From molecules to monitoring: integrating genetic tools into freshwater quality assessments

Beentjes, K.K.

CHAPTER 2

The influence of macroinvertebrate abundance on the assessment of freshwater quality in the Netherlands

Kevin K. Beentjes[1,2], Arjen G. C. L. Speksnijder[1], Menno Schilthuizen[1,2], Bartholomeus E. M. Schaub[3], Berry B. van der Hoorn[1]

Metabarcoding and Metagenomics 2:e26744 (2018)
https://doi.org/10.3897/mbmg.2.26744

[1] Naturalis Biodiversity Center, Leiden, The Netherlands
[2] Institute of Biology Leiden, Leiden University, Leiden, The Netherlands
[3] Hoogheemraadschap van Rijnland, Leiden, The Netherlands.

ABSTRACT

The use of molecular tools for the detection and identification of invertebrate species enables the development of more easily standardizable inventories of biological elements for water quality assessments, as it circumvents human-based bias and errors in species identifications. Current Ecological Quality Ratio (EQR) assessments methods, however, often rely on abundance data. Translating metabarcoding sequence data into biomass or specimen abundances has proven difficult, as PCR amplification bias due to primer mismatching often provides skewed proportions of read abundances. While some potential solutions have been proposed in previous research, we instead looked at the necessity of abundance data in EQR assessments. In this study, we used historical monitoring data from natural (lakes, rivers and streams) and artificial (ditches and canals) water bodies to assess the impact of species abundances on the EQR scores for macroinvertebrates in the Water Framework Directive (WFD) monitoring program of The Netherlands. By removing all the abundance data from the taxon observations, we simulated presence/absence-based monitoring, for which EQRs were calculated according to traditional methods. Our results showed a strong correlation between abundance-based and presence/absence-based EQRs. EQR scores were generally higher without abundances (75.8% of all samples), which resulted in 9.1% of samples being assigned to a higher quality class. The majority of the samples (89.7%) were assigned to the same quality class in both cases. These results are valuable for the incorporation of presence/absence metabarcoding data into water quality assessment methodology, potentially eliminating the need to translate metabarcoding data into biomass or absolute specimen counts for EQR assessments.

2.1 INTRODUCTION

Quality monitoring of freshwater ecosystems is prescribed under the European Union Water Framework Directive of 2000 (EU WFD; Directive 2000/60/EC) and focuses on monitoring of biological quality elements (BQEs) (European Union 2000). In Europe, benthic invertebrates are one of the most prevalently monitored BQE (Birk et al. 2012). Invertebrate communities are made up of species that represent a broad range of trophic levels, ecological functions and tolerances to stressors (Kenney et al. 2009). Traditional monitoring of freshwater macroinvertebrates, however, is labor-intensive and heavily dependent on expert knowledge of the assessors, making it slow, expensive and prone to human-induced bias and errors at all stages of collecting, sorting and identifying (Clarke & Hering 2006, Haase et al. 2010).

The incorporation of DNA barcodes (Hebert et al. 2003) into the identification process seems to have alleviated some of the human-induced issues. The use of these barcodes for identification of species has become more and more accepted, especially given the decline in traditional taxonomists (Hopkins & Freckleton 2002) and the ability of DNA barcodes to provide identifications of non-adult specimens and distinguish between cryptic clades (Sweeney et al. 2011, Jackson et al. 2014, Macher et al. 2016). Recent developments in DNA metabarcoding show high potential to assess biodiversity across many biomes (Taberlet et al. 2012a, Carew et al. 2013, Leray et al. 2013, Gibson et al. 2014, Stein et al. 2014, Pauls et al. 2014).

Now that the actual identification of species in bulk samples with high throughput sequencing (HTS) has shown its efficacy, the focus seems to shift towards solving some of the other issues concerning these novel strategies, especially the relationship between input biomass or specimen counts and output HTS sequence abundances (Amend et al. 2010, Deagle et al. 2013, Kelly et al. 2014, Elbrecht & Leese 2015, Piñol et al. 2015, Gibson et al. 2015, Hering et al. 2018, Aylagas et al. 2018)

The discussion, regarding the use of HTS read counts as an approximation of biomass or specimen abundances, is important for the biological components of the WFD as well. Abundance of (indicator) species or species groups is used in many European assessment metrics (albeit regularly as abundance classes) and is often part of multi-metric approaches (Birk et al. 2012, Hering et al. 2018, Pawlowski et al. 2018). While information on species abundances and evenness are generally considered important ecosystem properties, the often relatively simple WFD scoring systems may abide with a presence/absence-based methodology. Most traditional morphological monitoring relies on specimen count data, rather than biomass abundances, so even in situations where read abundances can be translated into

relative biomass, comparisons are difficult, considering also that most invertebrate taxa differ in biomass depending on their life stage. If presence/absence data can be as useful for WFD scoring as abundance data, it would allow for easier and faster incorporation of molecular techniques, especially now that efforts have been made to infer biotic indices from DNA data (Aylagas et al. 2014, Elbrecht et al. 2017a, Pawlowski et al. 2018).

In this study, therefore, we assessed the influence of species abundances on the Ecological Quality Ratio (EQR) scores for macroinvertebrates in the WFD monitoring program of The Netherlands. The Dutch system uses abundance data (in the form of abundance classes) for macroinvertebrates, where each species is scored as either a positive indicator, a negative indicator, a characteristic species or none of the aforementioned, depending on the type of water body (Evers et al. 2012, Van der Molen et al. 2016). A simple formula is used to calculate the ratio between normalized values for the indicators and expected reference values for the water type, which is expressed as a value between 0 and 1. Using historical records from traditional monitoring, we evaluated whether abundance data and presence/absence-based data produce comparable EQR scores.

2.2 MATERIALS AND METHODS

EQR scores for macrofauna were calculated on historical monitoring data from four Dutch water authorities (Hoogheemraadschap van Rijnland, Waterschap Aa en Maas, Waterschap Brabantse Delta and Waterschap Rivierenland), using morphological macroinvertebrate records from 2009 to 2017. These records are based on traditional macrofauna monitoring using kick-net sampling and morphological identification. The dataset included 877 monitoring locations spanning 23 different water types according to the Dutch classification system. Most locations were monitored more than once (some even annually), creating a total of 1780 macrofauna samples. An overview of the samples is provided in Table 2.1.

EQR macrofauna scores were calculated for all samples. The scoring system is based on the presence and/or abundance of positive indicator (DP), negative indicator (DN) and characteristic (KM) taxa. Most taxa are identified to species level in the Dutch macrofauna metrics, although for some "harder to identify" groups, species aggregates or higher-level taxonomic assignments are used (Evers et al. 2012, Van der Molen et al. 2016). In the most recent version of the Dutch WFD benchmarks, the absolute abundances of the dominant negative and the characteristics species used in the calculation are transformed into abundance classes (van der Hammen 1992). The

TABLE 2.1. Overview of samples. Distribution of samples used in this study, per water authority (includes survey time span), divided into the three categories defined by the EQR calculation: artificial ditches and canals, natural lentic (lakes) waters and natural lotic (rivers and streams) waters. No monitoring sites are present in rivers and streams for Hoogheemraadschap van Rijnland.

	Natural waters		Artificial waters	Total
	Lakes (type M12-M32)	**Rivers / streams** (type R04-R18)	**Ditches / canals** (type M01-M10)	
Hoogheemraadschap Rijnland (2009–2014)	198	n/a	173	371
Waterschap Aa en Maas (2011–2017)	9	221	150	380
Waterschap Brabantse Delta (2011–2016)	139	230	62	431
Waterschap Rivierenland (2011–2017)	8	56	534	598
Total	354	507	919	1780

EQR scores are calculated according to three different methods, based on the water type. Natural water bodies are divided into lentic and lotic. For lentic water bodies, such as lakes, the EQR is calculated according to the formula:

$$\mathbf{EQR_L} = (200*((KM\%)/KMmax)+(100-DN\%)+KMDP\%)/400$$

Where KM% is the percentage of characteristic taxa, KMmax is a constant value representing the expected fraction of characteristic taxa depending on the specific water type, DN% is the percentage of negative indicator individuals and KMDP% is the percentage of characteristic and positive indicator individuals (Van der Molen et al. 2016). Lotic water bodies, such as streams and rivers, are calculated slightly differently, with more emphasis on the negative indicators:

$$\mathbf{EQR_R} = (200*((KM\%)/KMmax)+(2*(100-DN\%))+KMDP\%)/400$$

For artificial water bodies, such as ditches and canals, the calculation is performed according to the following formula:

$$\mathbf{EQR_A} = ((2*PT/PTmax)+(1-(DN\%/DNmax)))/3$$

Where PT is the absolute number of positive indicator taxa, PTmax is the absolute number expected positive indicator taxa, DN% is the percentage of negative indicator individuals and DNmax is a constant value depending on the specific water type (Evers et al. 2012). The resulting score of all formulae is a value between 0 and 1, which is subdivided into five quality classes: "bad" (EQR <0.2), "poor" (0.2–0.4), "moderate" (0.4–0.6), "good" (0.6–0.8) and "high" (0.8–1.0). These scores also reflect how observed conditions compare to reference status (and thus target status) for the assessed water type, where the highest status shows no difference and the lowest status shows substantial differences (Birk et al. 2012). For artificial water bodies, there are only four quality classes, with "good" representing scores between 0.6 and 1.0, as artificial waters have no natural reference status for comparison.

For each of the 1780 samples, EQR scores were calculated using both original data with abundance classes and a manipulated dataset, converted to a presence/absence monitoring scheme by setting all specimen counts to 1. Any duplicate taxa in a given sample (e.g. where both adult and juvenile specimens were recorded separately) were removed to avoid aggregation into abundance classes other than 1 (abundance class 1 indicates a single specimen was found). QBWat software version 5.33 (Pot 2015) was used to compare and score the original and manipulated monitoring lists with predefined positive and negative indicator species lists, as well as the characteristic taxa list and the EQR based on the relevant formula for each water type was calculated. These predefined species lists (positive indicators, negative indicators and characteristic taxa) have been created specifically for EQR assessments and are based on species characteristics described in literature and expert judgements (Van der Molen et al. 2016). EQR scores with and without abundances were compared to determine the effect on the score, as well as the effect on the classification into the five quality classes. Dunn's test was used to investigate the difference between water types and between quality classes.

2.3 RESULTS

The investigated macrofauna samples had an average of 72.1 ± 0.8 (mean ± SEM) species (minimum 1, maximum 217) recorded, with an average of 1221.5 ± 25.8 specimens (minimum 1, maximum 11767). Mean EQRs, calculated with presence/absence-based data, were highly correlated to original EQRs based on abundance class data, for natural lentic sites (Pearson correlation $r = 0.993$, $p < 0.001$) (Figure 2.1A), natural lotic sites (Pearson correlation $r = 0.982$, $p < 0.001$) (Figure 2.1B) and artificial sites (Pearson correlation $r = 0.983$, $p < 0.001$) (Figure 2.1C). Neither of the EQRs,

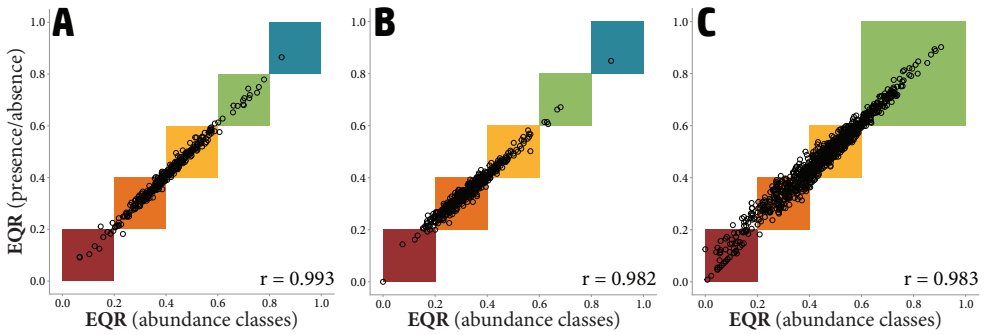

FIGURE 2.1. EQR (presence/absence) versus EQR (abundance classes). Comparison of macroinvertebrate EQR scores in standard assessment using abundance classes and EQR scores in simulated scenarios with presence/absence data for (A) natural lentic waters (lakes, n=354), (B) natural lotic waters (streams and rivers, n=507) and (C) artificial waters (ditches and canals, n=919). Coloured boxes indicate EQR quality classes: "bad" (red), "poor" (orange), "moderate" (yellow), "good" (green) and "high" (blue). For artificial water bodies, there are only four quality classes, with "good" representing scores between 0.6 and 1.0. For all comparisons, the EQR scores of abundance class data and presence/absence data was significantly correlated (Pearson correlation, $p < 0.001$). Pearson correlation values are provided in the panels.

nor the difference between the two scores (ΔEQR) followed a normal distribution. Mean EQR without abundance data was 0.424 ± 0.003, which was significantly higher than the mean EQR calculated with abundance classes (0.404 ± 0.003) (Wilcoxon signed-rank test, $p < 0.001$). The majority of EQRs were higher without abundances (1349 samples, 75.8%), 359 samples scored lower (20.2%) and only 72 out of 1780 samples (4.0%) scored exactly the same (based on scores with three decimal digits). The removal of abundance classes had significantly less impact on the scoring for natural lentic systems (mean ΔEQR 0.006 ± 0.001) than it had on both natural lotic systems (0.021 ± 0.001) and artificial water bodies (0.024 ± 0.001) (Dunn's test, $p < 0.001$). There was no significant difference between the lotic and artificial systems (Figure 2.2A). Removal of abundance data had a stronger effect on samples from the lowest quality class ("bad"), where the mean ΔEQR was significantly higher than all other quality classes (Dunn's test, $p < 0.001$). Mean ΔEQR of the "poor" class, in turn, was significantly higher than those of "moderate" and "good" (Dunn's test, $p < 0.001$), while there was no significant difference in the impact on "moderate" and "good". The "high" class was excluded from this analysis with only two of 1780 samples being assigned to that category (Figure 2.2B).

When assigning quality classes to the EQRs based on presence/absence data, 1596 (89.7%) of all samples were assigned to the same class, 22 (1.2%) were scored lower and 162 (9.1%) were scored higher. The change was most profound in the samples

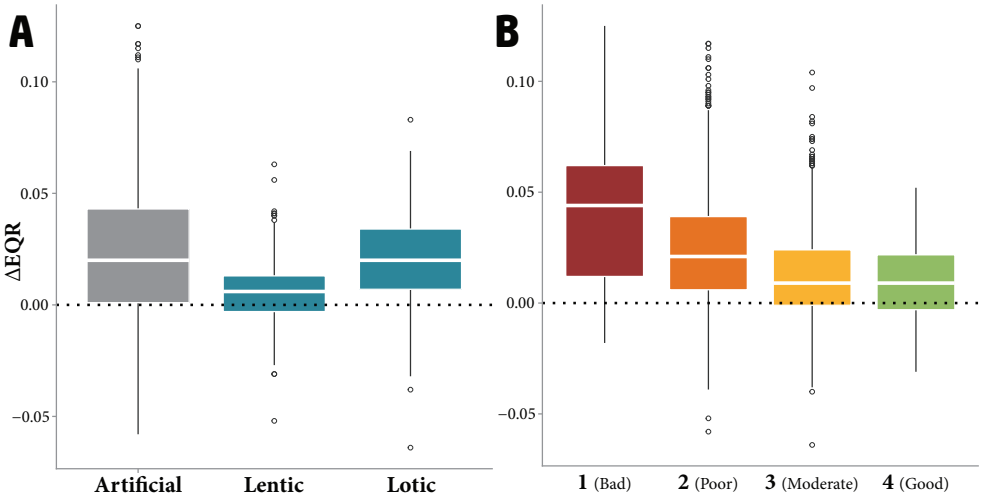

FIGURE 2.2. Factors influencing ΔEQR. Comparison of differences in EQR between assessment using abundance classes and using presence/absence data, (A) split by water type and EQR calculation method and (B) split per original assessment quality class ("high" was omitted, with only two samples in this data set). On average, classifications without abundance are higher than original assessments (ΔEQR positive). Removal of abundance resulted in significantly lower differences in natural lentic waters compared to natural lotic and artificial waters (Dunn's test, $p < 0.001$). There was no significant difference between lotic and artificial. Removal of abundance data has significantly more impact on EQR assessments for samples originally classified as "bad" compared to all other classes (Dunn's test, $p < 0.001$). ΔEQR was also significantly higher in "poor" samples compared to "moderate" and "good".

originally assigned to "bad", where 51 out of 117 (43.6%) were assigned to "poor", the class above. Results were comparable for the different water types: 95.2% of natural lentic samples, 89.9% of natural lotic samples and 87.4% of artificial samples were assigned to the same class. Samples assigned to a different quality class were assigned to a class either directly below or directly above its previous classification.

2.4 DISCUSSION

Our results show there is a strong correlation between traditional EQR based on freshwater macrofauna using abundance data and EQRs calculated without abundance data in the Dutch system. For most samples, scores were comparable between the abundance- and presence/absence-based methods, with the majority (89.7%) being assigned to the same quality class in both cases. The difference seems to be largest in samples at the lower end of the EQR score spectrum, with almost half (43.6%) ending up in a higher quality class ("poor" instead of "bad").

Based on the formulae used for the calculation of the EQRs, it can already be deduced that abundance is not a consideration for all components that determine the final score. For natural lakes, half the score is represented by the fraction of characteristic taxa, which does not take individual specimen counts into account. The fraction of the score defined by factors that use abundance data is slightly lower for natural streams and rivers (two fifths) and for artificial ditches and canals abundances are not used for two thirds of the final score (Van der Molen et al. 2016). Interestingly, in our analysis, the impact of removal of abundances was significantly smaller in lakes than it was in rivers and streams and artificial water bodies (Figure 2.2A). A larger impact on rivers and streams was expected, as 60% of the final EQR is based on the abundance of individuals scoring on each of the three indicator lists (positive, negative and characteristic). However, in the artificial systems, this only amounts to one third of the final score, so one would expect the impact to be smaller, especially considering that the quality classes most impacted by the removal of abundance ("bad" and "poor", Figure 2.2B) only account for 35.8% of the artificial water samples in the data presented in this paper, whereas those classes account for 49.4% and 81.1% of lakes and streams, respectively.

The parts of the EQR score that do rely on abundance data in the Dutch system use abundance classes rather than actual specimen abundances. This may be a major factor in why the removal of abundances has only a limited impact on the EQR scores. Abundance classes were introduced into the Dutch metrics to reduce the effect of extremely high abundances of a single species on the EQR. The abundance class system uses a total of nine classes, where class "1" represents a single specimen and class "9" represents abundances over 1808 specimens. When applied to the monitoring data, this means that an abundance of 20 specimens is translated to class "4", whereas an abundance of 200 specimens is assigned to class "6". Thus, whilst the actual abundance difference might be tenfold, in the calculation it would be only 1.5-fold, already reducing the effect of absolute abundances on the final EQR (van der Hammen 1992, Evers et al. 2012).

These observations are important when considering the incorporation of molecular techniques into WFD quality monitoring methodology. Given that techniques, such as metabarcoding, are proving their efficacy in the process of identification of species in bulk samples, incorporation of such techniques into the actual monitoring is only a matter of time (Zimmermann et al. 2015, Elbrecht et al. 2016, Pawlowski et al. 2018). Efforts have been made in trying to link amplification bias in HTS with amplification success and PCR efficiency of quantitative PCR (qPCR) methods, showing there may be a relationship between low read numbers in HTS and high Cq values in qPCR,

although PCR efficiency itself seemed unrelated (Pawluczyk et al. 2015). Even in case such an approach would yield usable information, it would not only require a priori knowledge of species present within a sample, but also seems cumbersome in complex monitoring samples, such as the ones used for this study (with an average of 72 species).

While our results imply that the technically difficult DNA-based quantifications might be avoided when calculating EQR scores, being able to measure species-abundance relationships from DNA data would nonetheless be desirable, since such relationships play an important role in understanding community composition and dynamics (Hubbell 2001). However, even for the relatively straightforward EQR scoring, the findings in this study cannot be translated into a conclusion that any bias can simply be ignored. These biases are an important consideration when generating taxon lists using HTS on bulk sample metabarcoding. Uneven distributions, paired with preferential amplification of certain taxonomic groups, will result in incomplete recovery of taxa from a sample. It is therefore still important to take the necessary steps to avoid primer bias as much as possible.

One of the main advantages of DNA-based identifications over traditional taxonomy is the ability to reliably identify larval stages and complicated taxonomical groups, for example in cryptic species, showing contrasting reactions to stressors (Macher et al. 2016, Beermann et al. 2018). The use of metabarcoding to replace morphological taxonomic assignment will bring changes to the species lists that can be used for EQR or other quality assessments because there will be more information on those groups that are currently underused due to identification difficulties, as well as potentially higher resolutions of the identifications. Such changes alone may already prove challenging to use in traditional EQR assessments, as these systems have been set up with known limitations in mind. The Dutch macrofauna metric, for instance, makes little to no distinction between genera and species in the family Tubificidae, and many Chironomidae genera have the same scoring for each of its species (Van der Molen et al. 2016). Any such changes alone would warrant a new system, rather than recalibration of the currently used methodologies that are partly built around the limitations of morphological identifications. However, until the lack of knowledge about species-level responses to stressors has been resolved, higher-resolution taxa lists can be merged into less resolved levels to allow for compatibility with current assessment systems.

Taking these considerations into account, together with the fact that expanding the DNA barcode repository for freshwater macroinvertebrates is one of the main focal points of the European DNAqua-Net collaboration (Leese et al. 2016), the

generation of reliable species lists based on molecular data rather than morphological assessments is no longer a vision for the future. EQRs have always been used as a way to quickly assess the ecological status of water bodies. Thorough knowledge of the potential caveats in molecular identification and/or detection techniques will allow for new EQR methods to be developed, methods better suited for use with molecular data. The transition towards DNA-based EQRs certainly has the potential to induce supranational standardization within water quality assessment. Especially with international collaborations such as those in DNAqua-Net, which states that its goal is to "develop a roadmap to include [DNA-based tools] in the standardized ecological assessment of aquatic ecosystems in Europe and beyond" (Leese et al. 2016). Any such molecular-based EQRs might benefit from using more easily generated presence/absence taxon lists instead of an abundance-based analysis, allowing for faster and more and easily standardizable water quality assessments.

2.5 CONCLUSIONS

We demonstrated the viability of adopting presence/absence data instead of specimen abundance data in a WFD water quality assessment program. Given all obstacles hampering the translation of HTS read data into biomass or absolute specimen counts, this paves the way for incorporating metabarcoding workflows into future assessment methodology. While species abundances are still valuable for a thorough ecological understanding of natural systems, the EQRs have been used more as a relatively quick assessment of ecological status of water bodies compared to reference situations. The EQR methodology used in this paper applies to the quality monitoring in The Netherlands and results may vary for other nations, based on the methods of EQR calculation. We urge researchers to look into the actual influence of abundance data on their WFD programs and in studies using metabarcoding data. With molecular techniques, such as metabarcoding of environmental DNA or bulk samples, proving to be successful, it is imperative that developments in routine EQR assessments, be they recalibrations or entirely new systems, strive to be more compatible with the potential lack of abundance data.

2.6 ACKNOWLEDGEMENTS

This study was part of the DNA Waterscan project funded by the Gieskes-Strijbis Fonds. Preliminary simulations of abundance data and EQR-scores were performed by Royal HaskoningDHV. We thank the Hoogheemraadschap van Rijnland and Aquon for providing data and instructions for the QBWat software programme and Waterschap Aa en Maas, Waterschap Brabantse Delta and Waterschap Rivierenland for additional monitoring data.

2.7 SUPPLEMENTARY MATERIALS

SUPPLEMENTARY FILE S2.1. Monitoring event details and EQR scores.
https://doi.org/10.3897/mbmg.2.26744.suppl1