

# Estimates of benthic invertebrate community variability and its environmental determinants differ between snapshot and trajectory designs

Kaisa-Leena Huttunen<sup>1,4</sup>, Heikki Mykrä<sup>2,5</sup>, Riku Paavola<sup>3,6</sup>, and Timo Muotka<sup>1,2,7</sup>

<sup>1</sup>Department of Ecology and Genetics, P.O. Box 3000, FI-90014 University of Oulu, Finland

<sup>2</sup>Finnish Environment Institute, Freshwater Centre, P.O. Box 413, FI-90014 Oulu, Finland

<sup>3</sup>Oulanka Research Station, University of Oulu Infrastructure Platform, Liikasenvaarantie 134, FI-93900 Kuusamo, Finland

**Abstract:** Long-term data sets are essential for biodiversity research and monitoring. Researchers use 2 major approaches in the study of temporal variability of biological communities: 1) the trajectory approach (monitoring sites across several consecutive years) and 2) the snapshot approach (comparing sites among few sampling events several years apart). We used data on benthic macroinvertebrate communities in 23 near-pristine forested streams to compare these 2 approaches for different study periods ranging from 3 to 14 y. We asked whether the level of temporal turnover and the identity of the best explanatory variables underlying it were comparable across studies based on differing approaches, study periods, or total duration. The 2 approaches yielded partly different stories about the level of community variability and its environmental correlates. With the snapshot approach, variation in community similarity and factors explaining it reflected short-term (e.g., year-specific) conditions, which could be misinterpreted as long-term trends, the difference being most evident for periods that began or ended in an extreme drought year. Our results imply that snapshot studies may lead to ambiguous conclusions, whereas the trajectory approach yielded more consistent results. Trajectory data of differing length showed minor differences, apart from studies with the shortest durations. Overall, our results suggest that time sequences of ~6 y of trajectory data (i.e., 6 generations for most benthic invertebrates in boreal streams) may be needed for the among-year similarity of macroinvertebrate communities in near-pristine streams to stabilize. If temporal replication is limited (snapshots/very short time sequences) the outcome depends strongly on the particular years included in a comparison. Based on our results, we advise caution when basing conclusions on a comparison of a few (e.g., just 2) occasions several years apart or on very short time sequences.

**Key words:** interannual variability, temporal beta diversity, temporal replication, study duration, biomonitoring, stream invertebrates, drought

Long-term data sets are a prerequisite for biodiversity research and monitoring, and they contribute to our understanding of temporal population and community dynamics (Magurran et al. 2010, Jourdan et al. 2018). The 2 major approaches to the study of variability of biological communities through time are the trajectory approach and the snapshot approach (Hildrew and Giller 1994). In the trajectory approach, study sites are monitored regularly across several consecutive years (e.g., Scarsbrook 2002, Bêche et al. 2006, Angeler and Johnson 2012), whereas in the snapshot approach, monitoring at each site is conducted on only a few (sometimes just 2) occasions, years apart (e.g., Town-

send et al. 1987, Johnson et al. 1994, Woodward et al. 2002). As a consequence, the trajectory data should be better able to detect annual differences and the effects of exceptional years than the snapshot approach, for which the outcome may depend strongly on the environmental conditions of the particular years included ('too few replicates in time'; Hildrew and Giller 1994). On the contrary, long-term trajectory studies typically are based on just a few monitoring sites, in which case, results may not be easily extrapolated beyond these particular sites ('too few replicates in space'). The 2 approaches yield different types of information about temporal turnover in community com-

E-mail addresses: <sup>4</sup>kaisa-leena.huttunen@oulu.fi; <sup>5</sup>heikki.mykra@ymparisto.fi; <sup>6</sup>riku.paavola@oulu.fi; <sup>7</sup>timo.muotka@oulu.fi

position, but to our knowledge, no freshwater study has been done to systematically compare their outcomes.

Snapshot studies vs temporally repeated sampling are the 2 extremes of a spectrum, and any approach between the 2 is possible. For example, personnel working as part of large-scale biomonitoring programs may compile data from several sources and sites, often sampled on a few separate occasions but rarely continuously (and concurrently). The decision regarding which sampling design to use is always related to the question of how best to allocate limited resources available for monitoring, i.e., should one include fewer sites but monitor those intensively through time or is it better to focus on extensive spatial replication at the cost of ignoring the potential importance of temporal community variability. This is a fundamental question for biomonitoring and for all research on long-term ecological variability and factors controlling it.

Results of a monitoring study also may vary depending on the length of the study. This possibility seems obvious when comparing studies where trends are measured at time scales that vary by orders of magnitude (e.g., days vs weeks, years vs decades; Jackson and Füreder 2006), but it is equally relevant when time scales differ only slightly. For example, Bengtsson et al. (1997) compared study lengths ranging from 2 to 22 y to examine the effects of temporal scale on woodland bird community variability, finding that variability increased with the length of time over which the communities were observed—a result that calls into question the comparability of studies of differing length. In another ornithological study, Banks-Leite et al. (2012) found that temporal distribution of sampling was more important than total sample size and argued that authors of studies with the same sampling effort but different temporal distribution of sampling may report different ecological patterns. Similar requests for allocating more resources to expanding study duration rather than investing in intensive short-term sampling campaigns have been made for butterflies (Mac Nally et al. 2004) and grassland vegetation (Baasch et al. 2010).

Quantification of the past and prediction of future trends in biodiversity are contingent on the availability of long-term data sets for different taxonomic groups and ecosystem types (Magurran et al. 2010). Freshwater ecosystems have suffered disproportionately high biodiversity loss (Vörösmarty et al. 2010), yet long-term data sets for freshwater communities are rare. Streams provide a fitting environmental template for studies on temporal community turnover because they are extremely variable and disturbance-prone environments, yet lotic communities often exhibit remarkable long-term stability (Sponseller et al. 2010, Huttunen et al. 2017). Stream macroinvertebrates are a diverse group of organisms filling a variety of ecological niches, with relatively short life-cycles that enable rapid responses to environmental change and inclusion of several generations within a relatively short time scale (Rosenberg and Resh 1993).

We evaluated the influence of survey design on the detection of interannual community variability. We used our

14-y data set from 23 near-pristine boreal streams to address 3 questions pertinent to any study of temporal patterns in biological communities: 1) do the level of temporal variation in community composition and the factors driving it depend on survey design; 2) are results from studies of differing duration comparable; and 3) how long should a temporal data set be to obtain a representative picture of community variability?

## METHODS

### Study sites and sampling

We collected macroinvertebrate samples across 14 consecutive years (2000–2013) in 23 1<sup>st</sup>-to-2<sup>nd</sup>-order streams in the Koutajoki drainage basin in northeastern Finland, just south of the Arctic Circle (lat 66–67°N, long 28–30°E). Koutajoki basin, especially Oulanka National Park—a nature conservation reserve within the basin—represents the westernmost remnants of pristine taiga forests. Many of our study streams are located within the national park and others drain areas of minimal anthropogenic influence (Malmqvist et al. 2009, Huttunen et al. 2012).

The same field crew sampled in autumn (September–early October) each year. At each site, we collected a 2-min kick-net sample (mesh size = 0.3 mm) covering most microhabitats present in a riffle. Such a sample covers ~1.3 m<sup>2</sup> of the stream bed and captures ~75% of species present in a riffle, missing mainly species that occur sporadically in streams (Mykrä et al. 2006). We sorted the samples and counted and identified all individuals to the lowest feasible taxonomic level, mainly species. We did not count chironomids every year, so we excluded them from all analyses.

We measured a set of environmental variables potentially related to temporal variation of macroinvertebrate communities at each site. We quantified habitat stability at a scale relevant to stream organisms by monitoring the movement of painted stones ('substrate movement intensity') twice a year for 5 successive years (2005–2009): after the spring flood and in the autumn concurrently with benthic sampling (Huttunen et al. 2017). We calculated substrate movement intensity for each site as the mean percentage of stones moved across the study period (Townsend et al. 1997). We quantified the amount of in-stream vegetation as macrophyte cover (%), estimated visually at 20 randomly placed 50- × 50-cm quadrats. We measured within-site habitat heterogeneity as the Simpson index of substratum diversity, calculated from the proportional distribution of different particle size classes in 10 similar-sized quadrats. We used a modified Wentworth scale from silt (1) to large boulder and bedrock (10) (Huttunen et al. 2012). We monitored water temperature with data loggers (WT-HR 1000 mm, TruTrack Ltd, New Zealand) deployed in each stream from late May to early October in 2009 to 2012. We used daily averages (recording interval = 30 min) to calculate mean water temperature across years for each site. We measured habitat

connectivity, describing dispersal potential from adjacent reaches, as the relative isolation of a site within a stream network by quantifying, based on site visits, the surface area of riffle habitat available for stream invertebrates within a 500-m buffer up- and downstream of a study site. In 4 of the 23 streams, this buffer included nearby tributaries.

The study period of 2000 to 2013 included a few climatically exceptional years. Summertime precipitation within a 2-mo period before the collection of macroinvertebrate samples was exceptionally low compared to long-term records in 2006 and, to a lesser degree, in 2003, whereas 2010 was wetter than the other study years (Fig. 1A). Based on the summertime temperature sum, 2003 and 2005 were the warmest and 2008 the coldest study years (Fig. 1B).

### Data analysis

The level of temporal variation in community composition and, consequently, the potential difference between the snapshot and trajectory estimates, may depend on the form of long-term community change, i.e., whether it is directional, cyclical, or erratic (see Collins et al. 2000). Therefore, we first determined the form of temporal variation in

community composition by illustrating change through time in ordination space (nonmetric multidimensional scaling [NMDS]). NMDS was run for the whole data set using Bray–Curtis dissimilarity index based on  $\log(x + 1)$ -transformed abundance data, after which we plotted the site-specific ordination scores against time, separately for each site.

We used the Bray–Curtis dissimilarity index based on  $\log(x + 1)$ -transformed abundance data as the measure of temporal variation of macroinvertebrate communities. In this analysis, low dissimilarity values represent low interannual variability in community composition, i.e., low temporal turnover. To assess whether the snapshot and trajectory approach resulted in similar outcomes, we calculated community dissimilarity by: 1) contrasting only the first and last year of each study period (snapshot approach), and 2) as an average across all consecutive year pairs for study lengths ranging from 3 to 14 y (trajectory approach). For example, for the 6-y study period from 2000 to 2005, we calculated temporal turnover as dissimilarity in community composition between years 2000 and 2005 (snapshot), and for all consecutive year pairs, i.e., 2000 and 2001, 2001 and 2002, etc., the final measure of temporal turnover then being the mean of dissimilarity values across all consecutive year pairs (trajectory). We calculated dissimilarities in the program R with the package *vegan* (version 2.2-0; Oksanen et al. 2014).

To reflect the real-life situation with limited resources for replication, we equalized the sampling effort between the snapshot and trajectory approaches and among different study durations as closely as possible (Table 1). For the snapshot approach, sampling effort was always 46 samples/study period (23 sites  $\times$  2 y). For the trajectory approach, the number of temporal replicates increased with study length. Therefore, we reduced the number of spatial replicates to keep the sampling effort about equal for both designs. We achieved this by randomly sampling (without replacement) a subset of study sites ( $n = 3$ –16) from the pool of 23 sites (Table 1). Thus, the total number of samples per study period ranged from 42 to 52 (3 sites  $\times$  14 y; 4 sites  $\times$  13 y, respectively). We repeated this procedure 100 times and calculated community dissimilarity for each study period as a mean across the 100 runs.

To explore the comparability of results from studies using trajectory data of differing duration, we calculated the level of temporal variation in community composition between randomly ordered nonoverlapping sequences of 2, 5, and 7 consecutive years, and the whole study period of 14 y (all study sites included; Bengtsson et al. 1997). We used nonoverlapping sequences to render different study durations independent of each other. In practice, the study period of 14 y was randomly split into 3 pieces of consecutive years, e.g., 2000–2004 (5 y), 2005–2011 (7), and 2012–2013 (2), or 2000–2006 (7), 2007–2008 (2), and 2009–2013 (5). A new random sequence was drawn for each study site. We tested how the removal of the drought year (2006) affected these comparisons by repeating the analysis by excluding year

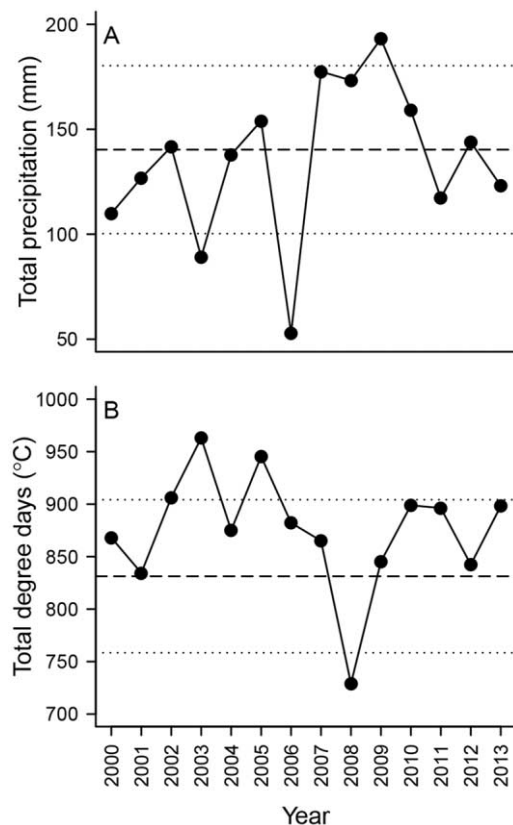


Figure 1. Summertime (July–August) total precipitation (A) and air temperature (degree-days, °C) (B) in the Koutajoki catchment area during the study period of 2000 to 2013. The dashed line indicates long-term (1969–2013) mean, and dotted lines show  $\pm 1$  SD.

Table 1. Study years and sampling effort (total number of samples) for different study durations. Different numbers of study sites were used in the trajectory and snapshot approaches to equalize the sampling effort. In the trajectory approach, the number of study sites decreased with longer study duration (time span) because of an increase in the number of site-specific temporal replicates, whereas in the snapshot approach both spatial and temporal replication is the same independent of study duration (23 and 2, respectively).

Time span (y)	Research years	Number of study sites		Sampling effort	
		Trajectory	Snapshot	Trajectory	Snapshot
3	2000–2002, 2001–2003, 2002–2004, 2003–2005, 2004–2006, 2005–2007, 2006–2008, 2007–2009, 2008–2010, 2009–2011, 2010–2012, 2011–2013	16	23	48	46
4	2000–2003, 2001–2004, 2002–2005, 2003–2006, 2004–2007, 2005–2008, 2006–2009, 2007–2010, 2008–2011, 2009–2012, 2010–2013	12	23	48	46
5	2000–2004, 2001–2005, 2002–2006, 2003–2007, 2004–2008, 2005–2009, 2006–2010, 2007–2011, 2008–2012, 2009–2013	9	23	45	46
6	2000–2005, 2001–2006, 2002–2007, 2003–2008, 2004–2009, 2005–2010, 2006–2011, 2007–2012, 2008–2013	8	23	48	46
7	2000–2006, 2001–2007, 2002–2008, 2003–2009, 2004–2010, 2005–2011, 2006–2012, 2007–2013	7	23	49	46
8	2000–2007, 2001–2008, 2002–2009, 2003–2010, 2004–2011, 2005–2012, 2006–2013	6	23	48	46
9	2000–2008, 2001–2009, 2002–2010, 2003–2011, 2004–2012, 2005–2013	5	23	45	46
10	2000–2009, 2001–2010, 2002–2011, 2003–2012, 2004–2013	5	23	50	46
11	2000–2010, 2001–2011, 2002–2012, 2003–2013	4	23	44	46
12	2000–2011, 2001–2012, 2002–2013	4	23	48	46
13	2000–2012, 2001–2013	4	23	52	46
14	2000–2013	3	23	42	46

2006 (the sequences compared being 2, 4, and 6 consecutive years, and the whole study period of 13 y).

We used multimodel inference in multiple linear regressions (Burnham and Anderson 2002) with the R package *MuMIn* (Barton 2012) to examine whether the same factors were identified as the best explanatory variables for temporal community variability for different study periods/durations in each approach and between the approaches for each study period. To restrict the number of candidate models, no interaction terms were allowed. We compared the identity of the factors included in the best models, i.e., those with the lowest AIC<sub>c</sub> score (Akaike Information Criterion with small-sample correction). In addition, we compared the relative importance of each explanatory variable among different survey designs. Importance values for each variable were obtained by summing model weights across all models that included that variable (Burnham and Anderson 2002). For the trajectory data, we assessed the identities and relative importance of each variable using both the whole data set of 23 sites and random subsets. For random subsets, we set the minimum number of spatial repli-

cates to 9 sites (vs Table 1) so that we could conduct regression analyses for surveys of longer duration.

## RESULTS

We processed 322 macroinvertebrate samples containing 613,159 individuals. The overall number of taxa across all sites and years was 129. Mean taxonomic richness per site and year, i.e., temporal  $\alpha$ -diversity, was 31 (range: 25–40). The form of variation in community composition through time did not show any obvious trend or directional change (Fig. S1), but was negligible (Fig. 2A, B), erratic (Fig. 2C, D), or dictated by certain years (Fig. 2E, F).

Temporal turnover (community dissimilarity through time) was generally higher when the snapshot approach was used (65 of 78 cases; Fig. 3). The difference was most evident for the periods that began or ended in the drought year 2006. The absolute difference between the 2 approaches was, on average, 0.09, but site-specific differences could be as high as 0.55 (Fig. 4). With the snapshot data, temporal turnover for the periods that began or ended in 2006 differed clearly from all others (Fig. 3). For the trajectory data, the

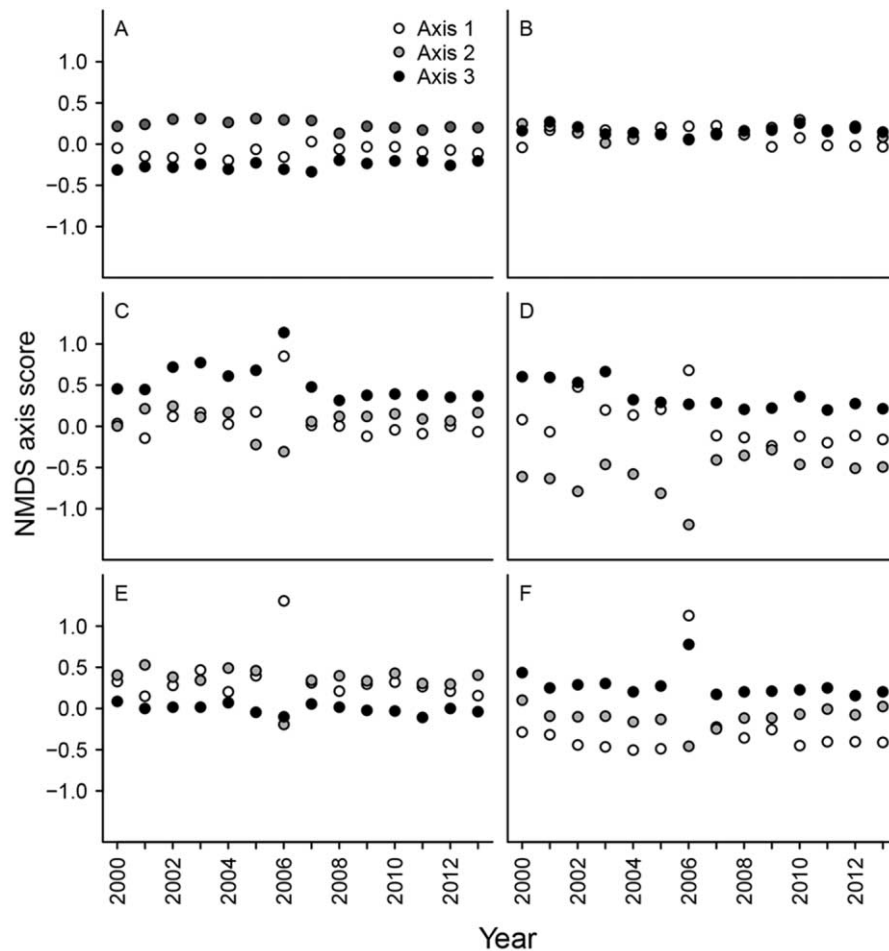


Figure 2. Nonmetric multidimensional scaling (NMDS) ordination scores for each study year for 6 sites representing negligible variation in community composition through time (A, B), erratic variation (C, D), and variation related to certain years (E, F). For all study sites, see Fig. S1.

effect of the drought year was much less evident, averaging out after 4 to 5 post-drought years (Fig. 3). For the trajectory approach, estimates of temporal turnover and, consequently, differences from the snapshot approach were site-specific, as indicated by variability among the random subsets for each study period (Fig. S2).

The drought year (2006) also had a disproportionate effect on the comparison (based on all study sites) of the non-overlapping sequences of 2, 5, or 7 y and of the whole study period of 14 y. The influence was strongest on 2-y sequences and resulted in higher turnover (mean = 0.334) than with longer data sets (means = 0.266–0.288) (Fig. S3A). Excluding year 2006, longer study durations (now 4, 7, and 13 y) showed no clear difference in the level of temporal turnover (Fig. S3B), and only the shortest sequence (2 y) differed from the 2 longest ones ( $F_{3,87} = 3.64$ ,  $p = 0.02$ ; Tukey's tests, 2 vs 7 y:  $p = 0.02$ ; 2 vs 13 y:  $p = 0.05$ ).

The 2 approaches rarely shared explanatory variables in the best models of temporal turnover (Fig. 5, Table 2), thus telling a partly different story about the drivers of commu-

nity variability. In the snapshot approach, water temperature was the most important explanatory variable. Temporal turnover typically increased with increasing temperature (i.e., the warmest sites supported the most variable invertebrate communities). However, study periods varied in both the identity and direction of change of the key environmental variables. In the trajectory approach, and especially for longer study durations, habitat connectivity was the most frequently selected explanatory variable. Temporal turnover decreased with increasing connectivity. The explanatory power and identity of the best explanatory factors varied strongly among random subsets, reflecting both small sample sizes and the influence of site identity. Nevertheless, patterns based on random subsets closely followed those based on all study sites (Table S1).

## DISCUSSION

Increasing the number of spatial and temporal replicates increases reliability of a study and makes the results more

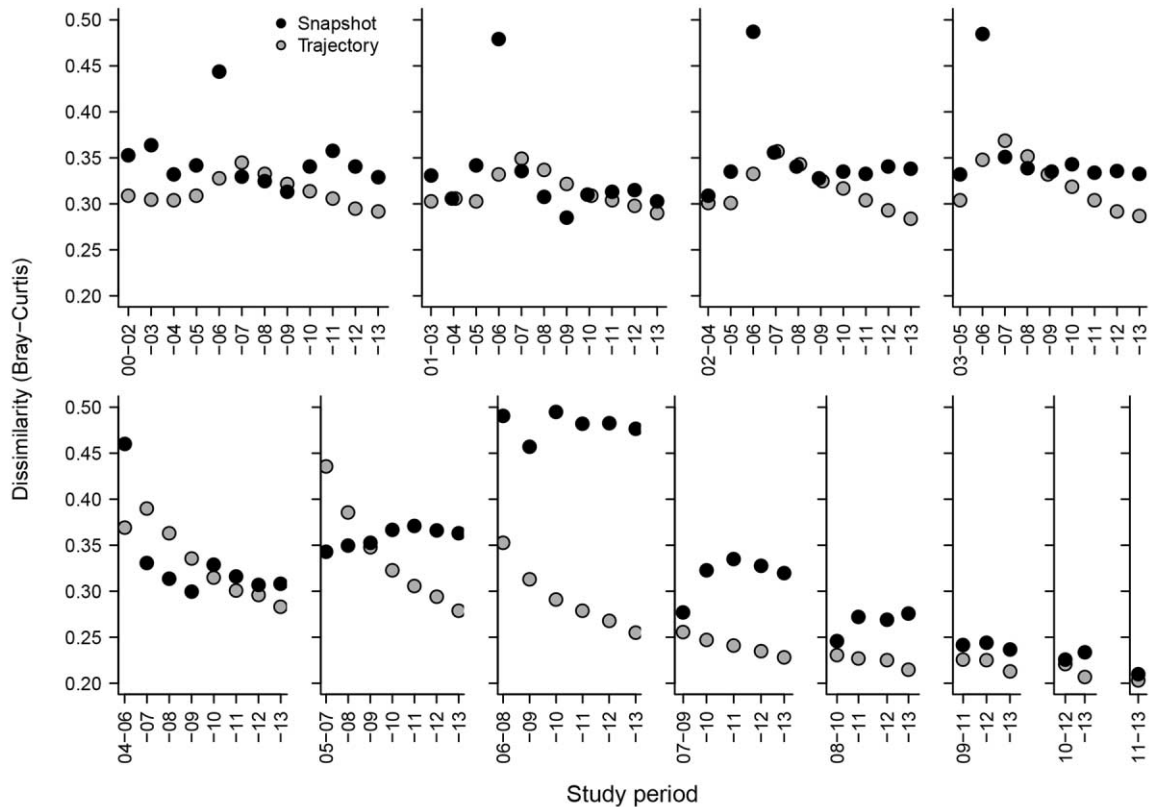


Figure 3. Bray–Curtis dissimilarity, i.e., temporal turnover, based on snapshot and trajectory approaches and equalized data sets for different study periods and durations. For the trajectory approach, mean was calculated for each study period as an average dissimilarity across 100 random subsets of sites (for deviation among subsets, see Fig. S2). Starting years range from 2000 to 2011, study duration from 3 (e.g., 2000–2002 and 2011–2013) to 14 y (2000–2013).

generalizable across time and space (Wiens 1981). However, resources available for sampling are often limited, forcing decisions on how to best allocate sampling events in space and time. We compared 2 approaches, trajectory vs snapshot, commonly used to assess the level of temporal variation in community composition. We tested whether the results were similar in terms of the level of temporal turnover and identity of the best explanatory factors, regardless of the approach, study period, and study duration. Our results demonstrate that different designs may yield different estimates of temporal turnover and factors explaining it, a possibility that should be carefully considered when comparing studies with different sampling designs.

The observed variation in community composition across time was generally higher with the snapshot than with the trajectory approach, but the 2 approaches yielded distinctly different outcomes only for study periods starting or ending in a climatically exceptional year with record-low summertime precipitation. This result highlights the importance of climatically exceptional years (Robinson et al. 2000, Huttunen et al. 2012) and, more generally, the major role of regional-scale climatic factors in driving stream community dynamics (e.g., Bradley and Ormerod 2001, Bêche and Resh 2007, Huttunen et al. 2014). However, of the hydrologically

exceptional years included in our study, only the drought year 2006 had a distinct effect on community composition that led to substantial differences between the 2 approaches. No major floods occurred during the 14-y study, and even the highest discharges, although clearly above long-term average, probably were too low to cause major shifts in species composition. High flows are considered less harmful than drought to aquatic communities because riverine organisms are likely to be better equipped for coping with an excess of water than with loss of it (Woodward et al. 2016). This drought effect may be particularly strong for boreal stream invertebrates, few of which possess any morphological or life-cycle traits to avoid the adverse effects of prolonged drought (Nilsson 1996). This vulnerability to drought is in sharp contrast with Mediterranean streams where wet and dry seasons alternate more or less predictably, favoring species with drought-tolerant traits (Bonada et al. 2007, Sarremejane et al. 2017). Our results suggest that to have a long-lasting impact on benthic communities and on monitoring outcomes in boreal streams, drought must be extreme in terms of both magnitude and duration.

The best explanatory factors related to temporal variation of macroinvertebrate communities differed between the snapshot and trajectory approaches. The fact that the observed

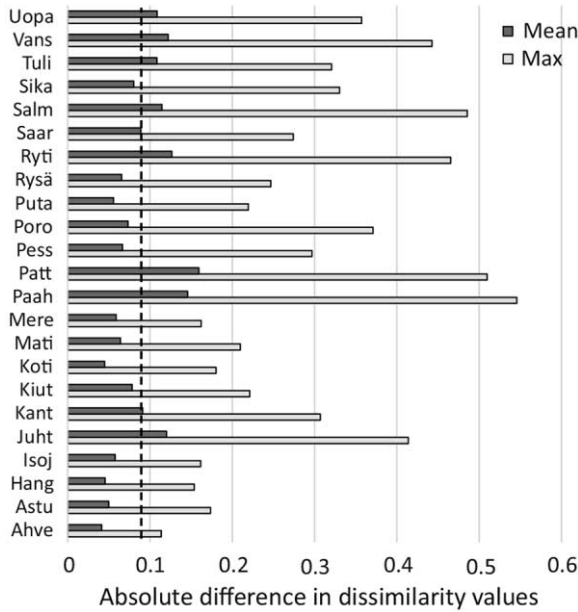


Figure 4. Site-specific mean and maximum values for the absolute differences in dissimilarity values between the snapshot and trajectory approach. The dashed line shows the mean across all sites. Max = maximum.

patterns of temporal turnover were related to different environmental variables is not problematic if one recognizes that the 2 approaches describe different aspects of community variability (Hildrew and Giller 1994). The snapshot approach provides a measure of the probability of finding a similar community in  $\geq 2$  y distant in time ('long-term stability'; Scarsbrook 2002), whereas the trajectory approach describes 'average stability' and provides a more comprehensive estimate of community variability. Thus, the approach chosen for a particular study reflects the definition of community variability or stability and the way it is being measured (Grimm and Wissel 1997, Donohue et al. 2016). For example, the probability of observing closely similar communities in 2 distant years (low temporal turnover) often increased with decreasing water temperature, whereas with the trajectory data, invertebrate communities were more stable at highly connected sites. However, connectivity contributes to community variability mainly via stochastic processes, particularly dispersal (Huttunen et al. 2017).

The identity and relative importance of the best explanatory variables also varied across study periods and durations, especially when the snapshot approach was used. With the trajectory approach, inconsistency in the identity of the best explanatory factors was evident only for short (<6 y) study durations. These results indicate that with limited temporal replication different environmental factors appear important, depending on which particular years are being compared, further supporting the view that temporally constrained sampling is more sensitive to individual events than are long-term or more intensive sampling protocols (Mac Nally et al. 2004).

Similarly, Vaughn and Young (2010) reported that the year of initiation may greatly affect the outcome of ecological field experiments (the 'year effect'). Conversely, we found that decreasing the number of spatial replicates with increasing study duration resulted in substantial variation among random subsets, highlighting the dependence of results on the identity of the sites included in a study.

An underlying assumption in bioassessment is that temporal variability of the monitored communities is negligible or at least significantly less than changes induced by anthropogenic stressors (Mazor et al. 2009). By showing the effect of individual years on monitoring outcomes, we join several previous authors (e.g., Mykrä et al. 2008, Mazor et al. 2009) in cautioning against neglecting interannual variation in community composition. In bioassessment, attempts to control interannual variation include, for example, resampling part of the reference sites in different years (Hargett et al. 2007), incorporating climatic variables as predictors into the models (Lawrence et al. 2010), and excluding the metrics that show the highest variability (Ofenböck et al. 2004). Uncertainty seems to be mostly related to extreme climatic conditions, so conclusions based on temporally constrained data should, at the very least, be supported

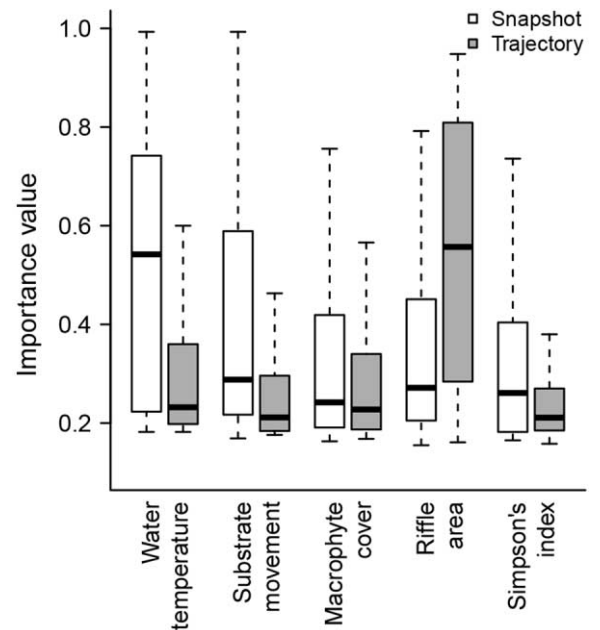


Figure 5. Box-and-whisker plot showing overall importance of explanatory variables among all candidate models explaining temporal turnover. Importance values for each variable were obtained by summing model weights across all models that included that variable. Each box plot is based on information from all possible study periods and durations. For the trajectory approach, only results based on all study sites are shown; patterns based on random subsets followed those based on all study sites (Table S1). Lines in boxes are medians, box ends are quartiles, and whiskers show the range.

Table 2. Standardized regression coefficients for the best models (i.e., the lowest Akaike Information Criterion with small sample correction [AIC<sub>c</sub>] score) explaining temporal turnover for different study periods (3, 6, 10, and 14 y) in the snapshot or trajectory approach based on the whole data set of 23 sites. ‘x’ denotes that a variable was not included in a model. See Table S1 for all study periods and equalized sampling effort.

Time span (y)	Study period	Snapshot										Trajectory (all sites)									
		Water temperature	Substrate movement	Macrophyte cover	Connectivity	Substrate diversity	Adj. R <sup>2</sup>	Water temperature	Substrate movement	Macrophyte cover	Connectivity	Substrate diversity	Adj. R <sup>2</sup>	Water temperature	Substrate movement	Macrophyte cover	Connectivity	Substrate diversity	Adj. R <sup>2</sup>		
3	2000–2002	0.604	x	-0.341	x	x	0.518	x	-0.432	-0.476	x	0.292	x	-0.432	-0.476	x	0.292	x	0.292		
	2001–2003	x	-0.322	x	-0.382	0.583	0.358	x	-0.429	-0.514	x	0.353	x	-0.429	-0.514	x	0.353	x	0.353		
	2002–2004	x	-0.453	-0.480	x	x	0.340	0.388	x	x	-0.387	0.310	x	x	-0.387	x	0.310	x	0.310		
	2003–2005	0.537	x	x	x	x	0.255	x	x	x	0.234	x	x	x	x	-0.519	x	0.234	0.234		
	2004–2006	x	x	0.355	-0.477	x	0.209	x	x	x	x	0.220	x	x	-0.506	x	0.220	x	0.220		
	2005–2007	x	-0.415	x	x	-0.336	0.294	x	x	x	x	0.000	x	x	x	x	0.000	x	0.000		
	2006–2008	x	x	0.332	x	x	0.068	x	x	x	0.384	0.107	x	x	x	x	0.107	x	0.107		
	2007–2009	-0.580	-0.629	x	x	x	0.274	-0.784	-0.573	x	x	0.406	x	x	x	x	0.406	x	0.406		
	2008–2010	x	-0.403	x	x	0.504	0.241	x	x	x	x	0.000	x	x	x	x	0.000	x	0.000		
	2009–2011	x	x	x	x	x	0.000	x	0.572	x	x	0.295	x	x	x	x	0.295	x	0.295		
6	2010–2012	x	x	x	x	0.000	x	0.374	x	x	0.099	x	x	x	x	x	0.099	x	0.099		
	2011–2013	x	x	x	x	0.000	x	x	x	-0.335	0.070	x	x	x	x	x	0.070	x	0.070		
	2000–2005	0.609	x	x	x	x	0.339	0.432	x	-0.368	0.320	x	x	x	x	x	0.320	x	0.320		
	2001–2006	x	x	x	x	x	0.000	0.317	x	x	0.386	x	x	x	x	-0.477	x	0.386	0.386		
	2002–2007	x	-0.615	x	x	x	0.348	x	x	x	0.278	x	x	x	-0.557	x	0.278	x	0.278		
	2003–2008	x	-0.396	x	-0.531	0.388	0.433	x	x	x	-0.483	0.196	x	x	-0.483	x	0.196	x	0.196		
	2004–2009	0.561	x	x	x	x	0.282	x	x	x	0.103	x	x	x	-0.379	x	0.103	x	0.103		
	2005–2010	0.544	x	x	x	x	0.263	x	x	x	0.000	x	x	x	x	x	0.000	x	0.000		
	2006–2011	x	x	x	x	-0.360	0.088	x	x	x	x	0.000	x	x	x	x	0.000	x	0.000		
	2007–2012	-0.697	-0.764	x	x	x	0.446	-0.449	x	x	0.164	x	x	x	x	x	0.164	x	0.164		
10	2008–2013	x	-0.525	x	x	x	0.241	x	0.348	x	0.079	x	x	x	x	x	0.079	x	0.079		
	2000–2009	0.472	x	x	x	x	0.183	x	x	-0.526	0.241	x	x	x	-0.526	x	0.241	x	0.241		
	2001–2010	x	x	x	x	x	0.000	x	x	x	0.236	x	x	x	-0.520	x	0.236	x	0.236		
	2002–2011	0.403	x	-0.338	x	x	0.277	x	x	x	0.169	x	x	x	-0.455	x	0.169	x	0.169		
	2003–2012	0.414	x	-0.425	x	x	0.381	x	x	x	0.138	x	x	x	-0.421	x	0.138	x	0.138		
14	2004–2013	0.611	x	x	x	x	0.344	x	x	x	0.000	x	x	x	x	x	0.000	x	0.000		
	2000–2013	0.471	x	x	x	x	0.183	x	x	-0.471	0.183	x	x	x	-0.471	x	0.183	x	0.183		



by evidence that the years being compared do not deviate strongly in terms of climatic and hydrological conditions.

What is then the optimal duration and sampling frequency for addressing temporal variability in benthic community composition? Furthermore, how should one distribute limited resources between spatial and temporal replication? Giving a universal recommendation other than “sampling decisions must be dictated by the study objectives” (Kenkel et al. 1989) would be unwise. Some study designs may benefit from pilot surveys or a priori power analysis to tune an optimal sampling effort, but these options are not always feasible. Piloting takes its share of limited resources, and no guarantee exists that a piloting period will include any extreme events. Mac Nally et al. (2004) recommended that, with limited resources, expanding study length is usually better than increasing sampling frequency. Baasch et al. (2010) also concluded that one generally should favor the long-term perspective over sampling frequency or sample size. These 2 studies differ from ours. The first concentrated generally on temporal variability in species composition (invasive plants and butterflies), the second on directional change (vegetation succession), whereas our focus was on the level of temporal turnover of invertebrate communities in near-pristine streams. Based on our study, the key challenge for effective sampling designs is the year effect, particularly as it relates to climatically exceptional years. In boreal streams with little human influence, time sequences of ~6 y of annual data (i.e., 6 generations for most of the taxa in our data) seem to be enough to detect long-term patterns of benthic community variability. This period matches closely Jackson and Füreder’s (2006) recommendation for 5 y as the minimum length for a study to encompass a meaningful range of climatically variable conditions. Our study period included only 1 climatically exceptional year that had an ecologically meaningful effect, and the recovery of communities even from such a rare event was rather fast. The effect of extreme events could be more dramatic if high- and low-flow years occurred more frequently without time for communities to recover (Woodward et al. 2016). This situation is exactly what various climate-change scenarios predict for the hydrological regime of boreal streams (Nilsson et al. 2015, Mustonen et al. 2018). Thus, the year effect is likely to be a bigger issue for snapshot and other short-term surveys in the future.

Wide consensus exists about the importance of biological long-term data, but availability of data sets that are representative spatially, temporally, and across different gradients (e.g., anthropogenic stress) is still limited. Our results are based on near-pristine streams. In human-affected streams, temporal variability can be much greater, with community composition varying rather erratically between years (Feio et al. 2010, Huttunen et al. 2012). Alternatively, stream communities in historically human-disturbed areas may undergo directional changes (Angeler and Johnson 2012). In both cases, community predictability may be much lower than in pristine streams.

This lack of predictability further underlines the undue influence of particular years on the assessment of temporal community variability, especially when the snapshot approach is used. Our results call for great caution when basing conclusions about interannual variability of communities on a comparison of a few (or just 2) occasions several years apart or on very short time sequences, at least until we know more about community variability in human-modified systems based on spatially replicated long-term data sets.

## ACKNOWLEDGEMENTS

Author contributions: K-LH was responsible for study design, data analysis, and writing of the manuscript, and she also participated in the fieldwork and species identification. HM assisted with study design and writing. RP contributed to data collection and species identification. TM supervised the work and contributed to study design and writing.

We thank several students and colleagues for their help in the field or in the laboratory along the years of data collection. We acknowledge logistic support by the Oulanka Research Station during the fieldwork. Financial support was received from the Academy of Finland, Maj and Tor Nessling foundation, and University of Oulu (Thule institute and University of Oulu Graduate School). We gratefully acknowledge Raphael Mazor and an anonymous referee for comments that greatly improved our manuscript.

## LITERATURE CITED

- Angeler, D. G., and R. K. Johnson. 2012. Temporal scales and patterns of invertebrate biodiversity dynamics in boreal lakes recovering from acidification. *Ecological Applications* 22:1172–1186.
- Baasch, A., S. Tischew, and H. Bruelheide. 2010. How much effort is required for proper monitoring? Assessing the effects of different survey scenarios in a dry acidic grassland. *Journal of Vegetation Science* 21:876–887.
- Banks-Leite C., R. M. Ewers, R. G. Pimentel, and J. P. Metzger. 2012. Decisions on temporal sampling protocol influence the detection of ecological patterns. *Biotropica* 44:378–385.
- Barton, K. 2012. MuMIn: Multi-model inference. R package version 1.7.11. R Project for Statistical Computing, Vienna, Austria. (Available from: <https://CRAN.R-project.org/package=MumIn>)
- Bêche, L. A., E. P. McElravy, and V. H. Resh. 2006. Long-term seasonal variation in the biological traits of benthic-macroinvertebrates in two Mediterranean-climate streams in California, U.S.A. *Freshwater Biology* 51:56–75.
- Bêche, L. A., and V. H. Resh. 2007. Short-term climatic trends affect the temporal variability of macroinvertebrates in California ‘Mediterranean’ streams. *Freshwater Biology* 52:2317–2339.
- Bengtsson, J., S. R. Baillie, and J. Lawton. 1997. Community variability increases with time. *Oikos* 78:249–256.
- Bonada, N., S. Dolédec, and B. Statzner. 2007. Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: implica-

- tions for future climatic scenarios. *Global Change Biology* 13:1658–1671.
- Bradley, D. C., and S. J. Ormerod. 2001. Community persistence among stream invertebrates tracks the North Atlantic Oscillation. *Journal of Animal Ecology* 70:987–996.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.
- Collins, S. L., F. Micheli, and L. Hartt. 2000. A method to determine rates and patterns of variability in ecological communities. *Oikos* 91:285–293.
- Donohue, I., H. Hillebrand, J. M. Montoya, O. L. Petchey, S. L. Pimm, M. S. Fowler, K. Healey, A. L. Jackson, M. Lurgi, D. McClean, N. E. O'Connor, E. J. O'Gorman, and Q. Yang. 2016. Navigating the complexity of ecological stability. *Ecology Letters* 19:1172–1185.
- Feio, M. J., C. N. Coimbra, M. A. S. Graça, S. J. Nichols, and R. H. Norris. 2010. The influence of extreme climatic events and human disturbance on macroinvertebrate community patterns of a Mediterranean stream over 15 y. *Journal of the North American Benthological Society* 29:1397–1409.
- Grimm, V., and C. Wissel. 1997. Babel, or the ecological stability discussions: an inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia* 109:323–334.
- Hargett, E. G., J. R. ZumBerge, C. P. Hawkins, and J. R. Olson. 2007. Development of a RIVPACS-type predictive model for bioassessment of Wadeable streams in Wyoming. *Ecological Indicators* 7:807–826.
- Hildrew, A. G., and P. S. Giller. 1994. Patchiness, species interactions and disturbance in the stream benthos. Pages 21–62 in A. G. Hildrew, P. S. Giller, and D. G. Raffaelli (editors). *Aquatic ecology: scale, pattern and process*. Blackwell Science, Oxford, UK.
- Huttunen, K.-L., H. Mykrä, A. Huusko, A. Mäki-Petäys, T. Vehanen, and T. Muotka. 2014. Testing for temporal coherence across spatial extents: the roles of climate and local factors in regulating stream macroinvertebrate community dynamics. *Ecography* 37:599–608.
- Huttunen, K.-L., H. Mykrä, and T. Muotka. 2012. Temporal variability in taxonomic completeness of stream macroinvertebrate assemblages. *Freshwater Science* 31:423–441.
- Huttunen, K.-L., H. Mykrä, J. Oksanen, A. Astorga, R. Paavola, and T. Muotka. 2017. Habitat connectivity and in-stream vegetation control temporal variability of benthic invertebrate communities. *Scientific Reports* 7:1448.
- Jackson, J. K., and L. Füreder. 2006. Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. *Freshwater Biology* 51:591–603.
- Johnson, P. D., K. M. Brown, and C. V. Covell. 1994. A comparison of the macroinvertebrate assemblage in Doe Run Creek, Kentucky: 1960 and 1990. *Journal of the North American Benthological Society* 13:496–510.
- Jourdan, J., R. B. O'Hara, R. Bottarin, K.-L. Huttunen, M. Kuemmerlen, D. Monteith, T. Muotka, D. Ozoliņš, R. Paavola, F. Pilotto, G. Springe, A. Skuja, A. Sundermann, J. D. Tonkin, and P. Haase. 2018. Effects of changing climate on European stream invertebrate communities: a long-term data analysis. *Science of The Total Environment* 621:588–599.
- Kenkel, N. C., P. Juhász-Nagy, and J. Podani. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83:195–207.
- Lawrence, J. E., K. B. Lunde, R. D. Mazon, L. A. Bèche, E. P. McElravy, and V. H. Resh. 2010. Long-term macroinvertebrate responses to climate change: implications for biological assessment in mediterranean-climate streams. *Journal of the North American Benthological Society* 29:1424–1440.
- Mac Nally R., E. Fleishman, and D. D. Murphy. 2004. Influence of temporal scale of sampling on detection of relationships between invasive plants and the diversity patterns of plants and butterflies. *Conservation Biology* 18:1525–1532.
- Magurran A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. J. Somerfield, and A. D. Watt. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology and Evolution* 25:574–582.
- Malmqvist, B., T. Muotka, C. Nilsson, and H. Timm. 2009. The Fennoscandian shield. Pages 297–336 in U. Tockner, U. Uehlinger, and C. T. Robinson (editors). *Rivers of Europe*. Elsevier, San Diego, California.
- Mazon, R. D., A. H. Purcell, and V. H. Resh. 2009. Long-term variability in bioassessments: a twenty-year study from two northern California streams. *Environmental Management* 43:1269–1286.
- Mustonen, K.-R., H. Mykrä, H. Marttila, R. Sarremejane, N. Veijalainen, K. Sippel, T. Muotka, and C. P. Hawkins. 2018. Thermal and hydrologic responses to climate change predict marked alterations in boreal stream invertebrate assemblages. *Global Change Biology* 24:2434–2446.
- Mykrä, H., T. Ruokonen, and T. Muotka. 2006. The effect of sample duration on the efficiency of kick-sampling in two streams with contrasting substratum heterogeneity. *Verhandlungen der Internationalen Vereinigung für theoretische und angewandte Limnologie* 29:1351–1355.
- Mykrä, H., J. Heino, and T. Muotka. 2008. Concordance of stream macroinvertebrate assemblage classifications: how general are patterns from single-year surveys? *Biological Conservation* 141:1218–1223.
- Nilsson A. (editor). 1996. *Aquatic insects of north Europe: a taxonomic handbook*. Apollo Books, Stenstrup, Denmark.
- Nilsson, C., L. E. Polvi, and L. Lind. 2015. Extreme events in streams and rivers in arctic and subarctic regions in an uncertain future. *Freshwater Biology* 60:2535–2546.
- Ofenböck, T., O. Moog, J. Gerritsen, and M. Barbour. 2004. A stressor specific multimetric approach for monitoring running waters in Austria using benthic macroinvertebrates. *Hydrobiologia* 516:251–268.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2014. *Vegan: community ecology package*. R package version 2.2-0. R Project for Statistical Computing, Vienna, Austria. (Available from: <http://cran.r-project.org>)
- Robinson, C. T., G. W. Minshall, and T. V. Royer. 2000. Interannual patterns in macroinvertebrate communities of wilderness streams in Idaho, USA. *Hydrobiologia* 421:187–198.
- Rosenberg, D. M., and V. H. Resh. 1993. *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman and Hall, New York.

- Sarremejane, R., M. Cañedo-Argüelles, N. Prat, H. Mykrä, T. Muotka, and N. Bonada. 2017. Do metacommunities vary through time? Intermittent rivers as model systems. *Journal of Biogeography* 44:2752–2763.
- Scarsbrook, M. R. 2002. Persistence and stability of lotic invertebrate communities in New Zealand. *Freshwater Biology* 47:417–431.
- Sponseller, R. A., N. B. Grimm, A. J. Boulton, and J. L. Sabo. 2010. Responses of macroinvertebrate communities to long-term flow variability in a Sonoran Desert stream. *Global Change Biology* 16:2891–2900.
- Townsend, C. R., A. G. Hildrew, and K. Schofield. 1987. Persistence of stream invertebrate communities in relation to environmental variability. *Journal of Animal Ecology* 56:597–613.
- Townsend, C. R., M. R. Scarsbrook, and S. Dolédec. 1997. Quantifying disturbance in streams: alternative measures of disturbance in relation to macroinvertebrate species traits and species richness. *Journal of the North American Benthological Society* 16:531–544.
- Vaughn, K. J., and T. P. Young. 2010. Contingent conclusions: year of initiation influences ecological field experiments, but temporal replication is rare. *Restoration Ecology* 18:59–64.
- Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global threats to human water security and river biodiversity. *Nature* 467:555–561.
- Wiens, J. A. 1981. Single-sample surveys of communities: are the revealed patterns real? *American Naturalist* 117:90–98.
- Woodward, G., N. Bonada, L. E. Brown, R. G. Death, I. Durance, C. Gray, S. Hladyz, M. E. Ledger, A. M. Milner, S. J. Ormerod, R. M. Thompson, and S. Pawar. 2016. The effects of climatic fluctuations and extreme events on running water ecosystems. *Philosophical Transactions of the Royal Society London Series B: Biological Sciences* 371:20150274.
- Woodward, G., J. I. Jones, and A. G. Hildrew. 2002. Community persistence in Broadstone Stream (U.K.) over three decades. *Freshwater Biology* 47:1419–1435.