Martin Schletterer[1]*, Leopold Füreder[1], Vyacheslav V. Kuzovlev[2],
Yuri N. Zhenikov[2], Juan Francisco Fuentes-Perez[3], Jeffrey A. Tuhtan[3]

[1] University of Innsbruck, Institute of Ecology; Innsbruck, Austria;
* **Corresponding author;** e-mail: martin@schletterer.co.at
[2] Tver State Technical University, Chair of Nature Management and Ecology; Tver, Russia
[3] Tallinn University of Technology, Centre for Biorobotics; Tallinn, Estonia

CLASSIFICATION OF BENTHIC BIOCENOSES OF THE LOWLAND RIVER TUDOVKA (TVER REGION, RUSSIA) USING COMMUNITY FEATURES

ABSTRACT. Within the joint Russian-Austrian monitoring programme "REFCOND_VOLGA (2006 – 20XX)", monitoring sites were established in the headwaters of the Volga (Tver Region). River Tudovka, a right tributary to the Volga River, was included within this monitoring programme as its catchment is partly protected and has only few anthropogenic activities. The monitoring activities include physico-chemical and hydraulic parameters as well as biota with a focus is on benthic organisms (diatoms and macrozoobenthos). In this work, the longitudinal patterns in community structure are classified in the lowland river Tudovka using a novel feature-based approach taken from signal processing theory. The method first clusters field sampling data into longitudinal classes (upper, middle, lower course). Community features based on the relative frequency of individual species occurring per class are then generated. We apply both generative and discriminative classification methods. The application of generative methods provides data models which predict the probability of a new sample to belong to an existing class. In contrast, discriminative approaches search for differences between classes and allocate new data accordingly. Leveraging both methods allows for the creation of stable classifications. On this basis we show how the community features can be used to predict the longitudinal class. The community features approach also allows for a robust cross-comparison of investigation reaches over time. In cases where suitable long-term data set are available, predictive models using this approach can also be developed.

KEYWORDS: lowland river, monitoring, LTERM, diatoms, macrozoobenthos, nearest neighbour (k-NN) algorithm

INTRODUCTION

Go-East: Intact Eastern European rivers like the headwaters of Volga River enable the definition of reference conditions (RC) for lowland rivers, according to the EU-Water Framework Directive [Schletterer et al. 2014]. In this context the joint Russian-Austrian monitoring programme "REFCOND_VOLGA (2006 – 20XX)", with monitoring sites in the headwaters of the Volga (Tver Region), was set up [Schletterer et al., 2016]. The reference condition approach became an important tool in aquatic assessments and therefore the identification of reference sites is an important goal [REFCOND, 2003]. However, to understand processes in lentic waters continuous assessments and analyses are needed on a long term perspective, to understand natural variability at reference sites. Thus long term ecological monitoring is of great importance in aquatic environments [Schletterer et al., 2016].

In this paper the longitudinal patterns in community structure are classified in the lowland river Tudovka using a novel feature-based approach taken from signal processing theory. We applied the k-NN algorithm – which comes from machine learning – for the analyses of long term data, in order to structure the dataset and reveal distribution patterns. We also summarize the results obtained within the first 5 years (2006–2010), with a focus on the macroinvertebrate as well as diatom communities, including an overview of the interannual variation in community metrics.

MATERIALS AND METHODS
Study area

The Tudovka River is a right hand tributary of the Volga. Its length is 106 km, with a mean altitude of 250 m and a gradient of 0.87 ‰. The coordinates of the source are 56°26′N 33°04′E and the mouth into the Volga is located at 56°25′N 33°49′E. The catchment exceeds over 1126 km² (Fig. 1), with a river network density of 0.69 km / km² and it is highly influenced by the surrounding mires (especially in the uppermost reaches). The forest cover (total) amounts to 62%, 12% of the catchment area are bogs and < 1% lakes.

Concerning the hydrological zonation of Tver region, the Tudovka River is included in the Rzhev-Staritskiy hydrological region, which is characterized by highly dissected surface of the catchment, including less water permeable loamy and podzolic soils, lack of lakes in the watershed, as well as representing an average for the Tver region regarding forested and marshy areas [Zagorski, 1967]. The hydrological regime is not influenced, with 68-70% of the annual runoff in spring, 25 % in summer / autumn and 5-6% in winter, i.e. > 50% of the runoff comes from snowmelt. According to the hydrological classification of B.D.Zaykova this is typical for a river of "Eastern European type with high flood in spring, low flows in summer and winter, as well as higher flows in fall". Only short term hydrological studies were carried out in the catchment of Tudovka River, i.e. at the village Priezdovo (near M. Tud) from 1898 till 1899 [Handbook..., 1936] and at the village Molodoi Tud from 1968 till 1970 [Surin, 1976].

The headwater of River Tudovka is located in the transition area of the "Tsentral'no Lesnoy Zapovednik" (Central Forest State Nature Biosphere Reserve), which was established in 1931 to protect "typical forest associations and animals of the central forest region" [Puzachenko et al., 2007]: On the moraine relief of the central part of the Russian plain, "southern taiga" is the predominant ecosystem type. On European scale, the last virgin spruce forests in the southern taiga – since their formation after the last glaciation, unaffected by felling – are found in this area. These forested areas of the reserve are dominated by spruce (Picea abies; 47 %). About 40 % of the forests are formed by birch and aspen; this forest type originated from of serial processes like windfalls or fires. The pine forest, which is characteristic on mires, amounts about 10 %. And along rivers and streams Black-alder forests are situated, that account for about 3 % of the total forest cover.

Also in the middle / lower course of the river, between Redkino and Molodoi Tud, an area of 80 km² ("Molodoitudskii Zakasnik"), is protected by the regional government since 1992, meeting IUCN criteria III (Natural Monument) and IV (Habitat/Species Management Area).
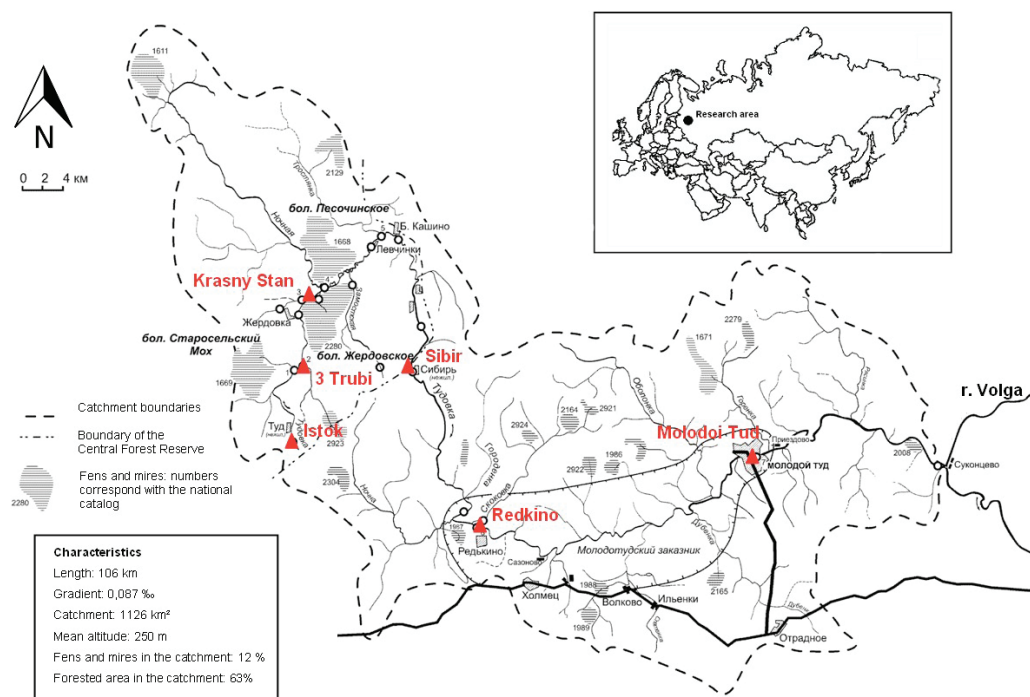
Fig. 1. Basin of river Tudovka: the six sampling sites are indicated

Field work

Within the joint Russian-Austrian monitoring programme "REFCOND_VOLGA (2006 – 20XX)", samples (zoobenthos and benthic diatoms) were taken along the course of the river (Fig. 1) in summer low flow period. Herein we present the date of the first 5 years (2006 – 2010) concerning zoobenthos and 2 years (2006 – 2007) concerning benthic diatoms. In Table 1 a brief characterization of the sampling locations is given.

The zoobenthos samples were taken with a multi-habitat-sampling (MHS) method (HERING et al 2003, modified). A standard frame-net (15x15 cm, 500 μm mesh size) was used and according to the "multihabitat-sampling" method, all available mesohabitats were sampled according to their coverage. As the substrate diversity is quite small in the lowland, we sampled ten squares according to percentages of mesohabitats (e.g. 60 % lithal and 40 % psammal = six samples from lithal and four samples from sand), in total 2250 cm². Afterwards the material was rinsed

Table 1. Characterization of the sampling sites at River Tudovka

| Site	rkm	km from source	Elevation [m]	Basin [km²]	Width* [m]	Depth* [m]
Istok	104	2	260 (270)	13,5	1 – 2.5	0.1 – 0.5
3 Trubi	96	10	238	65	2 – 6	0.2 – 1.0
Krasny Stan	87	19	232	192	2 – 3,5 / 7 – 15**	0.5 – 1.5
Sibir	67	39	223	350	15 – 20 (bridge 23)	0.5 – 1.0
Redkino	51	55	207	410	25 – 30	0.5 – 1.5
Molodoi Tud	20	86	187	926	30 – 35	0.5 – 1.5
Mouth into Volga	0	106	170.9	1126	20 – 25	0.5 – 1.5

* during summer low flow period
**upstream/downstream R. Nochnaja

through a 500 μm net and the invertebrates were preserved. All samples were conserved with ethyl-alcohol: in the field the samples were fixed with 96% ethanol and after sorting, the individuals were stored in 70% ethanol. Determination (the identification keys used are cited in [Schletterer, 2009]) and counting was carried out with a binocular, supported with light microscopy.

which can be represented in a parameter space (Fig. 2 (a)). Including new or additional samples is straightforward, as each new point is classified taking into account its nearest k-neighbors using a distance metric in the sample space (Fig. 2 (b)). A variety of different distance metrics can be used to establish the distance between neighbors.
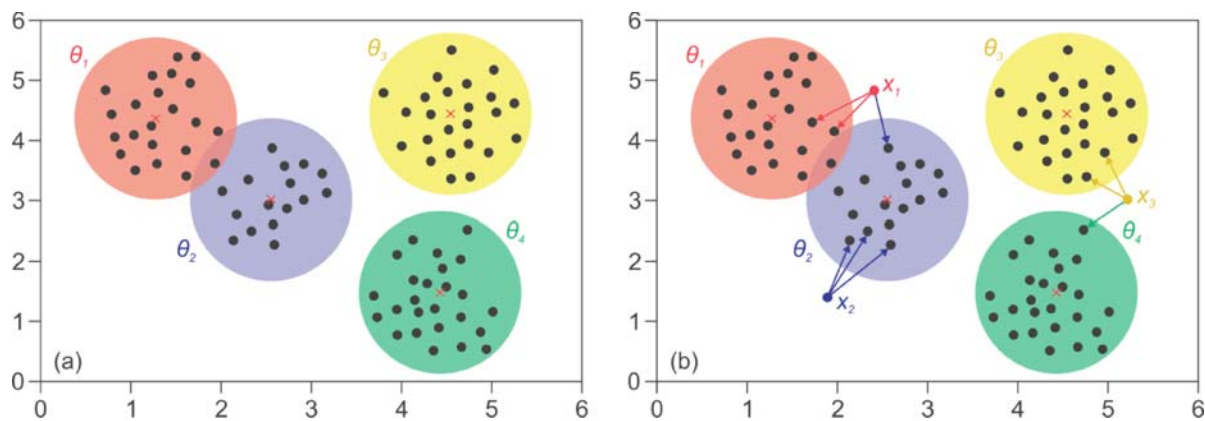
Fig. 2. (a) Distribution of data corresponding to four distinct classes (θ_1, θ_2, θ_3 and θ_4); (b) Classification of new data (x_1, x_2, x_3) using the nearest neighbor approach.

Also benthic diatoms (Bacillariophyceae) were sampled: In most cases at least 15 stones were brushed. In case a location was dominated by fine-sediment (e.g. mud or sand) the uppermost layer of the sediment (2-3 mm) was taken for analyses. The material was stored in ethanol (50 %). The diatom samples were prepared in the laboratory using the H2O2 method (protocol acc. to [Kingston, 1985] cit. in [Schiedele, 1987]). Determination and counting of the diatoms were carried out under the light microscope (1000x, oil immersion), using keys from Krammer & Lange-Bertalot [1986-2004], Prygiel & Coste [2000] and Kelly et al. [2005]. On each mount 300 – 500 valves were counted.

Theoretical background on community features using a species space

The nearest neighbor (NN) algorithm provides a simple nonparametric procedure to perform classification of a collection of sample points (x_i) to different categories (θ_j) [Cover and Hart, 1967]. The method is most commonly applied to perform supervised classification, i.e. when a given category, θ is known *apriori* from a collection of samples

It may be the case that the number and type of classes can be predetermined; however it is often necessary to define the number and division of individual classes using an algorithm such as k-means. This can be accomplished by letting the algorithm determine the definition of the distribution for each category during clustering. Most commonly this is done basing on the minimization of the sum of distances of each class, , to the centroid of its class. The location of the centroid typically becomes more stable with increasing number of samples, thus the k-means classification works best with large data sets.

Here we define an affinity matrix as an interesting particularization and simplification of the NN algorithm which can provide a useful way to study the linkages and interactions between different sampling points. The use of the affinity matrix allows us to uncover relations which may be otherwise difficult to determine using conventional ordination. For instance, if we represent our species distribution, per site as a set of multidimensional complex data points with a possible hidden link between them, we can treat each x as a separate category or class and can thus calculate the distance

to any other x. The resulting matrix contains linkages of all the distances between classes. We can then check to see if the selected distance metric is able to correctly evaluate the hidden connections. If so, then the generated matrix will be indicative of the affinity between classes (the lower the distance, the greater the affinity), and in turn it is generally possible to study the relations between the linkages between classes.

quantity, representative of the number of species sampled (Table 2). The percentage occurrence of each species, across the six study sites thus provides the base input data, where the species space neighborhood has as many dimensions as the total number of observed species.

Although it is possible to consider the relative rankings of the species as dimensions, this

Table 2. Example of data structure. The values are indicative of the frequency in percentage on the species in the sample side.

	Istok	3Trubi	Krasny Stan	Sibir	Redkino	Molodoi Tud
Species 1	0.4	0.0	0.0	0.0	0.0	0.0
Species 2	0.0	11.9	0.0	0.7	1.6	0.0
…						
Species n	0.0	0.6	0.0	0.0	0.3	0.0

Data analyses

As discussed in the previous section, the nearest neighbor algorithm provides a simple nonparametric procedure to perform classification of the sample points (x_i) to a fixed number of categories (θ_j). The method is most commonly applied for supervised classification, i.e. training the algorithm using apriori knowledge of the class membership of the training datasets. In this work, we also tested difference distance metrics on the Tudovka data to determine the most robust NN classification workflow.

1) Convert biological data to „species space" for a single site

2) Calculate inter point distance matrix for a single site

3) Calculate KNN between each site

4) Establish classification patterns between sites with additional variables (e.g. morphology)

In this work, each of the studied classes has a set of different dimensions, i.e. $\theta_1 = (a_1, a_2, …, a_n)$ and $\theta_2 = (b_1, b_2, …, b_n)$. For example, in the case under study the sampling sides can be considered as the individual classes, and each dimension will consist of a single scalar

can mask the relative influence of a given species. Thus the more logical selection for the dimensions will be the frequency of a species.

Once the classes and dimensions are defined, it is then necessary to select a distance metric in order to assess the affinity between the different sites. Considering distance metrics, there are multiple alternatives such as Euclidean distance, Mahalanobis distance, etc. The sum of the distance between two species is used here for simplicity. This distance is known as the Manhattan distance:

$$d(\theta_1, \theta_2) = \left| a_1 - b_1 \right| + \left| a_2 - b_2 \right| + …$$

$$… + \left| a_n - b_n \right| = \sum_{i=1}^{n} \left| a_i - b_i \right|$$

This metric is chosen over the Euclidean distance since considering high dimensionality of the species space vectors, the Manhattan distance usually generates better results, not only because the studied space is not Euclidean [Aggarwal et al., 2001], but also because Euclidean distance is very sensitive to noise [Zimek et al., 2012].

Upon calculating the interclass distance, we then calculate the affinity, defined as:

$$\alpha\left(\theta_1,\theta_2\right) = 1 - \frac{d(\theta_1,\theta_2)}{\max\left[d\left(\theta_1,\theta_i\right)\right]}$$

This affinity normalizes the results of the distances from 0 to 1, being 0 the most "far" site and 1 the most "closely" related. In contrast to a distance matrix, which is symmetric (see e.g. Table 4, where it can be seen that the species space distance between Istok to 3Trubi is identical to the distance between 3Trubi and Istok), the affinity matrix (see e.g. Table 5) is not symmetric and better reflects the variation of the interspecies distances within each individual site, and similarities to each other (see e.g. Fig. 6).

RESULTS

Physico-chemical characterisation of the Tudovka River

The hydrochemical type of water is "hydrogen carbonate magnesium-calcium" with a mineralization of up to 300 mg/ l in the summer low-flow period. The ionic composition of the water in the river at low water can be characterized by the Kurlov-formula [Kurlov 1928 cit. in Zaporozec, 1972) as follows:

$$M_{0.208} \frac{HCO_3 94 SO_4 5 Cl1}{Ca74 Mg21 Na4 K1}$$

at a distance of 3 km from the source;

$$M_{0.197} \frac{HCO_3 96 SO_4 3 Cl1}{Ca74 Mg20 Na5 K1}$$

at a distance of 19 km from the source;

$$M_{0.258} \frac{HCO_3 94 SO_4 4 Cl1 NO_3 1}{Ca65 Mg30 Na4 K1}$$

at a distance of 55 km from the source;

$$M_{0.270} \frac{HCO_3 94 SO_4 4 Cl1 NO_3 1}{Ca67 Mg29 Na3 K1}$$

at a distance of 86 km from the source.

In its headwaters the Tudovka river is highly influenced by the surrounding mires and in theses reaches organic matter increase: the parameter «color Cr-Co" reaches in some periods of the year up to 600 degrees (in tributaries coming from mires sometimes ≥ 1000 degrees), the permanganate oxidation up to 30 mg/l. Downstream of these river reaches groundwater sources contribute to the river, thus mineralization increases during low flow period up to 300 mg/l. The longitudinal variability of physico-chemical parameters along the course of the river during summer low flow period is presented in Table 3 and Fig. 3. Long-term data regarding physico-chemical parameters is summarized in Fig. 4.

ZOOBENTHOS

At the six sampling sites, from 2006 to 2010, a total of 198 macroinvertebrate taxa were identified, mostly at species level; Ostracoda, Hydrachnidae, Chironomidae and Oligochaeta were treated as one taxon each [Schletterer et al., 2010]. Additionally Chironomidae (46 taxa) and Oligochaeta (13 taxa) were determined for selected years. During the summer low flow period the average density of benthic macroinvertebrates was 1,659 ind. m^{-2} (range 815–2,956) at Istok, 1,959 ind. m^{-2} (range 933–2,880) at 3Trubi, 1,281 ind. m^{-2} (range 760–1,689) at Krasny Stan, 3,990 ind.m^{-2} (range 2,591–5,389) at Sibir, 1,669 ind. m^{-2} (range 1,387–1,844) at Redkino and at M. Tud 2,231 ind. m^{-2} (range, 1,107–3,189). Ephemeroptera (41 taxa), Trichopera (33 taxa), Diptera (21 taxa, exkl. Chironomidae), Coleoptera (19 taxa), Gastropoda (17 taxa) and Bivalvia (7 taxa) as well as Plecoptera and Odonata (each group had 11 taxa) were the most diverse groups. The average number of taxa at Istok was 19 (±4), at 3Trubi 18 (±2), at Krasny Stan 27 (±4), at Sibir 30 (±4), at Redkino 22 (±3) and at M. Tud 34 (±7). The total number of taxa increased along the continuum (Fig. 5 b). Abundances

Table 3. Hydrochemical parameters along the monitoring sites during summer low flow period (August, 2012)

Parameter	unit	№1 – Istok	№2 – 3 Trubi	№3 – Krasny Stan	№4 – Redkino	№5 – Molodoi Tud
distance from the source	km	3	10	19	55	86
date and time of sampling		23.08.2012 12:05	23.08.2012 15:35	22.08.2012 13:00	20.08.2012 18:40	20.08.2012 16:45
temperature	ºC	12,4	15,1	12,2	16,5	16,8
pH	ед. pH	7,18	6,95	7,48	8,61	8,31
conductivity	$\mu S\ cm^{-1}$	196	222	233	349	363
colour	º(Cr–Co)	303	117	153	56	36
HCO_3-	mg l^{-1}	98	183	128	195	201
SO_42-	mg l^{-1}	7,7	3,9	3,0	5,9	6,0
$Cl-$	mg l^{-1}	0,1	0,1	0,2	0,9	1,0
$Ca2+$	mg l^{-1}	28,2	30,9	32,6	45,9	46,7
$Mg2+$	mg l^{-1}	8,5	7,3	6,1	10,9	12,2
NH_4+	mg l^{-1}	0,43	0,31	0,33	0,22	0,57
NO_2-	mg l^{-1}	0,02	0,02	0,02	0,03	0,02
NO_3-	mg l^{-1}	0,4	0,4	0,3	2,8	2,7
PO_43-	mg l^{-1}	0,022	0,006	0,029	0,020	0,025
P(total P)	mg l^{-1}	0,139	0,114	0,131	0,107	0,141
Fe (total Fe)	mg l^{-1}	0,39	0,45	0,22	0,10	0,08
Si	mg l^{-1}	3,9	2,5	2,4	2,1	2,8
COD (KMnO4)	mg O l^{-1}	26,0	14,6	16,8	5,8	4,6
BOD5	mg O l^{-1}	6,1	0,5	0,7	0,4	0,6
$Mn2+$	mg l^{-1}	0,02	0,03	0,03	0,02	0,03
alkalinity	mEq l^{-1}	1,6	3,0	2,1	3,2	3,3
hardness	mEq l^{-1}	2,1	2,1	2,1	3,2	3,3
solid residual (mineralisation)	mg l^{-1}	128	134	124	150	174

Martin Schletterer, Leopold Füreder et al.   CLASSIFICATION OF BENTHIC BIOCENOSES...
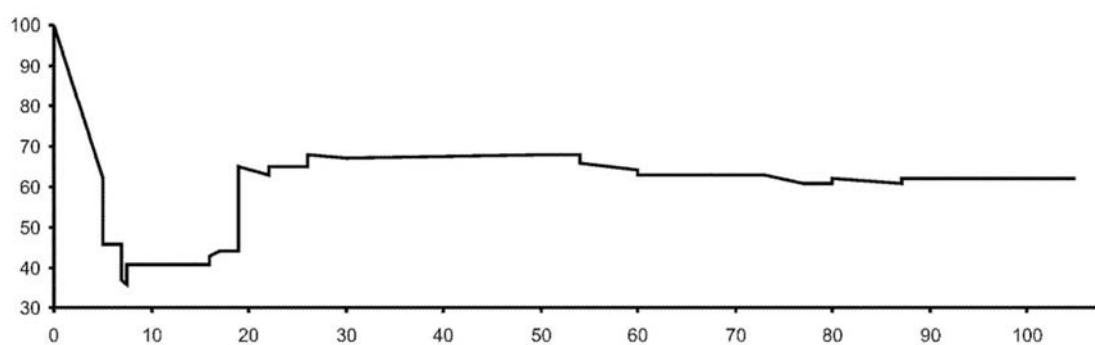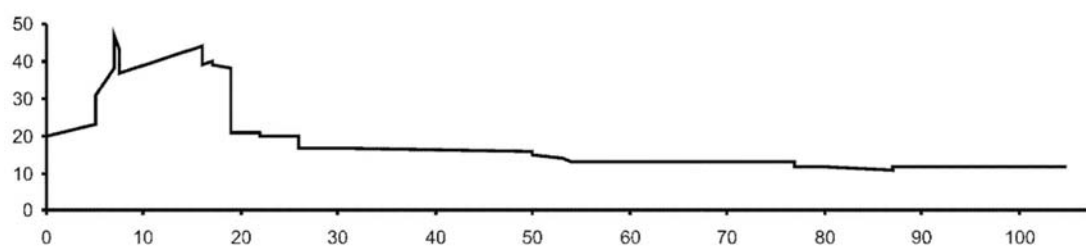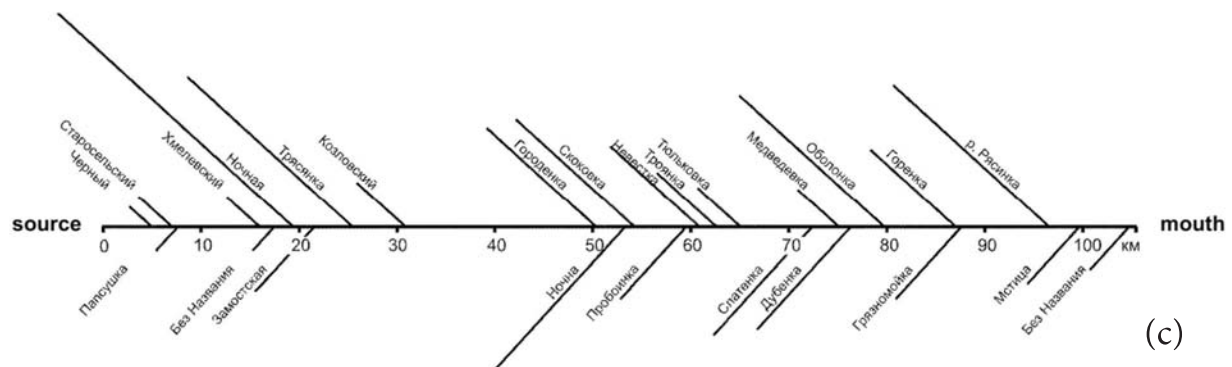
47 ENVIRONMENT
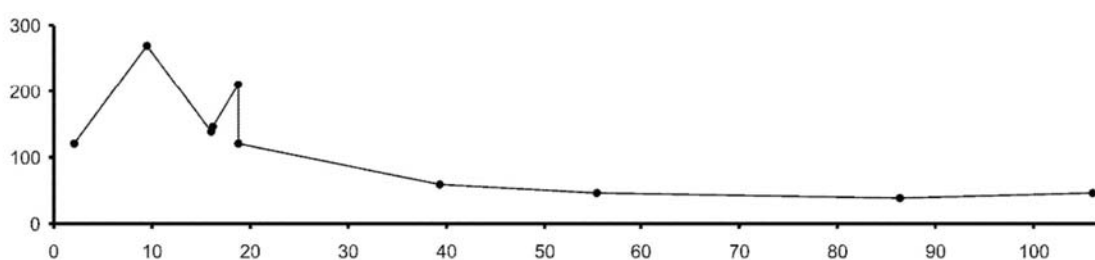
Fig. 3. The Tudovka River near its source "Istok" (a) and the lower course at Redkino (b); hydrographical schema of the Tudovka river with its main tributaries (c); the next graphs illustrate the changing parameters along the longitudinal course of the river: (d) amount of mires in the catchment (%), (e) forest in the catchment (%) and (f) the color (°) of the water (1.05.2005) (after: [Zhenikov et al 2007], modified)
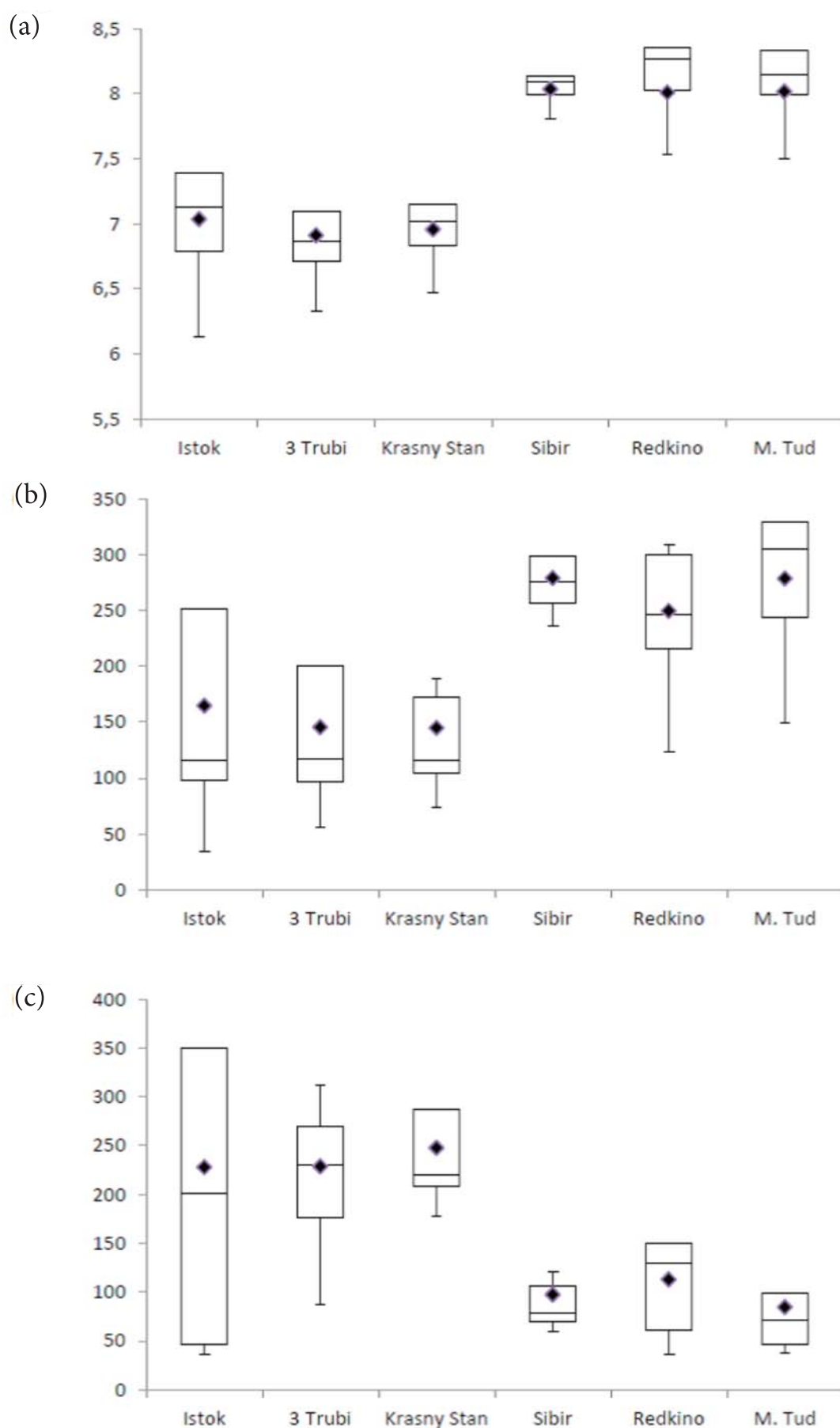
Fig. 4: Physico-chemical characterisation of the investigated sites, based on field measurements in the years 2004 – 2010: (a) pH, (b) conductivity, (c) colour. For the different stations, the amount of measurements are not the same, i.e. Istok (n = 10), 3 Trubi (n = 10), Krasny Stan (n = 9), Sibir (n = 4), Redkino (n = 9), M. Tud (n = 10).
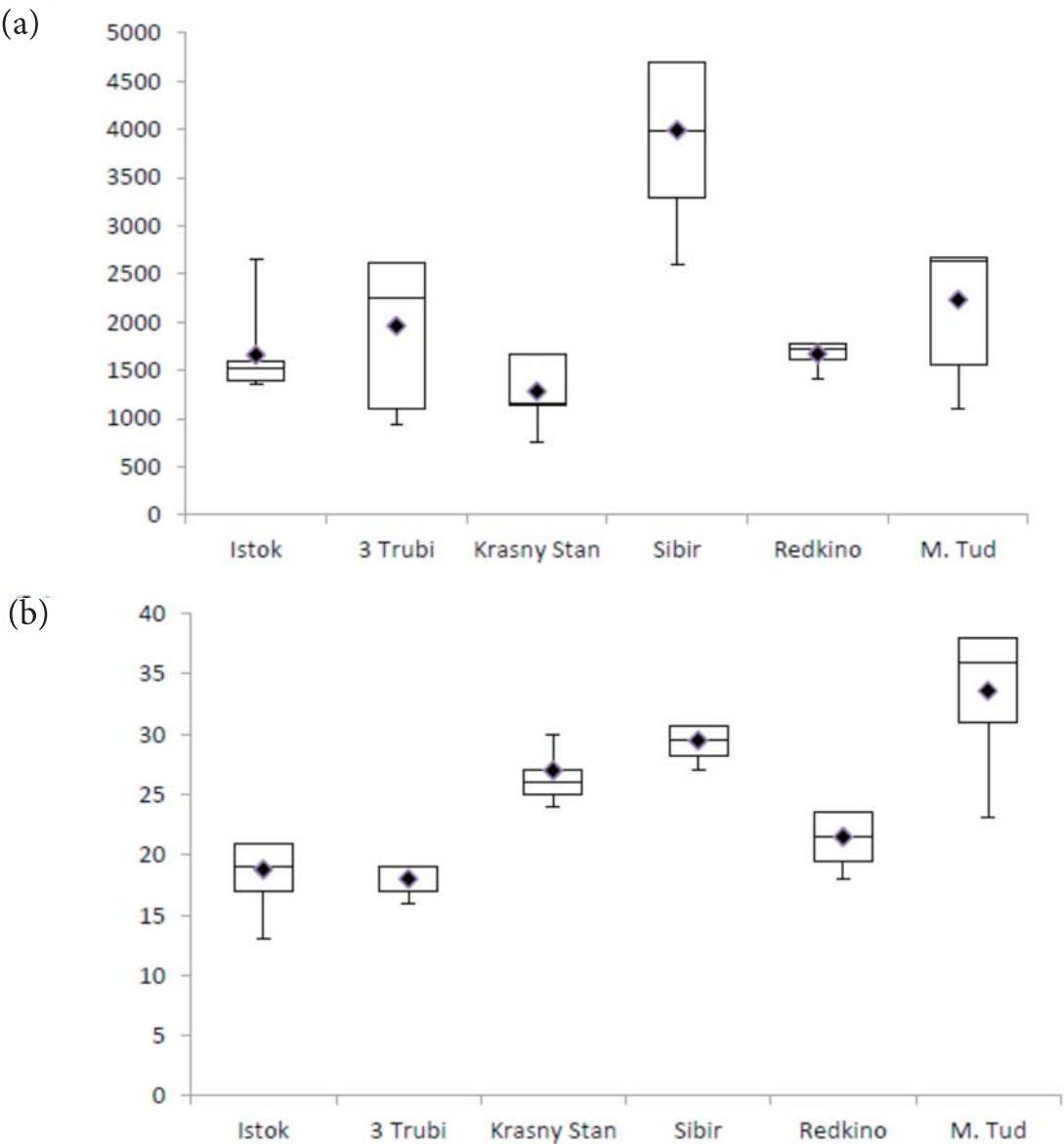
(a)

(b)

Fig. 5: Characteristics of the zoobenthos community during summer low flow period, based on samples from 2006 – 2010: (a) individual per m² and (b) number of taxa (chironomidae + oligochaeta handled as 1 taxon each). The amount of samples is usually 5, while less samples are available for Sibir (n = 2) and Redkino (n = 4).

Table 4. Distance matrix between all sites based on cumulative species data from 5 years [Schletterer et al., 2010]. Shortest distances corresponding to the nearest neighbours in species space are indicated (red / bold).

	Istok	3Trubi	Krasny Stan	Sibir	Redkino	Molodoi Tud
Istok	0.00	**114.33**	129.15	149.94	126.39	131.87
3Trubi	114.33	0.00	**110.00**	168.11	119.66	138.04
Krasny Stan	129.15	**110.00**	0.00	165.47	121.20	142.74
Sibir	149.94	168.11	165.47	0.00	165.93	**137.27**
Redkino	126.39	119.66	121.20	165.93	0.00	**109.24**
Molodoi Tud	131.87	138.04	142.74	137.27	**109.24**	0.00

Table 5. Affinity matrix used to evaluate the similarity of investigation sites to each other. The affinity matrix is not symmetric because affinity is normalized by the maximum distance observed across all other sites (Eqn. 2). Similar reaches are marked in red, the maximum affinity is always 1.0.

	Istok	3Trubi	Krasny Stan	Sibir	Redkino	Molodoi Tud
Istok	1.00	0.24	0.14	0.00	0.16	0.12
3Trubi	0.32	1.00	0.35	0.00	0.29	0.18
Krasny Stan	0.22	0.34	1.00	0.00	0.27	0.14
Sibir	0.11	0.00	0.02	1.00	0.01	0.18
Redkino	0.24	0.28	0.27	0.00	1.00	0.34
Molodoi Tud	0.08	0.03	0.00	0.04	0.23	1.00
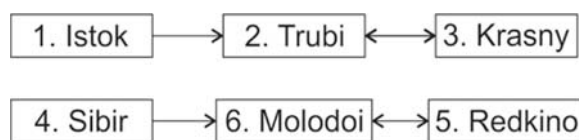
Fig. 6: Affinity diagram showing which sites are the most strongly related. In total, two distinct groups were formed, containing the sites 1 2 3 and 4 5 6. Double arrows indicate bilateral affinity, e.g. 3Trubi and Krasny Stan were found to be most strongly related to each other, whereas Istok was found to have the strongest affinity to 3Trubi.

were quite similar in the upper course (mean values = 1,659; 1,959 and 1,281 ind. m^{-2}, respectively), while they were higher in the middle and lower course (Fig. 5 a).

In total, six sites were compared using cumulative zoobenthos data from the years 2006-2010 [Schletterer et al., 2010], and the distance matrices and affinity matrices were generated using the KNN species space classification method. As shown in Table 4, the distance matrix for all six sites

reveals that Istok, 3Trubi and Krasny Stan can be considered as a group of nearest neighbors, and Sibir, Redkino and Molodoi Tud constitute a second group. After normalization, the affinity matrix results can be used not only to determine which sites are the closest matches in terms of the species space affinity (Table 5), but also how interrelated they are to one another (Fig. 6). Here it was found that the affinities between the six sites followed the same general pattern of two distinct groups, where it was found that for the first group, 3Trubi and Krasny Stan were more similar to each other, and in the second group Redkino and Molodoi Tud were reflexively similar as well. Different species assemblages are linked to certain biocoenotic regions along the continuum (crenal rhithral potamal) and certain longitudinal distribution patterns for instance are linked to different species [Moog, 2002]. A k-NN classification of the benthic community – based on presence / absence (p/a) of species clearly revealed longitudinal patterns, i.e. a classification between the source (Istok), the upper course

Table 6. Distance matrix between the sampling locations based on diatom samples collected in 2006 for five of the investigation reaches. The species space nearest neighbors are indicated (red / bold).

	Istok	3Trubi	Krasny Stan	Sibir	Molodoi Tud
Istok	0.00	185.10	178.10	171.70	**162.50**
3Trubi	185.10	0.00	**144.80**	191.80	183.00
Krasny Stan	178.10	**144.80**	0.00	194.20	180.20
Sibir	171.70	191.80	194.20	0.00	**114.20**
Molodoi Tud	162.50	183.00	180.20	**114.2**	0.00

Table 7. Affinity matrix of diatom similarity of five investigation sites.

	Istok	3Trubi	Krasny Stan	Sibir	Molodoi Tud
Istok	1.00	0.00	0.04	0.07	0.12
3Trubi	0.03	1.00	0.25	0.00	0.05
Krasny Stan	0.08	0.25	1.00	0.00	0.07
Sibir	0.12	0.01	0.00	1.00	0.41
Molodoi Tud	0.11	0.00	0.02	0.38	1.00

2. Trubi ←→ 3. Krasny

1. Istok → 6. Molodoi ←→ 4. Sibir

Fig. 7: Affinity diagram of the 2006 benthic diatoms. Two groups were also found, the major difference being that Istok was more closely linked with Molodoi and Sibir instead of 3Trubi and Krasny Stan.

(3Trubi and Krasny Stan), the middle course (Sibir) and the lower course (Redkino and Molodoi Tud).

BENTHIC DIATOMS

2006 Diatoms.

To compare the methodology using an additional species, diatom data sets from 2006 and 2007 were also collected, for five of the six sites, excluding Redkino. The distance (Table 6) and affinity matrices (Table 7) were calculated in the same manner as for the composite zoobenthos dataset. It was again observed that the KNN species space classification broke the five sites into two distinct groups, the first consisting of

3Trubi and Krasny Stan, the second group was made up of Istok, Sibir and Molodoi Tud. Considering the 2006 diatom dataset, it was found that once again 3Trubi and Krasny Stan were self-similar, however this must be the case with only two members in a group (Fig. 7). Considering the second group, the investigation reaches Sibir and Molodoi Tud were found to be the most similar.

2007 Diatoms.

The final data set compared was using 2007 diatom data. All six investigation reaches were available and were compared using the same methodology for the species space distances (Table 8) as well as the affinity matrix (Table 9). An interesting observation is that the same two groups emerged (Fig. 8) from the 2007 diatom data as for the zoobenthos and 2006 diatom data sets, indicating that the proposed KNN species space method may be able to provide robust classification across multiple sites and years. The affinities were also identical to the zoobenthos analysis, the only difference being that in the second group, it was found that Sibir and Molodoi Tud were the most self-similar, as was also found in the 2006

Table 8. Distance matrix between the sampling locations based on data collected in 2007 for five of the investigation reaches. The species space nearest neighbors are indicated (red / bold).

	Istok	3Trubi	Krasny Stan	Sibir	Redkino	Molodoi Tud
Istok	0.00	174.80	**143.60**	180.00	179.60	185.10
3Trubi	174.80	0.00	**134.60**	189.20	171.60	190.10
Krasny Stan	143.60	**134.60**	0.00	181.80	164.80	179.70
Sibir	180.00	189.20	181.80	0.00	109.40	55.10
Redkino	179.60	171.60	164.80	109.40	0.00	102.70
Molodoi Tud	185.10	190.10	179.70	**55.10**	102.70	0.00

52 ENVIRONMENT

Table 9. Affinity matrix of diatom similarity of the sampling locations.

	Istok	3Trubi	Krasny Stan	Sibir	Redkino	Molodoi Tud
Istok	1.00	0.06	0.22	0.03	0.03	0.00
3Trubi	0.08	1.00	0.29	0.00	0.10	0.00
Krasny Stan	0.21	0.26	1.00	0.00	0.09	0.01
Sibir	0.05	0.00	0.04	1.00	0.42	0.71
Redkino	0.00	0.04	0.08	0.39	1.00	0.43
Molodoi Tud	0.03	0.00	0.05	0.71	0.46	1.00

Fig. 8: Affinity diagram of the 2007 benthic diatoms. Two groups were also found, with the same reaches (1,2 3 and 4,5,6) as in the 2006-2010 zoobenthos classification.

diatom analysis. This also indicates that the differences between the community structures may be detected by the method, but that the overall grouping of the individual sites remains robust. Due to the small number of investigation reaches, a larger study would be required to substantiate this.

DISCUSSION AND OUTLOOK

Our analyses underline that the KNN algorithm is suitable for the classification of benthic biocenosis. Application is possible on species level as well as using traits (e.g. functional feeding types). The methodology is straightforward and relies only on the accuracy of the bioassay for each site. The proposed species space concept enables a detailed comparison of study sites and their interrelation, thus it is a useful tool for the comparison of large datasets, i.e. Long Term Ecological Research and Monitoring (LTERM) data. Using different biological components as well as the combination of all of them could even provide more robust classification and further interclass differentiation of investigation reaches.

ACKNOWLEDGEMENTS

REFERENCES

1. Aggarwal CC, Hinneburg A, Keim DA (2001) On the surprising behavior of distance metrics in high dimensional spaces. Proceedings of the 8th International Conference on Database Theory (ICDT): Lecture Notes in Computer Science 1973:420–434

2. Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans. Inform. Theory IT-13:21-7

3. Handbook of Water Resources of the USSR. (1936) V.3. Catchment of the Upper Volga and Oka. Part 2.

4. Hering D, Buffagni A, Moog O, Sandin L, Sommerhauser M, Stubauer I, Feld C, Johnson R, Pinto P, Skoulikidis N, Verdonschot P, Zahradkova S (2003) The development of a system to assess the ecological quality of streams based on macroinvertebrates – design of the sampling programme within the AQEM project. International Review of Hydrobiology 88: 345–361.

5. Kelly M.G., Bennion H., Cox E.J., Goldsmith B., Jamieson J., Juggins S., Mann D.G., Telford R.J. (2005): Common freshwater diatoms of Britain and Ireland: an interactive key. (http://craticula.ncl.ac.uk/EADiatomKey/html/, 22.05.2008), Environment Agency, Bristol.

6. Kingston J. (1985). Diatom analisis. Paleoecological investigation of recent lake acidification (Pirla). Interim version, Indiana University. In: Schiedele S. 1987. Indikatorwert benthischer Diatomeen in der Isar zw. Mittenwald u. Landshut, Diplomarbeit a. d. Universität München

7. Krammer K. & Lange-Bertalot H. (1986-2004). Bacillariophyceae. 1-5. Süsswasserflora von Mitteleuropa. Gustav Fischer Verlag, Stuttgart, New York.

8. Kurlov MG (1928) Classification of mineral Waters of Siberia. Tomsk, U.S.S.R.; cit. in: Zaporozec A (1972) Graphical Interpretation of Water-Quality Data. Ground Water Vol. 10, No.2

9. MOOG O (Ed.) (2002): Fauna Aquatica Austriaca, Lieferung 2002.- Wasserwirtschaftskataster, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien (Online: https://www.bmlfuw.gv.at/wasser/wasser-oesterreich/plan_gewaesser_ngp/umsetzung_wasserrahmenrichtlinie/FAA.html).

10. Prygiel J. & Coste M. (eds.) (2000). Tax'IBD – Logiciel de reconnaissance des diatomées retenues pour le calcul de l'Indice Biologique Diatomées (Frensh and English Version on CD). Cemagref

11. Puzachenko YG, Zheltukhin AS, Kozlov DN, Korablyov NP, Fedyaeva MV, Puzachenko MY, Siunova EV (2007) Central Forest State Biosphere Reserve – 75 years, pp. 80, Delovoi Mir.

12. REFCOND (2003) Guidance on establishing reference conditions and ecological status class boundaries for inland surface waters, 86 p., Final Report of the Working Group 2.3 (EU Common Implementation Strategy (CIS) for the Water Framework Directive).

13. Schletterer M (2009) Benthic invertebrates and reference conditions in East European running waters: case study in the headwaters of Volga River (Tver region, Russia) and adjacent water bodies. 302 pp., dissertation (PhD thesis), Faculty of Biology of the University of Innsbruck.

14. Schletterer M, Kuzovlev VV, Zhenikov YN (2010) Macrozoobentos and bottom diatoms of the Tudovka River (Never Oblast, Russia. In: Dynamics of phenomena and processes of the natural complex of the reserve: Letopis prirody 50: 412-425.

15. Schletterer M, Füreder L, Kuzovlev VV, Zhenikov YN, Grigorieva IL (2014) Lowland river reference condition: spatial and temporal patterns of the zoobenthos community in the Volga headwaters (2006–2010). Hydrobiologia 729:175–189

16. Schletterer M, Füreder L, Kuzovlev VV, Zhenikov KY, Zhenikov YN (2016) REFCOND_VOLGA: a monitoring programme for water quality in the headwaters of the Volga River (Tver region, Russia). Revista Eletrônica de Gestão e Tecnologias Ambientais (GESTA) 4/1: 18-40.

17. Surin ZS (Ed.) (1976): Resources on surface waters of the USSR. Vol. 10. The Upper-Volga area. (Resursy poverhnostnih vod SSSR, T.10 Verhne-Volgskiy Rayon), Moscow Dept. of Gidrometeoizdat

18. Zagorski GF (1967) Hydrological areas Kalinin region (by the nature of the water regime of the rivers) // Scientific notes of the Kalinin State Pedagogical Institute. Kalinin, Vol. 44. pp. 99-111.

19. Zimek A, Schubert E, Kriegel H-P (2012) A survey on unsupervised outlier detection in high-dimensional numerical data. Statistical Analy Data Mining 5:363–387

20. Zhenikov YN, Kuzovlev VV, Schletterer M. (2007) Obosnovaniye programmy monitoringa i rezultaty issledovanij reki Tudovki (po gidrologicheskim, gidrohimicheskim i gidrobiologicheskim pokazatelam) – Description of the monitoring programme and the results of research on River Tudovka (hydrological, hydrochemical and hydrobiological parameters). Trudy Central'no-Lesnogo gosudarstvennogo prirodnogo biosfernogo Zapovednika 5: 407 – 419.

Martin Schletterer studied biology, with a focus on zoology and limnology, at the University of Innsbruck (Austria) and graduated in 2007. He obtained his doctoral degree (PhD) for his research on "Benthic Invertebrates and Reference Conditions in East European Running Waters: Case study in the headwaters of Volga River (Tver Region, Russia) and adjacent water bodies" in 2010. He was working on national and international scientific projects since 2003 and his research interests are concentrated in the following areas: monitoring, biodiversity (zoobenthos and other aquatic quality components), investigation of anthropogenic stressors at large spatial scales and in the context of environmental factors as well as river basin management. Since 2009 he is working at TIWAG – Tiroler Wasserkraft AG (Hydropower Planning Department) as head of the group "Ecology" and in this role he is responsible for planning and management of ecological aspects as well as R&D projects. Additionally, Dr. Schletterer has teaching assignments at the Institute for Modelling Hydraulic and Environmental Systems at University of Stuttgart (Germany), the University of Vienna (Austria) as well as at the Universidade de Lisboa (Portugal).

Leopold Füreder, Associate Professor at the University Innsbruck. Holds a Master's degree in Limnology and Taxonomy (Master thesis on «Vertical distribution of Zoobenthos of the hard stubstrate areas in Lake Piburg, Tirol, Austria». Supervision: Prof. Dr. Roland Pechlaner) and a Ph.D in Tropical Ecology and Stream Ecology (Doctoral thesis at the University of Innsbruck and Drexel University, Philadelphia, as a Fulbright scholar, on «Drift patterns in neotropical streams in Costa Rica». Supervisors: Prof. Dr. Eugen Rott, Innsbruck, and Prof. Dr. Susan Kilham, Philadelphia). Scientific interest on the structure and function of stream ecosystems with special attention to (high) alpine areas and the taxonomy of high alpine aquatic insects, involved in basic and applied research topics. Currently involved in teaching at the undergraduate and graduate level as well as supervision of Master and Ph.D students

Vyacheslav V. Kuzovlev, candidate of technical sciences, associate professor. In 1994 he graduated from Tver State Polytechnic Institute (Russia) in «Open pit mining» and in 2002 he defended his thesis on «Development of Methods for the Environmental Evaluation of the Wise Use of Peatlands». Since 2000 he is working at Tver State Technical University. Research interests: evaluation of the impact of natural landscapes on the hydrochemical regime of riverine systems. Since 2011 he is also the head of the Laboratory of Environmental Monitoring of the Tver Center for Hydrometeorology and Environmental Monitoring (ROSHYDROMET division in Tver region)

Yuri N. Zhenikov graduated in 1973 from the Kalinin Polytechnic Institute (Russia), with the specialty «Technology and complex mechanization of peat deposits». In 2000 he defended his doctoral thesis «Geoecological basis for the rational use of peat bog resources of the Upper Volga Region». Since 1996 he works in international scientific projects. Research interests are focused in the following areas: the interaction of peat bogs and the environment, the composition and properties of the surface water as well as environmental impact assessment. Since 1973 he is working at Tver State Technical University, in 2004 he became head of the Chair of Nature Management and Ecology. He also supervises the department of Geology, exploration and technological design of the East European Institute of Peat Business

Juan Francisco Fuentes-Pérez is a junior researcher in the Centre for Biorobotics of the Tallinn University of Technology (Estonia). His research interests focus primarily on ecohydraulics, hydraulic modeling, fish passes and underwater robotics. Since 2009 he has collaborated in several research projects focused on design, simulation and evaluation of fishways (GEA-Ecohydraulics, University of Valladolid), and since 2014 he has been working with artificial lateral line technology and its applications for the evaluation of aquatic ecosystems as well as its use in underwater robotics

Jeffrey A. Tuhtan received the B.Sc. degree in civil engineering from California Polytechnic State University, San Luis Obispo, USA, in 2004, and a M.Sc. degree in water resources engineering and management, in 2007, and his Dr.-Ing. degree in hydraulics from the University of Stuttgart, Germany, in 2011. At the Center for Biorobotics, Tallinn University of Technology, Estonia, he leads the Environmental Sensing and Intelligence Group. His research areas include environmental intelligence, bio-inspired fluid mechanics and remote sensing