,50	0,15	0,51	7	2,45	10	11,1	103,7	226	10,7	112	5	0,6	0,8
Soca3	1,95	3,031	0,00	0,79	0,23	0,50	7	2,26	13,7	11,4	114,0	219	3,7	115	5	0,6	0,8
Soca4	3,57	3,524	0,06	0,98	0,30	1,04	7	3,07	18,3	7,1	78,0	297	21	131	5	0,7	2
Soca5	2,39	3,232	0,09	1,69	0,61	0,65	7	2,65	12,8	7,9	74,0	279	14,3	348	5	0,8	2,1
Soca6	2,36	3,288	0,02	1,04	0,35	0,53	7	3,47	15	4,7	67,7	335	14,7	150	5	0,9	2,1
Soca7	2,71	7,528	0,00	1,11	0,34	0,71	9	2,72	13	10,6	101,0	267	6	91	5	1	3,2
Soca8	1,65	8,754	0,08	1,45	0,37	0,53	10	2,85	13,3	6,8	65,7	307	23,7	136	5	1,2	3,5
Soca9	3,90	8,715	0,00	1,75	0,75	1,02	10	2,99	13,8	9,9	97,3	279	37,7	101	5	1,2	3,5
Idri1	3,32	4,841	0,00	1,34	0,62	1,14	1	3,42	10,3	7,8	72,0	319	17	98	5	1,2	0,5
Idri2	4,63	4,253	0,00	1,66	0,11	1,54	1	4,06	11,2	4,7	43,3	387	38	273	5	1,3	0,5
Idri3	5,90	34,405	0,00	1,93	0,40	2,00	10	3,83	13,9	1,9	19,3	496	49,8	223	5	1,4	1,2
Idri4	1,36	15,304	0,00	1,20	0,48	0,49	18	4,48	10,5	6,6	61,0	413	13,7	152	5	1,4	1,4
Idri5	5,08	26,143	0,00	1,77	0,49	1,17	37	3,59	11,5	3,6	33,7	412	51,3	106	5	1,5	2,6
Idri6	3,64	24,568	0,00	1,97	0,44	1,22	27	3,64	12	6,4	59,7	393	11	122	5	1,6	2,7
Idri7	3,15	26,814	0,00	2,15	0,53	1,02	28	3,56	11,9	7,8	73,0	371	12,7	107	5	1,7	2,3
Idri8	3,59	25,951	0,00	1,99	0,54	1,12	26	3,81	13,5	5,3	51,0	373	42,3	154	5	1,7	2,4
Belca1	4,10	3,038	0,00	0,74	0,44	1,12	1	3,17	10,2	7,9	74,0	321	31,7	89	5	1,2	0,5
Vipa1	6,38	5,24	0,00	1,89	0,44	2,21	1	3,39	11,7	7,6	50,7	343	11,3	228	5	1,2	1,1
Vipa2	5,92	7,347	0,00	4,18	0,68	1,61	1	3,48	12,7	7,8	118,0	348	28	194	5	1,2	1,2

Vipa3	5,56	7,239	0,00	4,20	0,86	2,20	1	3,45	12,5	11,6	109,7	349	39,7	109	5	1,3	1,1
Vipa4	4,61	8,702	0,00	4,04	0,71	1,83	7	3,45	14,3	1,5	15,0	458	14,3	86	6	1,4	2
Vipa5	7,18	8,55	0,00	4,80	1,19	2,16	51	2,19	15	9,4	94,0	360	39,7	99	6	1,5	3
Nadi1	3,48	5,58	0,00	2,41	0,49	1,06	16	3,06	13,9	9,8	96,3	278	7,2	117	5	0,5	0,5
Nadi2	2,92	5,233	0,00	2,33	0,44	0,98	32	3,31	14	10	98,3	296	10,3	80	5	1	0,5
Reka1	1,12	11,631	0,00	6,30	0,55	0,73	1	2,76	9,2	5	45,7	278	98	539	5	2,4	0,5
Reka2	0,53	10,192	0,00	4,88	0,73	0,69	1	2,61	9,8	9,5	87,7	291	93	108	5	2,4	0,5
Reka3	2,91	7,769	0,00	6,18	1,52	0,91	9	2,85	10,7	5,8	55,3	293	83,7	116	10	3,2	4,7
Reka4	3,13	7,045	0,00	4,53	1,51	0,35	8	3,38	11,3	1,3	12,3	321	57,3	396	10	3,2	10
Riza1	4,89	10,827	0,00	4,67	0,92	1,59	9	4,15	13,7	6,8	66,7	427	64,3	855	5	1,3	2,2
Drago1	3,60	53,569	0,00	12,84	1,94	0,64	1	4,93	15,4	2,8	28,0	515	11,7	68	5	1	2,1

Appendix E

Diversity indices for sampled locations. Nr.: consecutive number of sampling site; Type: river type according to Slovene national typology; Code: river type code; WFD: surface water quality according to WFD guidelines; Masl: meters above sea level; Ind.: number of individuals 10 l^{-1}; Taxa_S: number of taxa; α: Fisher's alpha diversity; H': Shannon-Wiener diversity; $e^{H'}$: modified Shannon-Wiener diversity; (1-D): Simpson's index (D); $E_{1/D}$: Simpson's measure of evenness; D_{Mg}: Margalef's diversity index; d: Berger-Parker index

Nr.	Location	River	Type	Code	WFD	masl	Ind.	Taxa_S	α	H'	eH'	(1-D)	$E_{1/D}$	D_{Mg}	d
1	Sava1	Sava	PA_Hrib_D	6	3	406	5	3	1,94	1,20	3,31	0,58	0,30	1,47	0,61
2	Sava2	Sava	PA_Hrib_D	6	3	383	1060	3	0,54	0,83	2,30	0,41	0,34	0,46	0,75
3	Sava3	Sava	VR1-AL-SA	9	1	366	249	5	1,15	1,18	3,24	0,53	0,23	1,00	0,67
4	Sava4	Sava	VR1-AL-SA	9	1	303	56	5	1,89	1,34	3,83	0,65	0,29	1,52	0,46
5	Sava5	Sava	VR1-AL-SA	9	1	291	67	5	1,47	1,52	4,55	0,74	0,65	1,12	0,37
6	Sava7	Sava	VR3-DN-SA	11	2	263	1372	12	1,96	1,86	6,41	0,77	0,34	1,64	0,41
7	Sava8	Sava	VR3-DN-SA	11	2	251	16	1	0,50	0,66	1,93	0,47	0,94	0,30	0,63
8	Sava9	Sava	VR3-DN-SA	11	1	237	10	3	1,65	1,32	3,74	0,71	0,87	1,06	0,41
9	Sava10	Sava	VR3-DN-SA	11	1	210	56	7	2,11	1,88	6,53	0,84	0,90	1,49	0,18
10	Sava11	Sava	VR1-AL-SA	9	4	197	9	3	1,22	1,01	2,75	0,60	0,84	0,78	0,54
11	Sava12	Sava	VR1-AL-SA	9	2	150	618	7	0,90	0,73	2,07	0,34	0,22	0,78	0,80
12	Sava13	Sava	VR1-AL-SA	9	3	146	86	4	1,27	1,56	4,77	0,77	0,73	1,01	0,28
13	SDol1	Sava Dolinka	KB_AL-D	2	2	727	0	0	0	0	1,00	0	0,00	0	0
14	SDol2	Sava Dolinka	KB_AL-D	2	3	660	9	2	0,58	0,38	1,46	0,18	0,41	0,43	0,90
15	SDol3	Sava Dolinka	KB_AL-D	2	2	583	19	3	0,98	1,00	2,71	0,61	0,84	0,67	0,50
16	SDol4	Sava Dolinka	PA_Hrib_D	6	2	418	20	1	1,04	1,40	4,07	0,70	0,56	0,86	0,42
17	SBoh1	Sava Bohinjka	PA_Hrib_D	6	3	530	82	4	1,29	1,31	3,70	0,67	0,50	1,02	0,47
18	SBoh2	Sava Bohinjka	PA_Hrib_D	6	3	481	342	3	1,01	1,43	4,18	0,73	0,52	0,86	0,35
19	SBoh3	Sava Bohinjka	PA_Hrib_D	6	2	417	962	4	0,66	1,24	3,47	0,69	0,64	0,56	0,40
20	KBis1	Kamniška Bistrica	KB_AL-D	2	2	586	41	1	0,83	1,29	3,63	0,70	0,84	0,65	0,40
21	KBis2	Kamniška Bistrica	PA_Hrib_D	6	3	403	172	5	1,01	1,16	3,18	0,63	0,39	0,86	0,44
22	KBis3	Kamniška Bistrica	PA_Hrib_D	6	3	333	189	5	0,95	1,09	2,97	0,55	0,32	0,82	0,62
23	KBis4	Kamniška Bistrica	PA_Hrib_D	6	3	269	20	1	0,33	0,46	1,59	0,29	0,70	0,20	0,83
24	Paka1	Paka	SI-Al	1	3	452	145	4	0,63	0,35	1,41	0,15	0,24	0,53	0,92
25	Paka2	Paka	PN_ZalVpliv	8	3	342	0	0	0,32	0,00	1,00	0,00	1,00	0,00	1,00
26	Krka1	Krka	VR1-AL-SA	9	1	145	24	2	0,51	0,55	1,74	0,36	0,79	0,31	0,76
27	Sora1	Sora	PA_Hrib_D	6	2	324	65	4	1,07	1,46	4,31	0,74	0,77	0,85	0,39
28	SSora1	Selška Sora	PA_Hrib_D	6	2	604	39	4	1,41	1,10	3,00	0,51	0,29	1,13	0,69
29	SSora2	Selška Sora	PA_Hrib_D	6	1	363	235	5	0,82	0,70	2,02	0,33	0,25	0,70	0,81
30	PSora1	Poljanska Sora	PA_Hrib_D	6	2	493	190	3	0,87	1,51	4,53	0,76	0,83	0,71	0,35
31	PSora2	Poljanska Sora	PA_Hrib_D	6	3	378	209	6	1,22	1,30	3,67	0,58	0,30	1,04	0,62
32	PSora3	Poljanska Sora	PA_Hrib_D	6	1	335	211	5	1,11	1,09	2,98	0,51	0,25	0,95	0,68
33	Savi1	Savinja	KB_AL-D_u	3	3	718	390	2	0,59	1,20	3,31	0,65	0,71	0,48	0,52

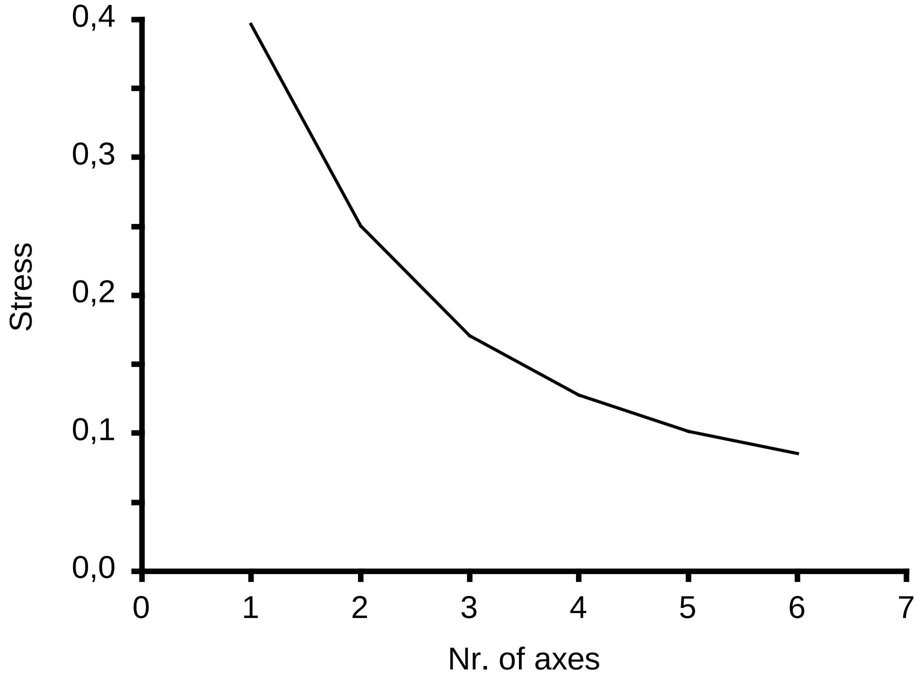
34	Savi2	Savinja	KB_AL-D_u	3	2	638	45	2	1,40	1,60	4,96	0,77	0,72	1,09	0,33
35	Savi3	Savinja	KB_AL-D_u	3	1	526	0	0	0,20	0,00	1,00	0,00	1,00	0,00	1,00
36	Savi4	Savinja	KB_AL-D_u	3	2	426	68	2	1,11	1,36	3,91	0,71	0,68	0,87	0,41
37	Savi5	Savinja	KB_AL-D_u	3	3	364	0	0	0,75	0,90	2,47	0,56	0,76	0,54	0,50
38	Savi6	Savinja	KB_AL-D_u	3	1	320	200	4	1,14	1,60	4,93	0,77	0,61	0,95	0,32
39	Savi7	Savinja	PN_ZalVpliv	8	3	276	65	4	1,62	1,52	4,56	0,72	0,45	1,29	0,41
40	Savi8	Savinja	PN_ZalVpliv	8	2	229	112	6	2,07	1,87	6,49	0,80	0,56	1,58	0,36
41	Savi9	Savinja	PN_ZalVpliv	8	3	213	25	2	1,34	1,42	4,15	0,72	0,72	1,00	0,42
42	Bela1	Bela	KB_AL-D_u	3	3	719	42	5	1,83	1,35	3,86	0,59	0,31	1,41	0,61
43	Iska1	Iška	PA_Hrib_D	6	3	356	0	0	0	0	1,00	0	0,00	0	0
44	Iska2	Iška	PA_Hrib_D	6	2	314	51	1	0,33	0,67	1,95	0,48	0,95	0,20	0,61
45	Kolpa1	Kolpa	PA_Hrib_D	6	1	280	73	4	1,05	0,91	2,49	0,44	0,30	0,86	0,73
46	Kolpa2	Kolpa	VR5-Ko	12	3	213	25	4	2,04	1,80	6,03	0,79	0,52	1,57	0,33
47	Kolpa3	Kolpa	VR5-Ko	12	3	186	53	4	1,63	1,54	4,67	0,73	0,53	1,26	0,40
48	Kolpa4	Kolpa	VR5-Ko	12	2	158	0	0	0	0	1,00	0	0,00	0	0
49	Kolpa5	Kolpa	VR5-Ko	12	3	141	8	1	0,93	1,05	2,85	0,55	0,55	0,71	0,64
50	Mura1	Mura	VR9-Mu-RavDr	13	4	224	0	0	0	0	1,00	0	0,00	0	0
51	Mura2	Mura	VR9-Mu-RavDr	13	1	201	3	1	0,53	0,00	1,00	0,00	1,00	0,00	1,00
52	Mura3	Mura	VR9-Mu-RavDr	13	3	191	16	1	0,24	0,00	1,00	0,00	1,00	0,00	1,00
53	Mura4	Mura	VR9-Mu-RavDr	13	2	180	52	1	0,17	0,00	1,00	0,00	1,00	0,00	1,00
54	Mura5	Mura	VR9-Mu-RavDr	13	1	169	6	1	0,34	0,00	1,00	0,00	1,00	0,00	1,00
55	Drava1	Drava	VR9-Mu-RavDr	13	2	254	31	3	0,81	0,87	2,39	0,45	0,45	0,63	0,72
56	Drava2	Drava	VR9-Mu-RavDr	13	2	242	46	1	0,60	0,95	2,59	0,58	0,79	0,45	0,53
57	Drava3	Drava	VR9-Mu-RavDr	13	1	232	0	0	0	0	1,00	0	0,00	0	0
58	Dravi1*	Dravinja	SI-Al	1	4	275	0	0	0,71	1,02	2,77	0,61	0,86	0,52	0,52
59	Dravi2*	Dravinja	SI-Al	1	4	250	0	0	0,53	0,68	1,97	0,48	0,97	0,32	0,59
60	Misl1	Mislinja	SI-Al	1	2	780	225	3	0,85	1,17	3,23	0,62	0,53	0,70	0,54
61	Oplo1	Oplotnica	SI-Al	1	1	697	61	6	1,39	0,99	2,68	0,47	0,27	1,12	0,70
62	Soca1	Soča	KB_AL-J	4	1	434	24	3	1,49	1,50	4,46	0,75	0,81	1,08	0,37
63	Soca2	Soča	KB_AL-J	4	1	398	0	0	0	0	1,00	0	0,00	0	0
64	Soca3	Soča	KB_AL-J	4	2	358	9	1	0,54	0,90	2,47	0,56	0,76	0,41	0,49
65	Soca4	Soča	KB_AL-J	4	2	323	4	1	0,62	0,64	1,89	0,44	0,90	0,37	0,67
66	Soca5	Soča	KB_AL-J	4	2	183	12	4	1,88	1,36	3,89	0,67	0,61	1,24	0,52
67	Soca6	Soča	KB_AL-J	4	3	171	11	2	0,74	0,76	2,13	0,42	0,58	0,54	0,74
68	Soca7	Soča	VR2-So	10	1	107	272	9	1,56	1,07	2,92	0,57	0,20	1,35	0,51
69	Soca8	Soča	VR2-So	10	3	90	31	5	1,42	1,23	3,41	0,62	0,38	1,14	0,52
70	Soca9	Soča	VR2-So	10	1	78	7	1	0,82	0,96	2,61	0,57	0,78	0,58	0,58
71	Idri1	Idrijca	PA_Hrib_J	5	4	436	780	9	1,56	1,82	6,18	0,81	0,52	1,32	0,30
72	Idri2	Idrijca	PA_Hrib_J	5	3	328	62	4	1,51	1,80	6,07	0,82	0,78	1,19	0,27
73	Idri3	Idrijca	PA_Hrib_J	5	1	304	105	9	2,56	2,07	7,93	0,83	0,48	1,96	0,34
74	Idri4	Idrijca	PA_Hrib_J	5	2	288	776	10	1,93	2,17	8,75	0,86	0,59	1,60	0,25
75	Idri5	Idrijca	PA_Hrib_J	5	2	255	1460	10	1,62	2,06	7,87	0,84	0,52	1,39	0,29
76	Idri6	Idrijca	PA_Hrib_J	5	2	223	94	7	1,35	0,57	1,76	0,22	0,14	1,15	0,88

77	Idri7	Idrijca	VR2-So	10	3	179	33	3	0,84	0,95	2,59	0,50	0,40	0,69	0,67
78	Idri8	Idrijca	VR2-So	10	1	158	8	2	0,86	0,69	2,00	0,50	1,00	0,48	0,50
79	Belca1	Belca	PA_Hrib_J	5	3	429	611	5	1,10	1,55	4,73	0,76	0,59	0,93	0,31
80	Vipa1	Vipava	PA_Hrib_D	6	3	97	293	4	1,24	1,18	3,26	0,56	0,32	1,02	0,63
81	Vipa2	Vipava	PA_Hrib_D	6	1	77	493	7	1,68	1,60	4,96	0,70	0,33	1,39	0,51
82	Vipa3	Vipava	PA_Hrib_D	6	3	69	417	3	0,78	1,21	3,37	0,65	0,57	0,65	0,45
83	Vipa4	Vipava	PA_Hrib_D	6	3	50	210	2	0,68	0,68	1,96	0,34	0,38	0,54	0,80
84	Vipa5	Vipava	PA_Hrib_D	6	3	36	125	3	0,97	1,12	3,06	0,55	0,44	0,78	0,64
85	Nadi1	Nadiža	PA_Hrib_J	5	3	287	110	6	1,89	1,98	7,24	0,83	0,67	1,48	0,31
86	Nadi2	Nadiža	PA_Hrib_J	5	3	239	38	5	1,41	1,16	3,19	0,58	0,40	1,09	0,60
87	Reka1	Reka	SM-Hrib-brez	7	2	516	198	5	1,10	1,19	3,30	0,60	0,41	0,90	0,58
88	Reka2	Reka	SM-Hrib-brez	7	2	446	62	4	1,22	1,35	3,85	0,65	0,47	0,98	0,54
89	Reka3	Reka	SM-Hrib-brez	7	4	369	12	2	0,90	1,08	2,95	0,59	0,62	0,69	0,56
90	Reka4	Reka	SM-Hrib-brez	7	2	331	7	1	0,30	0,00	1,00	0,00	1,00	0,00	1,00
91	Riza1	Rižana	PA_Hrib_D	6	4	24	222	4	0,83	1,17	3,21	0,64	0,56	0,69	0,47
92	Drago1	Dragonja	PA_Hrib_D	6	3	78	3	1	1,05	0,69	2,00	0,50	1,00	0,56	0,50

* Only juveniles were found.

Appendix F

Stress diagram of non-metric multidimensional scaling.



Shepard diagram.