This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

Feel free to email me at yuichiwsk@gmail.com

Predicting macroinvertebrate average score per taxon (ASPT) at water quality monitoring sites in Japanese rivers

Yuichi Iwasaki[1,*], Tomomi Suemori[1], Yuta Kobayashi[2]

[1] Research Institute of Science for Safety and Sustainability, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan

[2] Field Science Center, Faculty of Agriculture, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-tyo, Fuchu, Tokyo, Japan

*Corresponding author. E-mail: yuichiwsk@gmail.com, yuichi-iwasaki@aist.go.jp Tel: +81-29-861-4263

Abstract

Freshwater ecosystems provide essential services for human well-being but are impacted by multiple anthropogenic stressors. Biomonitoring with bioindicators such as river macroinvertebrates is fundamental for assessing the status of freshwater systems. In Japan, water quality and biomonitoring surveys are conducted separately, leading to a lack of nationwide information on the biological status of water quality monitoring (WQM) sites. In this study, we examined the co-occurrence of 983 biomonitoring sites with WQM sites to obtain a set of 237 "aligned" sites. Then, we developed a multiple linear regression model to estimate the average score per taxon (ASPT) from river macroinvertebrate data surveyed at these sites. The best model (i.e., with the smallest corrected Akaike

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

information criterion) included eight predictors: elevation, catchment area, biological oxygen demand, suspended solids, minimum pH, the proportions of paddy fields and urban areas in the catchment, and the proportion of urban areas within a 3-km radius. The best multiple linear regression model could predict ASPT with reasonable accuracy, i.e., with an error of ±1 for 96% of the aligned data (R^2 = 0.69; root mean squared error = 0.47) and 84% of the external validation dataset (R^2 = 0.55; root mean squared error = 0.75). Using the best multiple linear regression model, we estimated ASPT values at 2925 WQM sites in rivers nationwide. Although caution should be exercised because of uncertainties in the estimation, the WQM sites were categorized into four levels of river environment quality by estimated ASPT values: "very good" (29% of WQM sites), "good" (50%), "fairly good" (14%), and "not good" (8%). Furthermore, we observed statistically significant correlations (p < 0.05; 0.4 ≤ r ≤ 0.7) between ASPT and all eight macroinvertebrate metrics examined, such as mayfly (Ephemeroptera) and stonefly (Plecoptera) richness, providing valuable information on the ecological implications of changes in ASPT. Our study provides a valuable statistical model for estimating ASPT and contributes to further understanding of the biological status of rivers across Japan.

Keywords

Aquatic insect, Invertebrate, Water pollution, Land use, Bioindicator, Biomonitoring

Introduction

Freshwater ecosystems are essential for human well-being and provide vital material, non-material, and regulating services such as food, recreation, and water purification

(Lynch et al. 2023). However, these ecosystems are facing various anthropogenic stressors such as climate change, land-use change, and water pollution (Birk et al. 2020, IPBES 2019, Persson et al. 2022, Reid et al. 2019, Waite et al. 2021). In this context, biomonitoring with bioindicators such as algae, macroinvertebrates, and fish has a fundamental role in capturing the biological status of streams, rivers, and other freshwater systems (Aroviita et al. 2010, Barbour et al. 1999, Birk et al. 2012, Buss et al. 2014, Namba et al. 2020, Niemi &McDonald 2004, Wright 2000). Biomonitoring results can be used to assess and identify any adverse ecosystem impacts, contributing to conservation and sustainable management. This is particularly important for rivers because they are more often affected by multiple stressors than lakes (Birk et al. 2020). Consequently, bespoke management solutions are generally required to address the specific challenges faced by river ecosystems (Birk et al. 2020, Iwasaki et al. 2018).

Water pollution is a critical stressor affecting aquatic ecosystems, and streams and rivers are particularly susceptible to high pollution levels because of their limited capacity for dilution (Büttner et al. 2022, Johnson et al. 2020). In Japan, water quality monitoring has been conducted at approximately 6000 river sites nationwide to assess water quality variables including suspended solids (SS) and biochemical oxygen demand (BOD) (Iwasaki et al. 2022). In addition, the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) has initiated a nationwide biomonitoring program called the National Census on the River Environment (NCRE) to assess the biological and ecological status of rivers. The biomonitoring program involves the sampling of various taxa, including fish, benthic invertebrates, plants, birds, terrestrial insects, amphibians, reptiles, and mammals, from over 240 rivers across Japan at 5- or 10-year intervals (Feio et al. 2021). However, because water-quality and biomonitoring surveys are not necessarily conducted at the same locations, the biological status of water quality monitoring sites (hereafter, WQM sites) is largely unknown across Japan. This kind of large-scale comprehensive data on biological status is crucial for informing effective bespoke management strategies aimed at mitigating water pollution and safeguarding aquatic ecosystems (Abell et al. 2008).

Numerous biological metrics have been developed for biomonitoring and bioassessment (Birk et al. 2012, Eriksen et al. 2021). For Japanese rivers, the average score per taxon (ASPT) based on the occurrence of 71 macroinvertebrate taxa (mainly, families) is the only biological metric developed at the national level for assessing river health (MoE 2017, Nozaki 2012). The ASPT is calculated as follows:

$$\mathrm{ASPT} = \frac{\sum \mathrm{Score}_i}{\text{Total number of scoring taxa present}}, \qquad \text{Equation (1)}$$

where Score_i is the score assigned to macroinvertebrate taxon i, which is expected to represent the value inversely proportional to the taxon's perceived tolerance to water pollution (Yamasaki et al. 1996). ASPT was originally designed as a biological indicator of water quality (specifically, organic pollution) but is also correlated with the impacts of other anthropogenic pressures such as land-use change (Eriksen et al. 2021, Yamasaki et al. 1996). However, the relationships between ASPT and commonly used macroinvertebrate metrics, such as EPT richness (the total number of taxa in Ephemeroptera, Plecoptera, and Trichoptera), remain unexplored in Japanese rivers nationwide. Investigating these relationships could yield valuable insights into the implications of changes in ASPT and their ecological significance.

Thus, we first aimed to predict ASPT at 2925 WQM sites, which serve as environmental reference points that are officially used to assess the compliance/exceedance of environmental water quality standards in Japan. To achieve this, we examined the co-occurrence of NCRE biomonitoring sites with WQM sites to obtain a total of 237 “aligned” sites (i.e., those at which both macroinvertebrate and water quality monitoring data were available; Fig. 1; see Fig. 2 for our data analysis flowchart). We then developed a multiple regression model with physicochemical variables representing water quality, land use, and other factors to predict ASPT. The application of this model to all 2925 WQM sites would offer a more comprehensive understanding of biological status at streams and rivers across Japan. Furthermore, we investigated the relationships between variations in ASPT and variations in macroinvertebrate metrics such as EPT richness and mayfly richness at the aligned study sites.

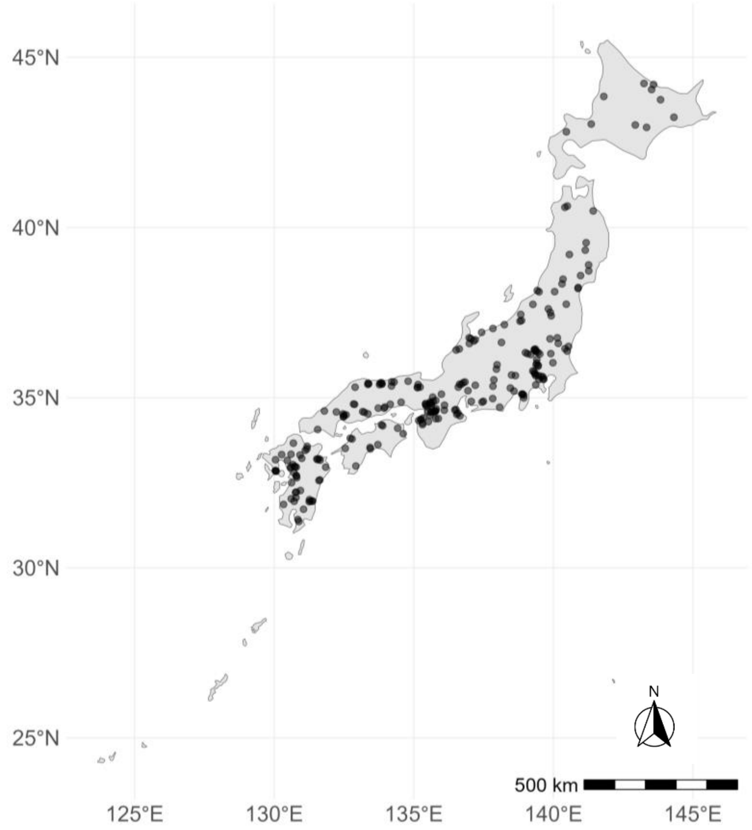

Fig. 1. Map of the 237 aligned study sites (filled circles).

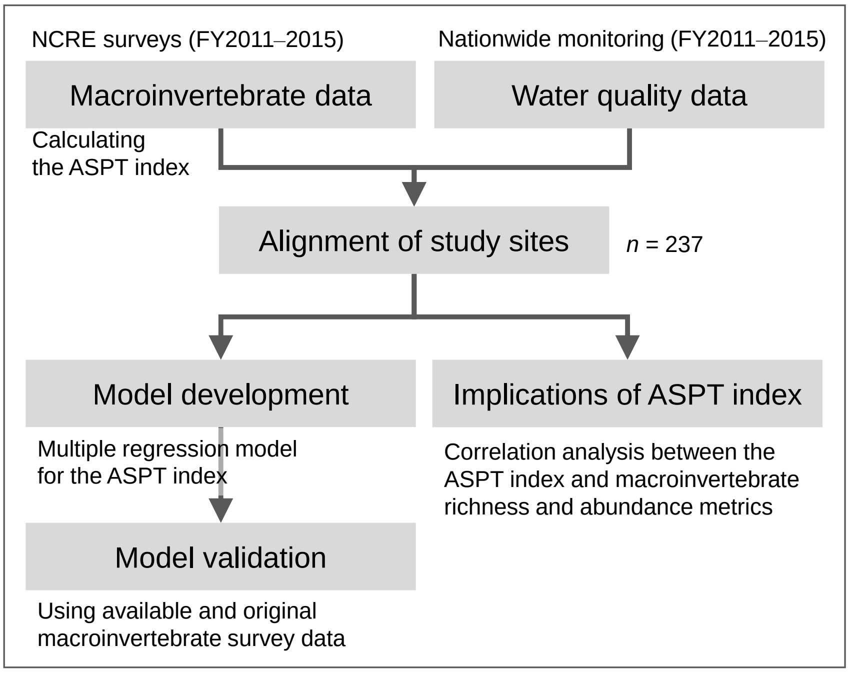

Fig. 2. Schematic diagram of the methodology used.

NCRE, National Census on the River Environment; ASPT, average score per taxon.

Materials and Methods

Alignment of study sites

We first extracted “aligned” study sites at which both macroinvertebrate and water quality monitoring data were available. The geographic coordinates (latitude and longitude) for a total of 983 river sites where quantitative macroinvertebrate sampling was conducted between April 2011 and March 2016 (fiscal years 2011–2015) were compiled by obtaining the original information, including results of macroinvertebrate monitoring from MLIT. Similarly, the geographic coordinates of a total of 5855 WQM sites where sampling was performed during fiscal years 2011–2015 were obtained from the Comprehensive Information Website for Water Environment (https://water-pub.env.go.jp/water-pub/mizu-site/, accessed November 21, 2023).

We then identified suitable pairs of biomonitoring and WQM sites to be matched for later analysis. Initially, we linked each biomonitoring site to the closest WQM site within a Euclidean distance of 2 km. We then examined the suitability of each linked site pair by using Google Earth Pro version 7.3 (https://www.google.com/earth/about/, accessed November 21, 2023). If the two sites were not identical, their suitability was assessed by considering whether the two sites were located within the same river, the absence of inflow of major tributaries and changes in land use between the sites, and the availability of other more suitable WQM sites. While we carefully matched the biomonitoring and WQM sites, it is impossible to guarantee the absence of any significant changes in physicochemical characteristics, such as pollution from an unknown point source, between the two sites. However, the inclusion of these few cases should not have been materially affected our findings. During this assessment process, we identified multiple sites with inaccurate geographic coordinates and made necessary corrections based on the available information such as river and site names. From this process, 409 pairs of biomonitoring and WQM sites were selected.

The 409 pairs of biomonitoring and WQM sites were further winnowed based on the following criteria: (1) quantitative sampling of macroinvertebrates was conducted using a Surber sampler with a 25 × 25 cm quadrat at each biomonitoring site during cold seasons (October to March), (2) macroinvertebrates were collected from cobble/gravel-dominated lotic environments (i.e., runs or riffles), (3) no macroinvertebrate species typical of estuarine environments were collected (to exclude sites influenced by salinity), and (4) all water quality variables used in the multiple

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

regression model (see below) were available. In total, 237 pairs of biomonitoring and WQM sites were selected for model development (Fig. 1).

Macroinvertebrates and ASPT

All field sampling and laboratory analysis (including sorting and identification) of river benthic macroinvertebrates were conducted by following the NCRE's Basic Survey Manual (MLIT 2016). Each of the three macroinvertebrate samples collected from riffles or runs at individual biomonitoring sites was washed through a 0.5-mm mesh sieve, and macroinvertebrates remaining on the sieve were sorted and identified generally to species or genus level. We calculated site averages of ASPT based on the presence of individual macroinvertebrate taxa and corresponding scores (Equation 1; see Table S1 for the scores). The scores used in the present study were initially developed by Yamasaki et al. (1996) and revised by the committee launched by the Ministry of the Environment, Japan (MoE 2017).

Sites are categorized into four groups of relative river environmental quality based on ASPT values (MoE 2017): "very good" (7.5 and above), "good" (6.0–7.5), "fairly good" (5.0–6.0), and "not good" (below 5.0). It should be noted that the calculation of ASPT (MoE 2017) involves 3 min of kick sampling (1 min of kick sampling at three locations per site) using a D-frame net in riffles or/and runs as well as the suggested use of a sieve with a mesh size of about 1 mm to filter macroinvertebrate samples. These methods differ from those employed in NCRE biomonitoring (MLIT (2016); see also above). However, we adopted the four categories described above despite these differences because, in general, the coarse taxonomic level (i.e., the family level) used to calculate ASPT likely mitigates any influence of methodological differences in the diagnostic evaluation (Armitage et al. 1983, Eriksen et al. 2021, Hawkes 1998; see also the section "*Model for predicting ASPT*").

Furthermore, to investigate relationships between APST and macroinvertebrate metrics, we calculated total taxon richness and total abundance as well as the taxon richness and abundance of three major insect groups (Ephemeroptera, Plecoptera, and Trichoptera). Correlations were examined by calculating Pearson product-moment correlation coefficients (r).

Physicochemical characteristics of sites

For the 237 aligned study sites, we complied a total of 11 physicochemical characteristics, which were used as predictors in the multiple regression model. These characteristics had been previously estimated for most of the aligned study sites by Iwasaki et al. (2022). However, 67 WQM sites corresponding to biomonitoring sites were not included in the 2925 WQM sites analyzed by Iwasaki et al. (2022). Thus, we obtained the physicochemical characteristics of these sites by using the methods described in Iwasaki et al. (2022). For the water quality variables included in the multiple regression model, we calculated 5-year averages of minimum pH, 5-day biochemical oxygen demand (BOD; mg/L), and suspended solids (SS; mg/L) at each WQM site. These averages were derived from measurements taken during fiscal years 2011–2015, using the available data in the Comprehensive Information Website for Water Environment. For pH, only minimum and maximum values for each fiscal year were available in the database, so we used the minimum pH as an indicator of river acidity. The 5-year averages of the water quality variables were used as representative values reflecting the general water quality conditions at individual WQM sites during the period that included the timing of macroinvertebrate sampling, although they may not accurately reflect the water quality at the specific time of the macroinvertebrate sampling (see the section "*Model development and validation*" for more discussion about temporal changes).

The catchment area (km^2) of each WQM site was estimated by delineating each catchment based on 30 × 30 m raster data of hydrologically adjusted elevations (Japan Flow Direction Map version 1.0; Yamazaki et al. 2018) using ArcGIS Pro (ESRI, version 2.6.0). In addition, the land uses both in the catchment areas and within a 3-km radius were estimated. Specifically, the proportions of urban areas, paddy fields, and dry fields (cropland) were derived as indicators of anthropogenic disturbance based on the High-Resolution Land Use and Land Cover Map (2014–2016; version 18.03) provided by the Japan Aerospace Exploration Agency (https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc_e.htm, accessed November 21, 2023). The land uses within a 3-km radius were included as an indicator of land use patterns in the immediate vicinity of WQM sites, although downstream land uses are unlikely to directly affect water quality or the biological status at WQM sites given natural flow patterns. Additionally, the average elevation (m) within a 100-m radius was calculated for each WQM site. Because of the high correlation between the proportions of forest and urban areas ($r = -0.85$), we chose not to include the proportion of forest as a predictor in the multiple regression model (see below).

Model development and validation

To predict ASPT, we employed multiple linear regression models with a normal error distribution and a total of 11 predictors, including the three water quality variables (minimum pH, BOD, and SS), catchment area, elevation, and six land use variables. Catchment area, elevation, BOD, and SS were $\log_{10}$-transformed to reduce skewness before analysis. Model selection was performed by using the Akaike information criterion corrected for small sample size (AICc; Burnham andAnderson 2004, Burnham et al. 2011), and the model with the lowest AICc was selected as the best among all possible models (i.e., 2048 models) considered. The coefficient of determination (R^2) and root mean square error (RMSE) were calculated to evaluate goodness of fit for the best model selected. All data processing and statistical analyses were performed in R version 4.2.0 (R Core Team 2022), and the model selection was carried out by using the function “dredge” in the “MuMIn” library (Bartoń 2022). We also employed a random forest, which is a machine learning algorithm (Ryo &Rillig 2017), as a preliminary analysis to model ASPT. However, during external validation, the best multiple linear regression model outperformed the random forest model, likely because of the limited coverage of the model development data.

To perform external validation for the best model using an entirely new dataset, we examined macroinvertebrate survey data collected at 75 river sites that were at or near WQM sites (environmental reference points) across Japan (Table S2; Fig. S1). Most of these surveys were conducted using a D-frame net following the sampling method described in MoE (2017) (see Table S2 for more details). Similarly, we conducted our own macroinvertebrate surveys at 28 WQM sites in different regions. Specifically, we surveyed 10 WQM sites in the Tohoku area (Iwate, Miyagi, and Fukushima prefectures) in January 2021, 9 WQM sites in Aichi prefecture in January 2022, and 9 WQM sites in Hokkaido (the city of Sapporo) in December 2022 (see Fig. S1 for map). These surveys were performed by following the NCRE’s Basic Survey

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

Manual (MLIT 2016). We then compared ASPT values predicted from the best model to those calculated from the macroinvertebrate survey data by assessing two diagnostic metrics, R^2 and RMSE. The timing of the surveys used for model validation (especially those conducted for this study) did not coincide with the data period (2011–2015) used for the multiple regression model. However, we assumed that marked changes in land use and water quality variables during these periods were relatively unlikely for the majority of WQM sites as compared to past changes observed between the 1980s and 2010s (see, e.g., Ye and Kameyama 2020) for temporal changes in SS and BOD). Given these variations in sampling methods and timing, validating the best model with the external datasets would be valuable to test the robustness of the best model as well as use of ASPT for the prediction. All data and R code used are available from a GitHub repository at https://github.com/yuichiwsk/predict_ASPT_Japan.

Results and Discussion

Model for predicting ASPT

The best multiple linear regression model included 8 predictors: elevation, catchment area, BOD, SS, minimum pH, the proportions of paddy fields (%Paddy) and urban areas (%Urban) in the catchment, and the proportion of urban areas within a 3-km radius (%Urban-3km) (Table 1). All predictors except catchment area and minimum pH were included in all of the top 10 models (see Table S3). The negative regression coefficients of BOD, %Urban, SS, %Paddy, and %Urban-3km in the best model, along with the positive coefficient of minimum pH, are all consistent with the expected adverse impacts of these factors on macroinvertebrates as reported in previous studies (Iwasaki et al. 2018, Larsen et al. 2009, Ormerod &Durance 2009, Roy et al. 2003, Schmidt et al.

2019, Waite et al. 2019), although inferring causal relationships is beyond the scope of the present study (Takeshita et al. 2022). The positive regression coefficient of elevation aligns with the general expectation that upland river sites at higher elevation would have lower water temperature (resulting in higher dissolved oxygen essential for aquatic organisms) and be less impacted by anthropogenic factors and disturbances. However, correlations between elevation and other predictor variables included in the present study were not evident ($|r| \leq 0.29$). Additionally, the positive yet nonsignificant coefficient of catchment area (an indicator of the magnitude of river discharge) might be associated with dilution capacity (Büttner et al. 2022, Johnson et al. 2020) for water pollution other than BOD and SS, although other factors, such as natural longitudinal changes in benthic macroinvertebrate communities in response to environmental conditions (Vannote et al., 1980), cannot be excluded.

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

Table 1. Estimated intercepts and coefficients of the best multiple linear regression model

Predictors	Estimates (SE)	Standardized Coefficients	p value
Intercept	5.65 (0.77)	NA	<0.001
Elevation	0.21 (0.06)	0.15	<0.001
Catchment area	0.09 (0.06)	0.07	0.110
BOD	–1.73 (0.21)	–0.41	<0.001
SS	–0.36 (0.11)	–0.14	0.002
Minimum pH	0.19 (0.11)	0.07	0.073
%Paddy	–0.021 (0.006)	–0.14	0.001
%Urban	–0.016 (0.003)	–0.27	<0.001
%Urban-3km	–0.004 (0.002)	–0.12	0.011

SE, standard error; NA, not available; BOD, biochemical oxygen demand; SS, suspended solids; %Paddy, proportion of paddy fields in the catchment; %Urban, proportion of urban areas in the catchment; %Urban-3km, proportion of urban areas within a 3-km radius. See the text for more details about predictors.

The best multiple linear regression model estimated ASPT values with an error of ±0.5 for 76% of the aligned data and an error of ±1 for 96% of the aligned data (R^2 = 0.69, RMSE = 0.47; Fig. 3a). There were two WQM sites where the observed ASPT value was lower than the predicted value by 1.5 or more. These two sites had catchments larger than 1500 km^2 that were predominantly covered by forest (>70%), and the proportions of urban areas, paddy fields, and dry fields were limited (<10%). These land use characteristics suggest that the two sites were only weakly affected by anthropogenic disturbances. However, the specific reasons for the deviation between observed and predicted ASPT values at these sites remain uncertain.

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

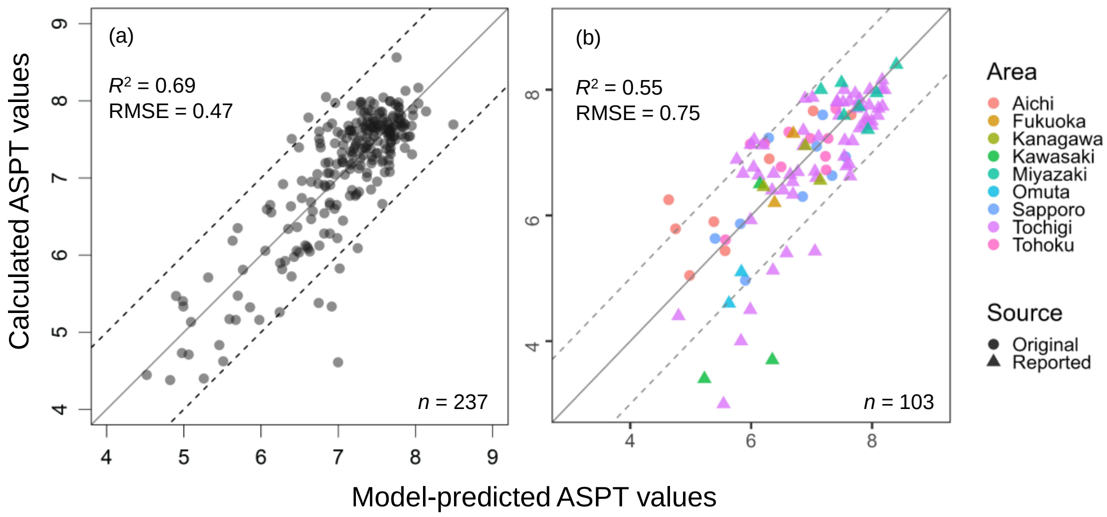

Fig. 3. Relationships between model-predicted average score per taxon (ASPT) and ASPT values calculated from macroinvertebrate survey data during model development (a) and validation (b). Solid lines indicate 1:1 lines, and dashed lines show ±1 of the 1:1 lines. Some ASPT values were calculated from macroinvertebrate surveys conducted as part of the present study (i.e., "original" surveys); these values are indicated by filled circles in panel (b).

R^2, coefficient of determination; RMSE, root mean square error

In the external validation, the best multiple linear regression model predicted ASPT values with an error of ±0.5 for 57% of the data and an error of ±1 for 84% of the data (R^2 = 0.55, RMSE = 0.75; Fig. 3b). These performance evaluation metrics for the validation data were somewhat worse than those obtained with the model development data. Although this is to be expected, this drop-off in performance can be attributed, at least partly, to the presence of several sites where the predicted ASPT value markedly deviated from the ASPT value calculated from macroinvertebrate survey data. For instance, there were six sites where the predicted ASPT value was >1.5 greater than that calculated from the macroinvertebrate data. Based on an examination of aerial photographs, we suspect that suitable habitats for macroinvertebrate surveys (i.e., riffles and runs) might have been absent at these sites, possibly due to factors such as straightened and channelized watercourses. Considering the deviations observed with the model development data as well, it is likely that other unmodeled factors, including the morphological alternation and episodic changes in water quality, contributed to these deviations. Importantly, despite the inconsistency in macroinvertebrate sampling methods (kick and quadrat sampling; see Table S2), no systematic deviations (i.e., under/overestimation) were observed for the validation dataset (Fig. 3b). This is consistent with our initial assumption that differences in macroinvertebrate sampling technique do not materially affect the calculation of ASPT.

ASPT values at water quality monitoring sites across Japan

Based on the predictor values obtained from our published database (Iwasaki et al. 2022), the ASPT values for all 2925 WQM sites (i.e., environmental reference points) were estimated by using the best multiple linear regression model (Fig. 4). These ASPT values indicated that 29% of the WQM sites should be classified as "very good," 50% as "good," 14% as "fairly good," and 8% as "not good." Iwasaki et al. (2022) classified all 2925 WQM sites into four groups based on physicochemical characteristics such as those used in our modeling. As expected, the majority (84%) of the WQM sites categorized as "fairly good" and "not good" were characterized by a high prevalence of paddy and dry fields or urban land uses associated with poor water quality (Iwasaki et

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

al. 2022). Although our categorizations based on ASPT could provide valuable information for screening-level assessments of WQM sites, caution is required when interpreting these results given the relatively large 95% prediction intervals of ASPT values (approximately 2), as well as the reliance on some extrapolation in the estimation process (see Fig. S2 for the distributions of predictor variables in different datasets).

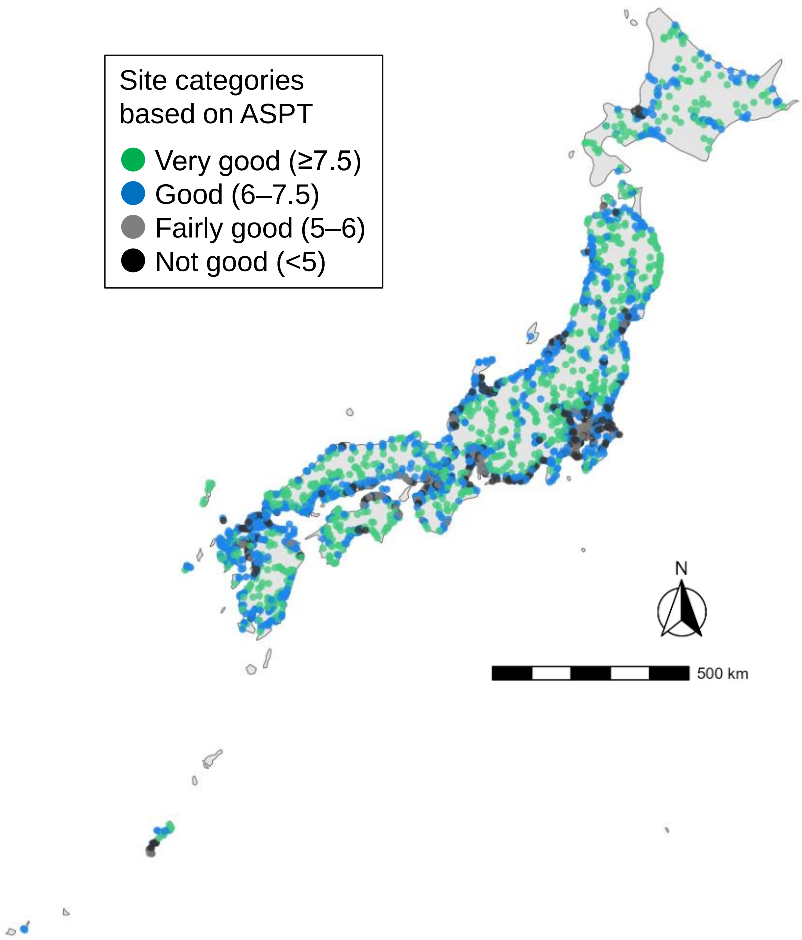

Fig. 4. Categorization of 2925 water quality monitoring sites (environmental reference points) into four river environmental quality categories based on average score per taxon (ASPT) as estimated by using the best multiple linear regression model.

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

Relationships between ASPT and macroinvertebrate metrics

Statistically significant correlations were observed between ASPT and all eight macroinvertebrate metrics examined for taxon richness and abundance, although the correlation coefficients varied between 0.17 and 0.70 (Fig. 5). Richness metrics such as stonefly (Plecoptera) and mayfly (Ephemeroptera) richness had especially high correlation coefficients with ASPT ($r > 0.6$). Compared to mayflies and caddisflies (Trichoptera), stoneflies were rarely found even at the "good" status WQM sites with ASPT values of 6–7 (Fig. 5). Regarding responses of mayflies, the richness and abundance of Baetidae, which is relatively tolerant to water pollution (assigned score = 6; Table S1), had remarkably weak correlations with ASPT ($r = 0.22$ and 0.17, respectively), and stronger correlations were observed between ASPT and the richness abundance of mayflies except Baetidae ($r > 0.7$; Fig. S2). Despite some variations in different richness and abundance metrics, these results indicate strong associations between variations in ASPT and variations in macroinvertebrate richness metrics that are commonly used for biological assessments in rivers (Carlisle &Clements 1999, Namba et al. 2020). Together with the observed considerable correlations between ASPT and other stressor-specific indices designed to detect the impacts of e.g., fine sediment, pesticides, and low flow (Jones et al. 2023), our results support the idea that ASPT should not be used as the sole indicator of water pollution.

Conclusions

In the present study, we developed a multiple linear regression model based on model selection with AICc to estimate ASPT at 2925 WQM sites. The best model included elevation, catchment area, three land use variables (%Paddy, %Urban, and %Urban-

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

3km), BOD, SS, and minimum pH. The model performed well and was able to estimate ASPT values with a reasonable level of accuracy (i.e., an error of ±1 for most sites). Use of the multiple linear regression model to estimate ASPT values for all 2925 WQM sites enabled the categorization of these sites into four groups ("very good," "good," "fairly good," and "not good"), providing the first nationwide categorization of WQM sites in terms of relative river environmental quality. However, these site categorizations should be interpreted with caution because of the uncertainties in the estimation process as well as observed overestimation at certain sites. In addition, there may be some river sites that have naturally poor benthic communities, resulting in lower ASPT values, thereby leading to, for example, a "not good" status. Our approach does not take account such reference/historical conditions, as is done in systems like the River InVertebrate Prediction and Classification System (RIVPACS; Wright (2000) and see Aroviita et al. 2009 for the example application in Finland). To address this issue, it is fundamental to develop a RIVPACS-type predictive model that can predict the presence/absence of macroinvertebrate taxa, preferably using data from least impacted reference sites (see Torii et al. 2023 for a similar modeling attempt in Japan).

The ASPT showed significant correlations with macroinvertebrate metrics frequently used for assessing the biological status of river sites. Therefore, despite the caveats mentioned above, the categorization based on ASPT provides initial but valuable information to capture the biological status of rivers across Japan and can inform effective river management strategies. Particularly in Japan, the compliance of environmental water quality standards in freshwater is assessed at environmental reference points, which correspond to the WQM sites where ASPT values were predicted in the present study. Based on this assessment, the need for countermeasures such as the establishment or refinement of effluent standards is further examined for regulating a given chemical (Naito et al. 2010). In this process, the biological status at the environmental reference points has not been considered even if the objective of environmental water quality standards is the protection of aquatic organisms. Yet, river ecosystems are subject to multiple influences (Birk et al. 2020), and thereby the regulation of individual chemicals may result in limited conservation benefits, particularly in areas where biological communities are already severely impacted (Iwasaki et al. 2018). Information about biological status, such as that can be inferred from ASPT in the present study, should be valuable as foundational knowledge for implementing the effective managements in freshwater ecosystems.

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in Environmental Science and Pollution Research, and is available online at https://doi.org/10.1007/s11356-024-33053-y.

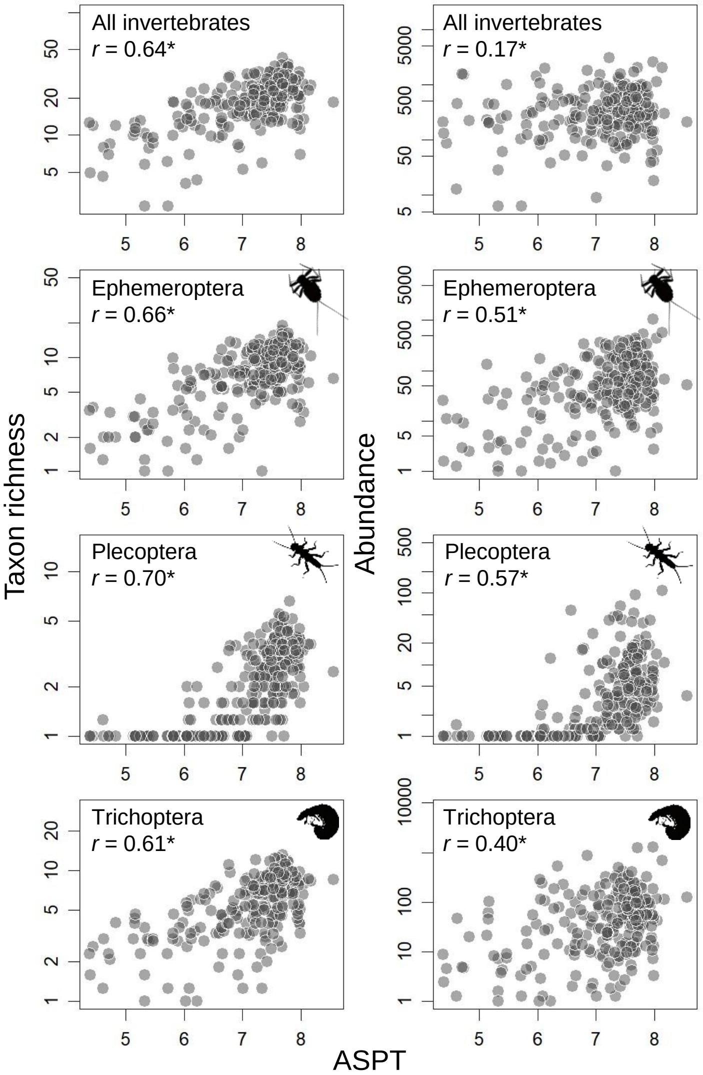

Fig. 5. Relationships between average score per taxon (ASPT) and macroinvertebrate metrics for taxon richness (number of taxa per 625 cm^2) and abundance (number of individuals per 625 cm^2).

Asterisks indicate $p < 0.05$. Macroinvertebrate metrics are averages of three 25 × 25 cm quadrat samples collected per site. Note that for the illustration on a $\log_{10}$-scale, we added 1 to each observed value (i.e., X + 1) to avoid any zero values.

Acknowledgments

We thank Terutaka Mori, Hidetaka Ichiyanagi, Takeshi Mizukami, Noriyoshi Shimura, and Takashi Yamasaki for their help and advice during data collection and handling. During the preparation of this work the authors used ChatGPT to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author contribution

Yuichi Iwasaki: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Tomomi Suemori**: Investigation, Data curation, Writing – review & editing. **Yuta Kobayashi**: Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing.

Funding

This study was supported by JSPS KAKENHI [grant numbers JP18H04141 and JP20K12213]. The funder had no role in the study design, data collection and analysis, interpretation of data, manuscript preparation, or decision to submit.

Data availability

All data and R code used are available from a GitHub repository at https://github.com/yuichiwsk/predict_ASPT_Japan.

Declarations

Ethical approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abell R et al. (2008): Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. Bioscience 58, 403–414. https://doi.org/10.1641/b580507

Armitage PD, Moss D, Wright JF, Furse MT (1983): The performance of a new biological water-quality score system based on macroinvertebrates over a wide-range of unpolluted running-water sites. Water Res. 17, 333–347. https://doi.org/10.1016/0043-1354(83)90188-4

Aroviita J, MykrÄ H, Muotka T, HÄMÄLÄInen H (2009): Influence of geographical extent on typology- and model-based assessments of taxonomic completeness of river macroinvertebrates. Freshw. Biol. 54, 1774–1787. https://doi.org/10.1111/j.1365-2427.2009.02210.x

Aroviita J, Mykrä H, Hämäläinen H (2010): River bioassessment and the preservation of threatened species: Towards acceptable biological quality criteria. Ecol. Indic. 10, 789–795. https://doi.org/10.1016/j.ecolind.2009.12.007

Büttner O, Jawitz JW, Birk S, Borchardt D (2022): Why wastewater treatment fails to protect stream ecosystems in Europe. Water Res. 217, 118382. https://doi.org/10.1016/j.watres.2022.118382

Barbour MT, Gerritsen J, Snyder BD, Stribling JB 1999: Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish (second edition), Office of Water, U.S. Environmental Protection Agency, Washington, DC, USA

Bartoń K (2022): MuMIn: Multi-Model Inference. R package version 1.47.1. https://CRAN.R-project.org/package=MuMIn

Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund WV, Zampoukas N, Hering D (2012): Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic. 18, 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009

Birk S et al. (2020): Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. Nature Ecology & Evolution 4, 1060–1068. https://doi.org/10.1038/s41559-020-1216-4

Burnham KP, Anderson DR (2004): Multimodel inference - understanding AIC and BIC in model selection. Sociol. Methods. Res. 33, 261–304. https://doi.org/10.1177/0049124104268644

Burnham KP, Anderson DR, Huyvaert KP (2011): AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. 65, 23–35. https://doi.org/10.1007/s00265-010-1029-6

Buss DF, Carlisle DM, Chon T-S, Culp J, Harding JS, Keizer-Vlek HE, Robinson WA, Strachan S, Thirion C, Hughes RM (2014): Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environmental Monitoring and Assessment 187, 4132. https://doi.org/10.1007/s10661-014-4132-8

Carlisle DM, Clements WH (1999): Sensitivity and variability of metrics used in biological assessments of running waters. Environ. Toxicol. Chem. 18, 285–291. https://doi.org/10.1002/etc.5620180227

Eriksen TE, Brittain JE, Søli G, Jacobsen D, Goethals P, Friberg N (2021): A global perspective on the application of riverine macroinvertebrates as biological indicators in Africa, South-Central America, Mexico and Southern Asia. Ecol. Indic. 126, 107609. https://doi.org/10.1016/j.ecolind.2021.107609

Feio MJ et al. (2021): The biological sssessment and rehabilitation of the world's rivers: An overview. Water 13, 371. https://doi.org/10.3390/w13030371

Hawkes HA (1998): Origin and development of the biological monitoring working party score system. Water Res. 32, 964–968. https://doi.org/10.1016/S0043-1354(97)00275-3

IPBES (2019): Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany.

Iwasaki Y, Kagaya T, Matsuda H (2018): Comparing macroinvertebrate assemblages at organic-contaminated river sites with different zinc concentrations: Metal-sensitive taxa may already be absent. Environ. Pollut. 241, 272–278. https://doi.org/10.1016/j.envpol.2018.05.041

Iwasaki Y, Kobayashi Y, Suemori T, Takeshita K, Ryo M (2022): Compiling physicochemical characteristics of water quality monitoring sites (environmental reference points) in Japanese rivers and site grouping. J. Japan Soc. Water. Environ. 45, 231–237. https://doi.org/10.2965/jswe.45.231

Johnson AC, Jin X, Nakada N, Sumpter JP (2020): Learning from the past and considering the future of chemicals in the environment. Science 367, 384–387. https://doi.org/10.1126/science.aay6637

Jones JI, Lloyd CEM, Murphy JF, Arnold A, Duerdoth CP, Hawczak A, Pretty JL, Johnes PJ, Freer JE, Stirling MW, Richmond C, Collins AL (2023): What do macroinvertebrate indices measure? Stressor-specific stream macroinvertebrate indices can be confounded by other stressors. Freshw. Biol. 68, 1330-1345. https://doi.org/10.1111/fwb.14106

Larsen S, Vaughan IP, Ormerod SJ (2009): Scale-dependent effects of fine sediments on temperate headwater invertebrates. Freshw. Biol. 54, 203–219. https://doi.org/10.1111/j.1365-2427.2008.02093.x

Lynch AJ et al. (2023): People need freshwater biodiversity. WIREs Water, e1633. https://doi.org/10.1002/wat2.1633

MLIT (2016): Basic Survey Manual for the National Census on the River Environment [River version, Benthic Macroinvertebrate Survey Edition] (in Japanese). River Environment Division, Water Management and Land Conservation Bureau, Ministry of Land, Infrastructure, Transport and Tourism. https://www.nilim.go.jp/lab/fbg/ksnkankyo/ (accessed November 21, 2023)

MoE (2017): Manual of Water Quality Assessment Method by Aquatic Organisms - Japanese Version of Average Scoring System- (in Japanese). Water and Atmospheric Environment Bureau, Ministry of Environment.

https://www.env.go.jp/water/mizukankyo/hyokahomanual.pdf (accessed November 21, 2023)

Naito W, Kamo M, Tsushima K, Iwasaki Y (2010): Exposure and risk assessment of zinc in Japanese surface waters. Sci. Total Environ. 408, 4271–4284. https://doi.org/10.1016/j.scitotenv.2010.06.018

Namba H, Iwasaki Y, Heino J, Matsuda H (2020): What to survey? A systematic review of the choice of biological groups in assessing ecological impacts of metals in running waters. Environ. Toxicol. Chem. 39, 1964–1972. https://doi.org/10.1002/etc.4810

Niemi GJ, McDonald ME (2004): Application of ecological indicators. Annual Review of Ecology Evolution and Systematics 35, 89–111. https://doi.org/10.1146/annurev.ecolsys.35.112202.130132

Nozaki T (2012): Biological assessment based on macroinvertebrate commnities -average score system for Japanese rivers- (in Japanese). J. Japan Soc. Water. Environ. 35, 118–121.

Ormerod SJ, Durance I (2009): Restoration and recovery from acidification in upland Welsh streams over 25 years. Journal of Applied Ecology 46, 164–174. https://doi.org/10.1111/j.1365-2664.2008.01587.x

Persson L, Carney Almroth BM, Collins CD, Cornell S, de Wit CA, Diamond ML, Fantke P, Hassellöv M, MacLeod M, Ryberg MW, Søgaard Jørgensen P, Villarrubia-Gómez P, Wang Z, Hauschild MZ (2022): Outside the safe operating space of the planetary boundary for novel entities. Environ. Sci. Technol. 56, 1510–1521. https://doi.org/10.1021/acs.est.1c04158

R Core Team (2022): R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ (2019): Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev. 94, 849–873. https://doi.org/10.1111/brv.12480

Roy AH, Rosemond AD, Paul MJ, Leigh DS, Wallace JB (2003): Stream macroinvertebrate response to catchment urbanisation (Georgia, U.S.A.). Freshw. Biol. 48, 329–346. https://doi.org/10.1046/j.1365-2427.2003.00979.x

Ryo M, Rillig MC (2017): Statistically reinforced machine learning for nonlinear patterns and variable interactions. Ecosphere 8, e01976. https://doi.org/10.1002/ecs2.1976

Schmidt TS, Van Metre PC, Carlisle DM (2019): Linking the agricultural landscape of the Midwest to stream health with structural equation modeling. Environ. Sci. Technol. 53, 452–462. https://doi.org/10.1021/acs.est.8b04381

Takeshita KM, Hayashi TI, Yokomizo H (2022): What do we want to estimate from observational datasets? Choosing appropriate statistical analysis methods based on the chemical management phase. Integr. Environ. Assess. Manag. 18, 1414–1422. https://doi.org/10.1002/ieam.4564

Torii T, Abe E, Tare H, Tsuzuki T, Myosho T, Kobayashi T (2023): Prediction of average score per taxon in Japan using mega data from the national census on river environments. Limnology. https://doi.org/10.1007/s10201-023-00729-2

Vannote RL, Minshall GW, Cummins KW, Sedell JR (1980): The river continuum concept. Can. J. Fish Aquat. Sci. 37, 130–137. https://doi.org/10.1139/f80-017

Waite IR, Munn MD, Moran PW, Konrad CP, Nowell LH, Meador MR, Van Metre PC, Carlisle DM (2019): Effects of urban multi-stressors on three stream biotic assemblages. Sci. Total Environ. 660, 1472–1485. https://doi.org/10.1016/j.scitotenv.2018.12.240

Waite IR, Van Metre PC, Moran PW, Konrad CP, Nowell LH, Meador MR, Munn MD, Schmidt TS, Gellis AC, Carlisle DM, Bradley PM, Mahler BJ (2021): Multiple in-stream stressors degrade biological assemblages in five U.S. regions. Sci. Total Environ. 800, 149350. https://doi.org/10.1016/j.scitotenv.2021.149350

Wright JF (2000): An introduction to RIVPACS. In: Wright JF, Sutcliffe DW , Furse MT (Editors), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association,, Ableside, UK, pp. 1–24

Yamasaki M, Nozaki T, Fujisawa A, Ogawa T (1996): Researches on the establishment of the standard method to evaluate lotic environments based on the biological condition of macrobenthic invertebrates in Japan -the results of the collaborative studies by the Environmental Biology Group of Environmental Laboratories Association-. Journal of Environmental Laboratories Association 21, 114–145. https://dl.ndl.go.jp/info:ndljp/pid/11641873

Yamazaki D, Togashi S, Takeshima A, Sayama T (2018): High-resolution flow direction map of Japan. J. Jpn. Soc. Civil Eng. Ser. B1 74, I_163–I_168. https://doi.org/10.2208/jscejhe.74.5_I_163

Ye F, Kameyama S (2020): Long-term spatiotemporal changes of 15 water-quality parameters in Japan: An exploratory analysis of countrywide data during 1982–2016. Chemosphere 242, 125245. https://doi.org/10.1016/j.chemosphere.2019.125245