Department of Geography and Environmental Management

Faculty of Environment and Technology

University of the West of England, Bristol

The Anglers' Riverfly Monitoring Initiative: evaluation of data collected through structured monitoring by amateurs and comparative review of the methodology

Daniele Di Fiore

Student Number: 14039321

April 2017

Dissertation submitted in partial fulfilment for the degree of

Master of Science in Environmental Management

University of the West of England

Attestation

I understand the nature of plagiarism, and I am aware of the University's policy on this.

I certify that this dissertation reports original work by me during my University project

………………………………………………………………………………………

Word Count = 11954

(Excluding title page, contents page, list of figures, references, data source list and figure labels)

Acknowledgements

The author would like to thank Dr Mark Everard for his patience and help over these years and during this project. Much gratitude to Ben Fitch and all the people involved in the Angler's Riverfly Monitoring Initiative for the help and for releasing the data they had available. Another big thank you to Dr Chad Staddon, everyone in International Water Security Network and the University colleagues that helped during the MSc. Finally, the author would like to thank his family and his friends that always supported him no matter what.

Abstract

The Angler's Riverfly Monitoring Initiative is a network of organisations and individuals that are interested in the conservation of the riverine ecosystem. The network engages amateurs to collect data about invertebrates in the freshwater systems through structured monitoring as a form of citizen science. A score is produced and compared to a trigger level set by statutory bodies. If the score is equal of lower than the threshold this indicates possible water quality problems.

The study evaluated the scientific validity of the method through the analysis of its theoretical background and an in-depth literature review around biological monitoring. An assessment of the results collected since 2011 by the network in the Severn and Thames River Basin District areas has been undertaken. Particular attention has been dedicated to the analysis of seasonality and variance in the monitoring of taxa identified. Results indicate that volunteers successfully identified variation in the macrobenthic invertebrates, with statistical significant difference in the composition between warm and cold months and high variance of single taxa by month.

Comparison of volunteers results versus results collected by professionals from statutory bodies confirmed that a positive correlation exists between the scores. Nevertheless, the proportion of the variance explained reflects the simplified nature of the ARMI method. This suggest that the employment of the trigger level as baseline is of paramount importance. Finally, the study found that using single taxa scores and the number of taxa identified by volunteers during a sample to predict the scores employed by professionals increases the amount of variance explained.

The results suggest that the ARMI methodology can be employed to assess river quality as it recognises changes in the structure of the invertebrate communities. Further

improvement in the scoring system are possible if more data sampled concurrently by volunteers and professional will be collected.

Table of Contents

List of Figures

List of Tables

1 Introduction

This project analysed data about river invertebrates collected through structured monitoring by interested amateurs as a form of Citizen Science. Results from volunteer monitoring were assessed and then compared with data collected by professionals from statutory agencies.

The importance of citizens' involvement in protecting and managing the environment is increasing as the relationship between communities and their environment has changed significantly due to a range of factors. Some of these issues include urbanisation and consumerist lifestyles, disconnecting people from the ecosystems sustaining their needs (Rosewarne *et al.*, 2013), urbanisation itself is a significant driver of degradation and destruction of the river systems instrumental in their founding (Everard and Moggridge, 2012). Furthermore, increasing population density and complexity of environmental issues have resulted in a situation where local communities, formerly able to 'micromanage' their natural resources for example by limiting overexploitation, are no longer able to take ownership devolving analysis and response to problems to trained specialists (Dunlap & Michelson, 2002).

Extreme weather events due to climate change, overexploitation of resources and poor understanding of long-term impacts over the course of the last few decades have compounded environmental problems, to which regulatory responses are perceived as not having kept pace. Consequently, there is greater public mobilisation to address environmental issues, for example with the emergence in the UK of the Rivers Trust movement (Newson, 2011), engendering a stronger sense of stewardship towards local habitats with local communities demanding a more active role in their protection. This trend is also reflected in government aspirations to devolve responsibility from the state to the public in the UK and elsewhere. Sustainability has become a guiding principle behind this public activism to restore a viable balance between local populations and the environment, strengthening local identities, reconstructing broken co-evolutionary relations and seeking new rules and behaviours that concur to maintain local homeostasis (Dunlap & Michelson, 2002; Hannigan, 2012).

Citizen science, a process of enlisting the public in collecting data across an array of habitats and locations over long spans of time, have proven successful in advancing scientific knowledge particularly about species occurrence and distribution around the world as well as increasing scientific literacy (Bonney *et al.*, 2009). It is therefore a useful tool to help reconnect people and the environment. The term defines scientific research conducted, in whole or in part, by amateur or nonprofessional scientists (Hand, 2010).

The Angler's Riverfly Monitoring Initiative (ARMI) is a national citizen science scheme across the UK under which volunteers collect data in a structured way on river invertebrates (Di Fiore & Fitch, 2016) . The initiative, under the Riverfly Partnership network, enables interested groups and individuals to actively monitor, protect and have a greater sense of ownership of local river monitoring sites more widely and at greater frequency than is possible given the limited monitoring resources of the Environment Agency (EA). This data is used to support River Basin Management Plans developed in response to the EU Water Framework Directive in England and Wales.

Figure 1 The Riverfly Partnership logo

There are several documented occurrences of successful actions, such as prosecution of polluters and identification of incidents, based on data generated under the ARMI programme (BART, 2015; Peacock, 2008; The Riverfly Partnership, 2015).

This research will examine the ARMI method results in order to:

- evaluate the scientific validity of the ARMI method;
- analyse ARMI results in comparison with the monitoring data obtained by statutory bodies;
- identify possible issues of the technique.

1.1 Background and Context Overview

The use of biological indices to assess river quality is widespread in Europe (Abbasi & Abbasi, 2012; Chave, 2001; Ziglio, Siligardi & Flaim, 2006). With the publication of Water Framework Directive (WFD) in 2001, biological monitoring became an important topic in the conservation of waterbodies. The WFD included general guidelines and framework for biological monitoring, which would permit Member States to address the quality of their water bodies. Following the publication of the directive, many methods were researched and implemented, with principles generally harmonising existing schemes across Europe. Remarkably, different EU countries worked out similar methods, as the majority of the indices fall into the category of Biotic Index (Abbasi & Abbasi, 2012; Chave, 2001). In details, the method involves sampling taxa that are known to have a different degree of tolerance/sensibility to pollution and/or water quality parameters (Davis & Simon, 1995). Therefore, the presence and/or relative abundance of certain taxa would indicate distinct water characteristics and/or pollution levels.

In UK, the most used indices are the BMWP (Biological Monitoring Working Group) score and the ASPT (Average Scores per Taxon), used alongside the RIVPACS (River Invertebrate prediction and classification system) computer system. The system enables the comparison of observed versus expected scores based on river characteristics.

Expected scores are calculated based on historical data and assessment of general condition of the waterbodies (Calow & Petts, 1984). Lately, the Whalley, Hawkes, Paisley and Trigg (WHPT) metric has been developed and introduced in response to the requirements of the WFD. The technique involves species level identification and the calculation of an abundance weighted matrix. The WHPT enables the assessment of invertebrates in rivers with relation to general degradation, including organic pollution.

All these indices require trained professionals to sample water bodies for macroinvertebrate taxa, and exploit their presence to assess waterbodies health. Results from these surveys are taken into account in the drafting of WFD River Basin Management Plans.

1.2 Study Rationale

The more recent results for the UK show that the combined effort of statutory bodies and voluntary initiatives was not enough to achieve the "good status" for all waterbodies by the 2015 timeframe (Environment Agency, 2016). The UK may therefore extend the deadline (to a maximum of 2027) or meet less stringent environmental objectives; the EA and the Department for Environment, Food and Rural Affairs (Defra) aim to achieve good status in at least 60% of waters by 2021 and in as many waters as possible by 2027 (Priestley, 2015).

Furthermore, there is a common criticism around citizen science in that the generated data are potentially of lower quality for reasons such as the lack of standards, limited technical capacity, and lower-quality equipment (Cohn, 2008). Some authors (Nerbonne & Nelson, 2008) argue that different groups may have different goals, thus pursuing methods not adequately matched to the purpose of research. Concurrently, quality issues has also been raised about the ability of different indices to detect the correct health status of the environmental medium being assessed (Ruaro & Gubiani, 2013). The main issue of concern has been the creation and definition of reference conditions, which could result in different biotic indices giving dissimilar results and inconsistencies within the same environment.

The ARMI monitoring technique, developed in collaboration with the EA and utilised by the Riverfly Partnership, avoids the technical difficulties of the methods used by experts by building upon and simplifying the BMWP methodology, which has proven scientific validity and extensive use in the UK for the assessment of the ecological status of waterbodies (Di Fiore & Fitch, 2016). Furthermore, the method has been quite successful in many locations, with the Initiative increasingly setting up monitoring schemes(The Riverfly Partnership, 2015).

The numbers of volunteers, sites and monitoring frequency have greatly increased over the years; nevertheless, the majority of data collected by the volunteers is used primarily for the identification of sites with trigger levels below the one set up by the EA. Additional activities can be actioned when volunteers are able to visually identify "out of normal" condition of the water environment even without actually performing the monitoring.

As all the info are recorded into a rich national dataset, there is an excellent amount of data available for further examination and longitudinal analysis on freshwater communities with the possibility of cross validation with other studies conducted by statutory agencies.

The objective of this study is explicated in the following section.

1.3 Study Objectives

The aim of the research is to investigate the application of the ARMI method, evaluate the reference conditions and examine the pertinence of the technique to evaluate the ecological assessment of waterbodies.

Specifically, the research focuses on:

- Evaluating literature regarding other commonly used river quality indices and analyse the foundation of the ARMI Method;
- Analysing and identifying spatial and temporal variability within ARMI results for the Severn and Thames Basins;
- Investigating ARMI results "concurrence" versus related macroinvertebrate data from government agencies;
- Analysing the weaknesses and strength of the ARMI technique in light of the previous results.

2 Literature Review

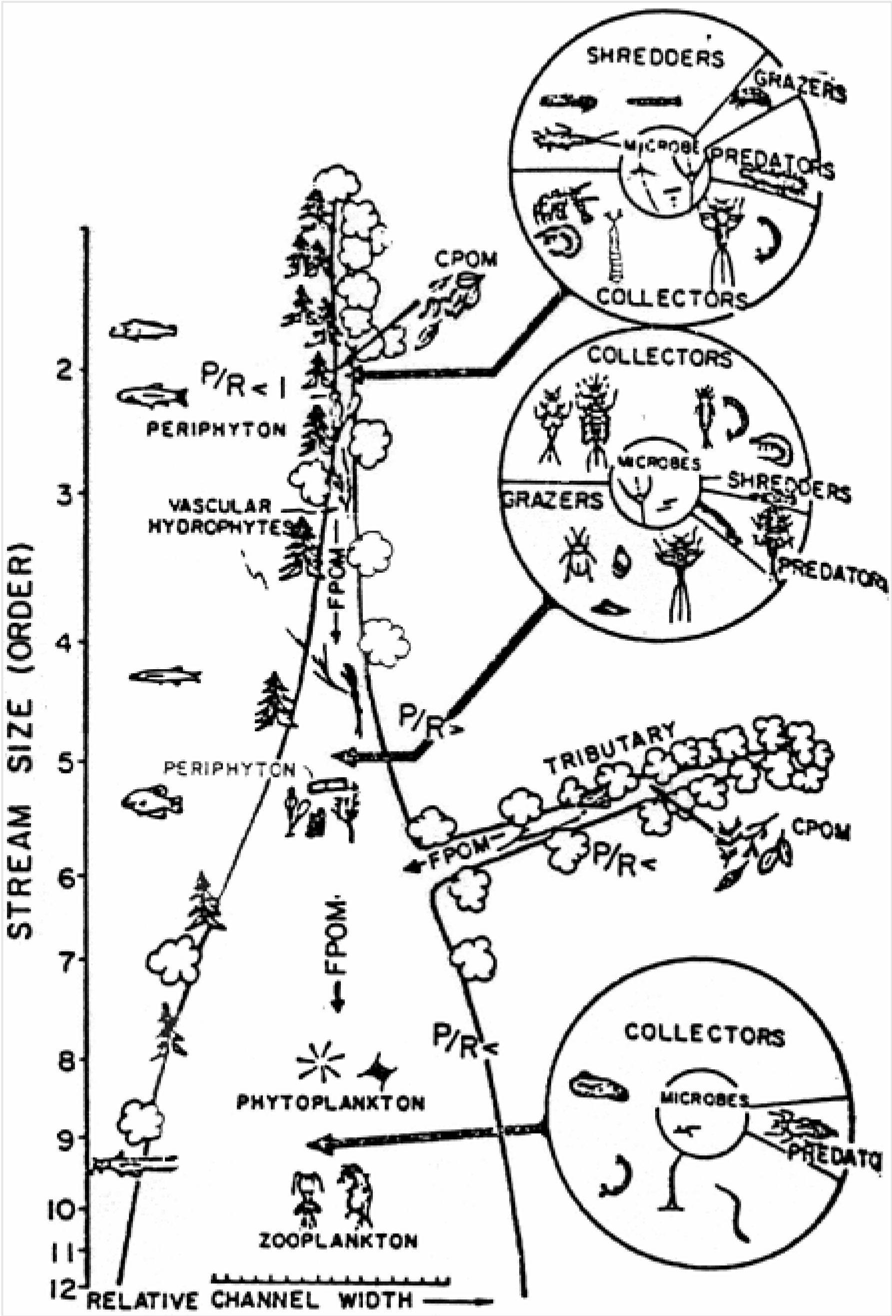
2.1 The composition of freshwater communities

2.1.1 Freshwater Ecosystem Characteristics

The freshwater communities are influenced by aspects such as hydrodynamic conditions, habitat composition, trophic and biotic factors. The combination of these factors creates a multitude of habitats, which are themselves affected by hydrology, morphology, and riparian vegetation. However, the hydrodynamic conditions of waterbodies also vary in function of the flow; which is influenced by climate, erosion, anthropic activity etc. It follows that the freshwater ecosystem is generally not characterised by a great stability (Bailey, Norris & Reynoldson, 2004).

Nevertheless, the freshwater ecosystem is composed by some homogenous area in which conditions are more stable. In landscape ecology, an area with these characteristics is called "patch", which is defined as a relatively homogeneous area that differs from its surroundings. Patches can change and fluctuate over time, a process called "patch dynamics" (Forman, 1995; Wu & Hobbs, 2007). The alternans of stable and heterogenous conditions is quite important in the freshwater ecosystem, where all the elements are subject to physical and chemical disturbances but stable communities can develop.

Figure 2 A depiction of the relationship between stream size and the progressive shift in structural and functional attributes of lotic communities (from Vannote et al., 1980)



2.1.2 Freshwater Macroinvertebrates

In regards to the aims of this research, the community of macroinvertebrates is the most relevant for the purpose.

The classification of macroinvertebrates has not systematic meaning, but only functional and practical (Scaglia, 2009). According to one of the most reliable definitions (Cummins, 1974), the group includes all invertebrates whose last stages of development reach at least 3-5 mm in length. Another classification assigns to the macroinvertebrates organisms that exceed the millimetre in length and are therefore visible to the naked eye (Rosenberg & Resh, 1993). Both classifications define a set of aquatic invertebrates belonging to different taxonomic groups such as insects, crustaceans, molluscs, *Hirudinea, Tricladida, Oligochaeta, Nematomorfa* and many more.

The lifespan of macroinvertebrates can be from a few weeks up to some years. Most macroinvertebrates (crustaceans, molluscs, *Hirudinea, Turbellaria, Oligochaeta, Porifera, Cnidaria* and *Bryozoa*) spends the entire life cycle in the aquatic medium; while others, like almost all the insects, spend in water just a part of theirs (Rosenberg & Resh, 1993). Among all these taxa, there are organisms that have more than one generation year, others that reproduce only once a year, and others with time intervals greater than a single year. Macroinvertebrates can colonise water by passive dispersion through transportation by others organisms such as birds or fish, or by active dispersion, with the ascent of the current or through the flight of adult organisms, such is the case of insects, thus offsetting the phenomenon of drift (Wallace & Webster, 1996).

There are four feeding groups of macroinvertebrates: shredders, filter, collectors, grazers, and predators (Wallace & Webster, 1996). Shredders such as stoneflies (*Plecoptera*) feed on plant material and animal material (generally dead) in minimal part. Collectors, such as caddisflies (*Trichoptera*) and blackflies (*Diptera*), they filter from the water fine organic material. Grazers, such as snails and beetles, feed on algae and other plant material living on rocks and on plant surfaces. Predators such as dragonflies (*Odonata*) feed on other macroinvertebrates. Some species are more generalist than others and can fit into more groups.

Apart from many mechanisms developed by aquatic organism to survive, such as osmoregulation, ability to avoid predation, find food etc., many adaptive strategies of the macroinvertebrates living in aquatic environments are aimed to resisting changes of chemical/physical conditions of the water (Scaglia, 2009). Therefore, they have developed several morphological and functional structures with which they assimilate and/or avoid the absorption of certain organic material (and/or pollutants) present in the stream (Rosenberg & Resh, 1993; Wallace & Webster, 1996; Ziglio, Siligardi & Flaim, 2006).

Furthermore, in order to resist the force of the current, many organisms have acquired the ability to adhere to the substrate. The mechanisms include: presence of structures anchoring the body to the substrate, tapered or flattened body shapes, behaviours such as positive "thigmotaxis" (movement of an organism toward any object that provides a mechanical stimulus) and "rhetoaxis" (orientation of an organism in a stream of liquid, with its long axis parallel with the direction of flow, moving in the opposite direction) (Giller & Malmqvist, 1998; Ziglio, Siligardi & Flaim, 2006; Scaglia, 2009). This is particularly important in the freshwater ecosystem as the zones protected by the current are interconnected and can serve as shelters for many organisms. This spatial organization allows the coexistence of species with very different habitat preferences, life cycles and strategies.

2.2 The Biological Monitoring Method

2.2.1 The rationale for the method

The foundation of the method is that every waterbody collects the contaminants released and/or draining in the hydrological catchment (Laws, 2000). Such pollutants can wind up being so diluted to the point that it is uneconomical and/or technically challenging to detect them through standard chemical/physical analyses.

On the other side, biological monitoring, which employs the analyses of the macroinvertebrate community to reveal habitat modifications, is less challenging and more cost-effective (Karr, 1999).

The analysis of aquatic macroinvertebrates is the most used method for the assessment of the environmental quality (Rosenberg & Resh, 1993; Davis & Simon, 1995; Ziglio, Siligardi & Flaim, 2006; Everard *et al.*, 2011). By assessing species, diversity and functional groups of the benthic macroinvertebrate community, it is possible to determine water quality.

This is feasible as many macroinvertebrate taxa have relatively long life cycles and play different ecological roles. Generally, these organisms are vagile but with limited movements during certain parts of the life cycle. As they tend to remain in their original habitat, these organisms are affected by local changes in water quality. Some are capable of tolerating higher concentration of pollutants and organic material than others. On the other hand, specialist species tend to favour more stable chemical and physical conditions (Rosenberg & Resh, 1993; Wallace & Webster, 1996; Ziglio, Siligardi & Flaim, 2006).

If a pollution event is severe, or is moderate but sustained over time, the whole community structure may simplify in favour of tolerant species. When pollutants are emitted into the system, the abundance of certain species may increase; however, the diversity, and species richness (the number of different species in a given area) decreases (Ziglio, Siligardi & Flaim, 2006).

Biological monitoring is also useful alongside standard chemical and physical analysis, as "*physical and chemical analyses give a measurement which is valid only for the instance in time when the sample was collected, whereas some biological methods reflect the effects of the physical and chemical conditions to which the organisms were exposed over a period of time*" (Chapman & Jackson, 1996).

All the aforementioned characteristics are very valuable to the biological monitoring method. Thus, the use of these taxa as indicators is favoured by a series of characteristics that make macroinvertebrates suitable for the purpose (Scaglia, 2009):

- High sensitivity to pollution and ability to react quickly to its effects;
- Good understanding of the scientific community of morphological and physiological adaptations of many taxa;
- Presence of long life cycles that allows to link their presence to environmental conditions over time;

- Their disappearance is easily attributable to stress conditions;
- Relative low mobility of many taxa, which live on the substrate;
- Aptitude to immediately reflect changes in quality conditions of water and sediments;
- Easiness of collection, and possibility to identify taxa to the naked eye;
- They are preferred food for fish and are a critical component of the food chain of rivers.

2.2.2 The Biological Monitoring Method

Statistical sampling, carried out following precise protocols, can identifies a change in the constitution of these communities; thus, indicating a clear signal of change of environmental characteristics. The activities for the application of biological monitoring can be generally grouped into the following phases:

- **Preliminary investigations**: collection of information material for the proper positioning of sampling stations;
- **Field activities:** identification of ideal sub-locations to perform the sampling, compilation of field data relating to environmental information, use of specific sampling methods to capture the organisms, initial separation and classification of taxa;
- **Laboratory activities:** final classification of the sampled community with the use of optical instruments and tables, assignment of a score to each taxon, and calculation of a biological index value.

2.2.3 Biological Monitoring Limitations

Biological monitoring shows some limitation due to its own nature (Conti, 2008).

In general, the full review of all the species that make up a single community is an extremely challenging task; however, the compilation of checklists is rarely a priority (Bartram & Ballance, 1996; Scaglia, 2009).

The working standard is to perform detailed studies on selected taxa. However, it is clear that the richness in the taxa depends on many variables such as the size of the sample collected. The area from which the sample is collected can also have a profound effect on the number of taxa. Another issue that emerges from the study of the community macroinvertebrates is that relatively few are the common taxa, while most are quite rare. As a consequence, further sampling continues to produce an increase in specie. Furthermore, the precision of the result may be affected by a number of variable, such as the composition and diversity of the substrate, the different attitude of the organisms in the community to be dispersed in the water column and their ability to be anchored to substrate (Conti, 2008).

Other errors that can affect the accuracy of the result may originate by the planning decision, such as the number of samples (transects) carried out on the area under study, the degree of experience of the operators that apply the procedure, and the level of accuracy with which the substrate is explored during the sampling, the diligence and ability of each operator in sorting and identifying the taxa. Additional disadvantages lie within the identification of the taxa to be monitored. The species included in the method might be not sufficiently specific to a particular chemical; might suffer by a combination of other chemicals in the biological medium; might be not sufficient for the assessment of acute (limited in time) pollution phenomena. As per any other monitoring operation that involve the collection of information on a selected sample, the variable that acts mostly on the accuracy of the result is represented by the size of the sample. Thus, a sensible number of samples is necessary to represent correctly the whole community (Conti, 2008; Scaglia, 2009). Finally, competition theory, niche theory and disturbance theory still play large roles in streams, although it is difficult to quantitatively measure their effect on taxa's populations.

2.3 Biological Indices and Legislation

In 2000, the Water Framework Directive (WFD) was published, setting broad frameworks for the use of biological monitoring in assessing freshwater ecosystems ecological status (Di Fiore & Fitch, 2016). Since then, several indices have been developed, employed and subsequently amended, with diverse methods applied in different Member States (Abbasi & Abbasi, 2012). This variance reflected the highly unlikely possibility that a single index could be used in all the different freshwater ecosystems present in Europe, where communities are structured differently and organisms respond in a different way and are subjected to diverse perturbations.

2.4 Most Used Indices in UK

In UK, the two most common indices used were the Biological Monitor Working Party (BMWP) Score and the Average Sore per taxon (ASPT). These indices were included in the river basin plans for the first cycle of the WFD. However, more recently, the Whalley, Hawkes, Paisley & Trigg (WHPT) metrics have replaced these indices as the main metrics for the second cycle.

Biological Monitor Working Party Score

The BWMP was introduced in 1980 to provide an index of river water quality for England and Wales based on aquatic macroinvertebrates. The method is based on the same principles described in the sections above. For the BMWP, the sensitivity is scored versus organic pollutants, with taxa more sensible getting a higher score (Hawkes, 1997; Paisley, Trigg & Walley, 2014).

The method is based on kick sampling, which involves placing a small net downstream (mesh size: 1 mm) from the sampler and agitating with the foot the river bed for at least three minutes so that organisms can be caught by the net. Macroinvertebrates trapped are then stored and preserved with an alcohol solution, and subsequently identified to the family level in order to obtain a final score.

The original process included allocating different scores for eroding and depositing zones. However, the system was simplified with the elimination of habitat scores so that, now, the final score number is just the sum of the tolerance scores of all macroinvertebrate families in the sample, Values greater than 100 are associated with clean streams, while the scores of heavily polluted streams are usually less than 10. Lately, the method has been revised with the reintroduction of the habitat score and new scores for some taxa (Paisley, Trigg & Walley, 2014). A summary of the old and new score system is reported in the appendix.

Average Score Per Taxa (ASPT score)

The effect of sampling effort is usually considered a weakness of biological indices. Usually, it is expected that prolonged sampling period would produce a higher final score than a sample taken in a short time (Hawkes, 1997). The Average Score Per Taxa (ASPT) was introduced to try and overcome this weakness. The score is produced by dividing the BMWP Score by the number of taxa, a high ASPT score usually suggests a clean site containing large numbers of high scoring taxa (Armitage *et al.*, 1983). The value of ASPT for a given site, being an average of such scores, provides the best available estimate of the state of the site with respect to pollutional stress.

Whalley, Hawkes, Paisley & Trigg Metric and RIVPACS model

For WFD cycle 2 the WHPT metrics have replaced the BMWP scores that were used in the first river basin planning cycle.

The WHPT classification comprises two metrics that are assessed separately and then combined in a "worst of" approach to provide the overall invertebrate classification:

- WHPT ASPT (Average Score Per Taxon)
- WHPT NTAXA (Number of taxa contributing to the assessment)

The metrics are then entered into a RIVPACS (River Invertebrate Prediction and Classification System) model, which is a database built using reference sites across the UK.

This is a multivariate approach defined as the use of different multivariate analysis techniques, such as different sorting techniques, usually followed by multiple regression analysis when groups are correlated with environmental variables. The use of this approach has led to the development the RIVPACS system (River InVertebrate Prediction and Classification System), which is the first predictive system concerning the expected macrobenthic fauna in absence of environmental stresses (Wright, 2000). The observed fauna in each site is compared with the expected pattern of wildlife recorded in the RIVPACS database assessing the degree of deviation.

The RIVPACS models use reference-sites metrics and species together with WHPT metrics obtained for the non-reference sites, to compute statistical comparisons and provide a WFD probabilistic classification for the non-reference sites.

The method requires two samples and associated environmental measurements carried out per year for each location. Samples should be collected in the spring (01-March – 31-May) and autumn (01-September – 31 November). Sites may be classified using invertebrate data from one, two or three years (Clarke, R & Davy-Bowker, 2014).

The more recent RIVPACS predictive models have been incorporated into a web-based tool called RICT, which has become the official tool for the WFD macroinvertebrate classification by the UK Agencies.

2.5 The Role of Citizens in Biological Monitoring

In relation to water management, many citizen science programmes have employed biotic indices originally developed by scientists and statutory bodies to investigate water quality. The growing number of initiatives that involve citizen science is resulting in a multitude of monitoring data registered in databases, comprising complementary and non-complementary datasets, which subsequently become available for scientific analysis. (Blossom, 2012; Di Fiore & Fitch, 2016; Roy *et al.*, 2015).

In this context, many biotic indices have been developed *ad hoc* and amendments are regularly investigated to achieve better accuracy (Everard, 2008; Everard *et al.*, 2011). Positive results have been achieved in many research projects where volunteers were involved (Korycińska & Królak, 2006; Moffett & Neale, 2015; Rech *et al.*, 2015). As well as providing valuable data with extended spatial and temporal resolution, citizen science results in other benefits such as improved education about environmental issues and a stronger sense of making a difference.

Where utilised, citizen science has often been exploited as a "first alert", meaning that it has acted as the first alarm of possible pollution events and/or other issues; subsequently, the information has been passed to the appropriate statutory bodies, which pursued more in-depth analysis if necessary.

Finally, citizen science is gaining favourable attention as an approach that "*can inform natural resource management and has some promise for solving the problems faced by adaptive management*" (Aceves-Bueno *et al.*, 2015). Adaptive management is an approach that focuses on identifying critical uncertainties with the aim of reducing risks over time via experiments and system monitoring (Holling, 1978). Buytaert *et al.* (2016) recognise that the involvement of citizens on water resources is increasingly mutating the relation between risks, monitoring and decision making processes. Specifically, the participation of the general public in monitoring initiatives and science-related projects results in the generation of new scientific knowledge. (Buytaert *et al.*, 2014).

Figure 3 Volunteers sampling a site (ARMI)

2.6 The ARMI Method

2.6.1 The Riverfly Partnership Initiative

The history of the Riverfly partnership initiates in 1980, when Dr Cyril Bennett pioneered angler flylife monitoring and entomological courses for anglers (Di Fiore & Fitch, 2016). The initial courses were managed by Steve Brooks and Peter Barnard at the Natural History Museum, in London, however, in the following decade, Riverfly identification courses were also run in Hampshire by Warrant Gilchrist, Dr Bennett and other colleagues at the John Spedan Lewis Trust for the Advancement of the Natural Sciences (JSLTANS). The importance of riverflies became paramount with the publication of the "*Report on the millennium chalk streams fly trend study*" (Frake, Hayes & Region, 2001) which highlighted the decline of flylife across chalk streams in Southern England. Concurrently, the scientific world started to stress the need of volunteer help, particularly for monitoring purposes.

Riverfly identification and monitoring workshops were organised in Hampshire as part of a collaboration between the NHM/EN Partnership, JSLTANS and the Ephemeroptera and Trichoptera Recording Schemes. Riverfly workshops continued around the country in subsequent years.

The Riverfly Interest Group, with key partners including the EA and Salmon and Trout Conservation UK (S&TC – formerly Salmon and Trout Association), was established by "Buglife", the NHM/EN Partnership and others. The first national Riverfly conference entitled "Riverflies: a beacon of environmental quality" was held during November 2004, thereby launching the Riverfly Recording Schemes and establishing the Riverfly Partnership at the same time. With EA collaboration, the Anglers' Monitoring Initiative (AMI) pilot began in 2005. AMI launched nationally in 2007 and has been referred to as the ARMI since 2012 (Di Fiore & Fitch, 2016).

2.6.2 ARMI methodology

In detail, the ARMI method enables trained volunteers to carry out a three-minute kick sample every month, using the same sampling technique and specification equipment used by UK agency ecologists. Presence and abundance of the larval stage of eight invertebrate groups (seven of which are riverflies) is recorded according to an abundancy table and then registered in an online repository.

Abundance	Score	Estimate Numbers
1 to 9	1	Quick Count
10 to 99	2	Nearest 10
99 to 1000	3	Nearest 100
over 1000	4	Nearest 1000

Figure 4 Equipment needed to sample a site

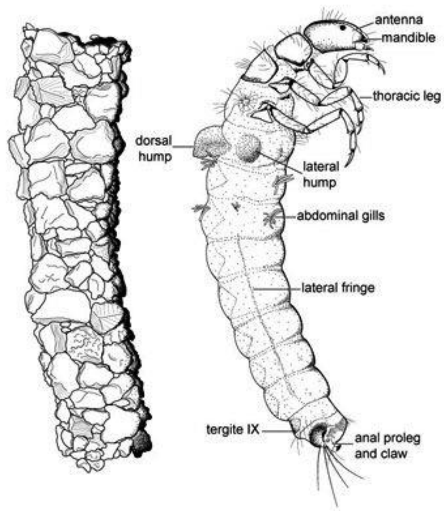
The eight "target groups" of invertebrates used in ARMI are:

- Cased caddis *Trichoptera*;
- Caseless caddis *Trichoptera*;
- Mayflies *Ephemeroptera*;
- Blue-winged olives *Ephemerellidae*;
- Flat-bodied *Heptageniidae*;
- Olives *Baetidae*;
- Stoneflies *Plecoptera*;
- Freshwater shrimps *Gammarus spp*.

The taxa used in the method are detailed below.

Cased and Caseless caddis - *Trichoptera*

Figure 5 Cased Caddis



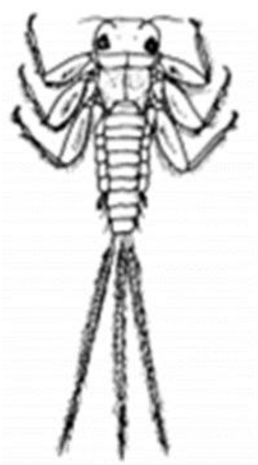
The caddisflies are holometabolous insects that mainly colonise current waters, even if there are families adapted to stagnant waters. Having a broad spectrum of diverse ecological specializations, they are good indicators of water quality.

Eggs are laid in water and generate larvae capable of producing an adhesive silky substance with which they build cases, using material found on the bottom of the riverbed. Caddis larvae are usually divided in the non-taxonomic categories of "cased" caddis, which are more vagile and transport their case when they move, and the so-called "case-less" caddis, which indicates both net-making caddisflies and real caseless species. Case-making caddisflies build cases of silk, that holds together substrate materials such as small fragments of rock, sand, small pieces of twig or aquatic plants. Net-making caddisflies usually live hidden in shelters, with nets that serve both as a means to collect algae, detritus, and animal food and as retreats. Real caseless species are rarer.

Most Trichoptera have an annual cycle, with some species being polyvoltine. They populate different freshwater environments: some species live in environments wet only by a film of water, some occupy running water, distributing along the various zones of the river, other species populate lakes. As for the habitat, the food regime of caddisflies is most varied, with herbivore, scavenging and carnivore species. Some species scrape and graze in the periphyton, other shred the debris, suck the sap of the algae or capture other small invertebrates.

Mayflies- *Ephemeroptera*

Figure 6 Mayfly larvae



http://www.state.nj.us/dep/wms/bfbm/mayfly.gif

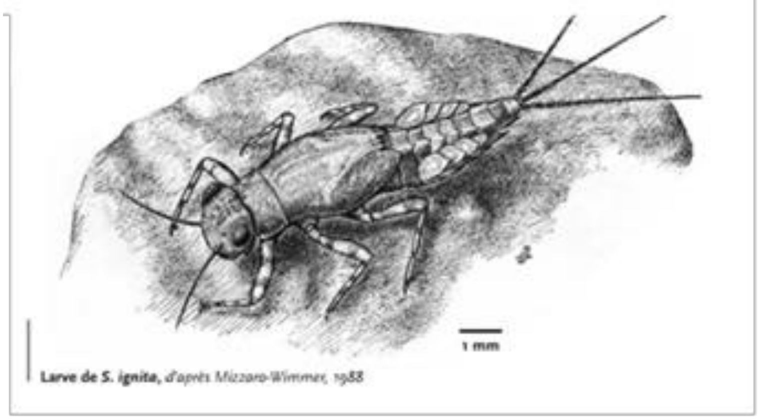
The Ephemeroptera are an order of insects with incomplete metamorphosis that spend most of their lives as larvae, while the adult stage is very short, just long enough to take the life cycle reproducing and spawning (hence the name of the order, from the Greek *ephemeros* - live one day). The adults do not feed, having a non-developed mouthpart. The development cycle is a kind of hemimetabolous-paurometabolous, unique in the world of insects. The life cycle can be univoltine, polyvoltine, or semivoltine (species whose larvae develop in a period of two years). The long stay in water gives the Mayflies an important role as bio-indicators. These insects are mainly herbivorous and detritivore, while they are only occasionally predators. The Ephemeroptera are spread mainly in flowing waters, but a few genera are also adapted to lentic environments. The nymphs have sub cylindrical and tapered body, more or less compressed dorsal-ventral. They possess 5-7 pairs of abdominal tracheal gills, three cerci (two cerci and a cerci-like appendices). Most Ephemeroptera feeds grazing the surface exposed to the stream, and, despite the anchoring mechanism to the bottom, many of them are likely to utilise passive transportation (drift transportation).

Blue-winged olives - *Ephemerellidae*

Ephemerellidae are a family of the order Ephemeroptera, known as Blue-winged olive and Spiny Crawler Mayflies. Larvae and adults can be found in a variety of habitats, especially in flowing waters of streams and rivers, where they feed as collector and gatherers. They possess four pairs of plate-like gills which are held over back.

A well-known characteristic of the Ephemerellidae is that when threatened, they raise their three tails up to frighten the possible predator. If the behaviour is not successful, they curl their abdomen over the body so that the tails project in front of the head and are used to attack the intruder.

Figure 1 Blue-winged olive larvae



Flat-bodied – *Heptageniidae*

The Heptageniidae are a family of Ephemeroptera that are generally rather small with three long tails. The body is usually dark brown and noticeably flattened with broad head, thorax, and femora. Nymphs cling to surface of stones in shallow, rapid streams or in lakes.

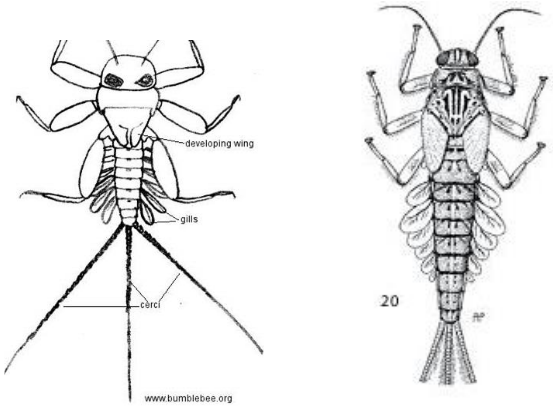


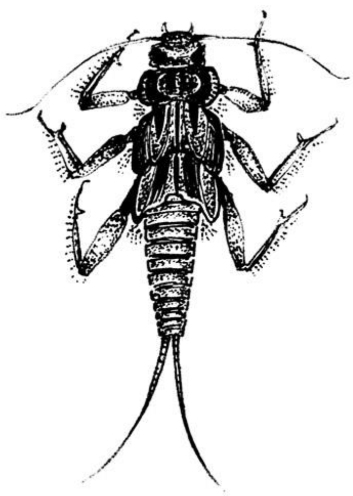
Figure 7 Flat-bodied (left) and Olives (right) nymphs

Olives - ***Baetidae***

The Baetidae are again another family of Ephemeroptera, which are known as Small Minnow Mayflies or Olives. They are usually small, streamlined larvae with long antennae, which are usually two or three times longer than the head's width. Oval or heart-shaped gills are present on abdominal segments, with posterior abdominal segments that usually lack spines pointing backwards.

Stoneflies - ***Plecoptera***

Figure 8 Stonefly larvae

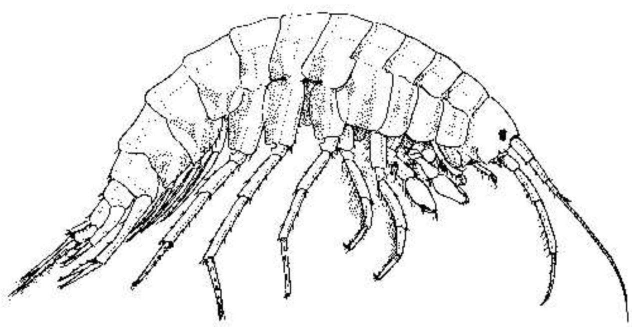


The Plecoptera are an order of heterometabolous insects, of medium or large size, with aquatic larvae and adults living out of the water. They are sensitive to organic pollution and lowering levels of oxygen due to decomposition processes. Plecoptera live in cold, clear and turbulent waters, typical of courses of the lower order. They prefer environments with substrates of boulders and pebbles, where there is high retention of organic matter coarse, which settle in the gaps of stones and trapped leaves. The Plecoptera are mostly detritivore with a life cycle synchronized with the fall of the leaves; other groups are carnivorous and feed on other invertebrates. The nymphs are characterized by two long cerci and the absence of extensive tracheobronchial apparatus, which reflect their relative intolerance to deficiency of dissolved oxygen.

Freshwater shrimps - *Gammarus spp*.

Gammarus spp. are the only non-insect taxon used in the ARMI method. The freshwater shrimps belong to the crustacean order Amphipoda. Their body is flattened from side to side, with seven pairs of thoracic walking legs and six pairs of abdominal limbs, which are used for swimming. They are good swimmers with ability and tendency to drift, which allows them to easily invade and colonize ecosystems, hence they tend to occupy all habitats available in rivers from source to mouth where they feed on fragmented organic matter. The genus has a high reproductive capacity with several broods per female per year, a high number of offspring, and relative longevity.

Figure 9 Freshwater shrimp



Each target group, included in the ARMI methodology, was selected based upon sensitivity to (largely organic) pollution, distribution and status in rivers across the country, and presence throughout the year. Key identification and morphological characteristics were also taken in consideration in the selection.

Workshop are run throughout the year to ensure that volunteers can be trained to identify, sort and record invertebrates according to each target group. The Riverfly Partnership hands out qualifications only after the volunteer has successfully attended a workshop, so that only trained citizen can contribute to the monitoring.

The sampling is ideally carried out each month in the same locations. Volunteers are provided with a key Guide to freshwater macroinvertebrates identifications, a high magnification jeweller's loupe (10x + 20x) and at tray to identify the taxa. After having recognised the macroinvertebrates present in the sample, the volunteer refers to a specific scoring table based on the estimated number of individual taxa, thus producing an ARMI score. The scoring obeys the following table where each taxon is given an abundance score and total score is obtained summing up all single values:

Figure 10 ARMI abundance scores

Recording data

13. Record the category and *estimate* the numbers of each invertebrate group as noted on the recording sheet.

Abundance	Score	Estimated number
1-9	1	Quick count
10-99	2	Nearest 10
100-999	3	Nearest 100
over 1000	4	Nearest 1000

Figure 11 ARMI recording sheet

Riverfly monitoring for anglers and conservation volunteers - Recording sheet

ARMI group	
Site name	
River	
Grid reference	
ARMI group coordinator	

The Riverfly Partnership

		Example month		Month 1		Month 2		Month 3	
	Date	27/06/2015							
	Recorded by	B Fitch & A Menzies							
		Score	Est. number*	Score	Est. number*	Score	Est. number*	Score	Est. number*
Caddisflies	Cased caddisfly	2	20						
	Caseless caddisfly	1	2						
Up-wing flies	Mayfly (Ephemeridae)	2	10						
	Blue-winged olive (Ephemerellidae)	2	20						
	Flat-bodied stone clinger (Heptageniidae)	3	100						
	Olives (Baetidae)	1	4						
Stoneflies	Stoneflies (Plecoptera)	1	3						
Freshwater shrimp	Freshwater shrimp (Gammaridae)	1	8						
	Additional observations/notes	Fly hatches observed. River level: LHB 200mm, Middle 350mm, RHB 150mm.							

This score is recorded and then compared to the site-specific "trigger level" (expected population abundances), which has been established beforehand by the UK agencies as a result of ecological expected parameters. This value is set on a local scale and it's subject to professionals' judgment.

If invertebrate numbers drop below the trigger level, the agency is notified immediately, so that more detailed investigations and appropriate response action can take place. The agency then provides the relevant ARMI monitor with feedback concerning any actions taken, thus validating the volunteer's efforts and maintaining ongoing motivation.

An online data repository enables registered users to track survey results over time, from every registered UK ARMI site.

Figure 12 Map of ARMI sites

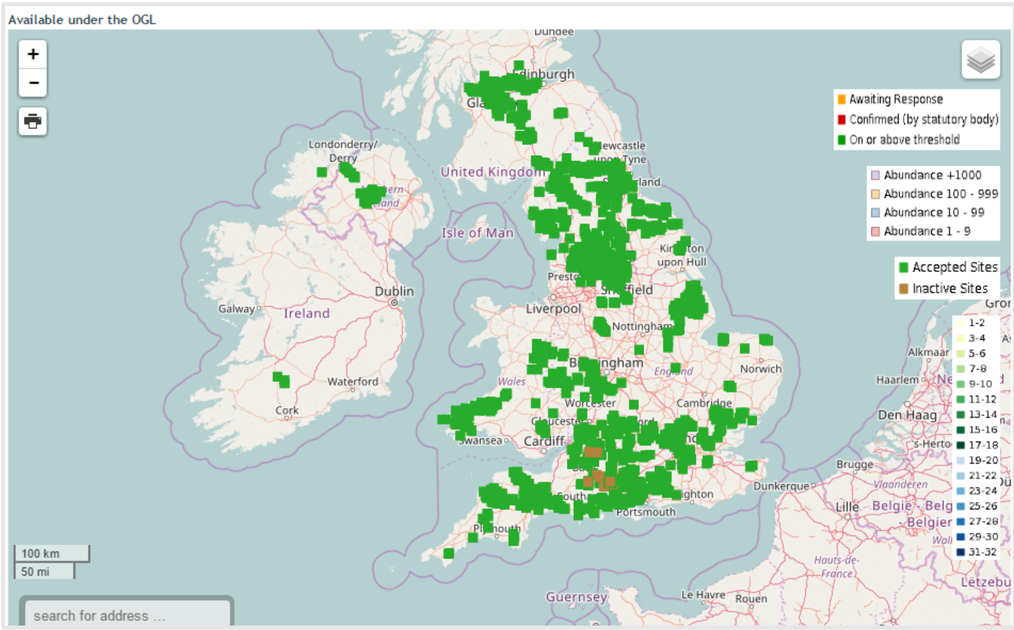
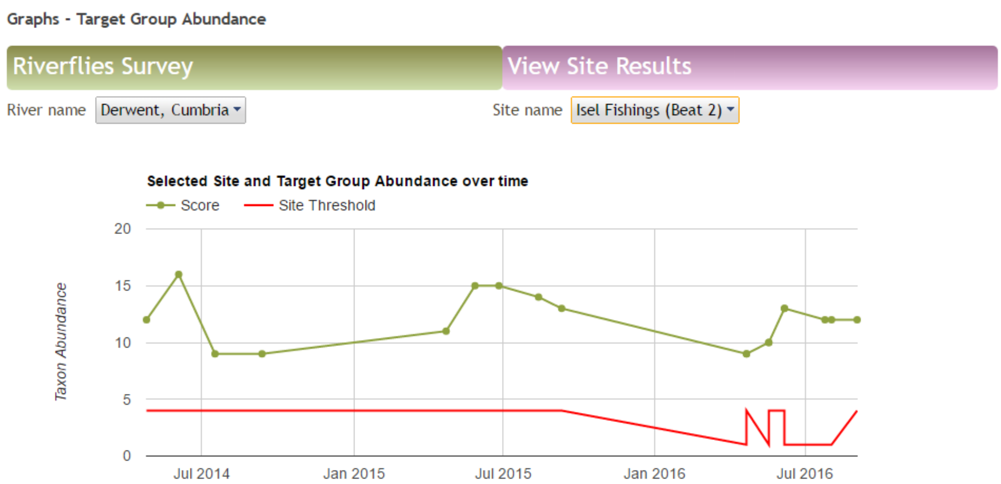


Figure 13 ARMI online repository - Example Chart



3 Methodology

The research is designed as an evaluation of ARMI monitoring technique through literature review and exploratory investigation into the data collected by the Initiative. Further comparative analysis of data collected by the ARMI volunteers and monitoring results obtained by professional from statutory bodies was undertaken.

The aim of this study was:

- to determine whether data collected by ARMI volunteers can be used to assess changes associated with river ecological status
- to compare the precision of assessments made by volunteers with those made by professionals, such as scientist from statutory agencies.

The number of data available (i.e. total number of monitoring stations and complementary data from statutory bodies) has been collated, manipulated, processed and analysed through the use of the following commercial and freeware software:

- ArcGIS suite;
- QGIS suite;
- Microsoft Excel and other software of the Microsoft Office suite;
- R Software and RStudio.

3.1 Data Collection

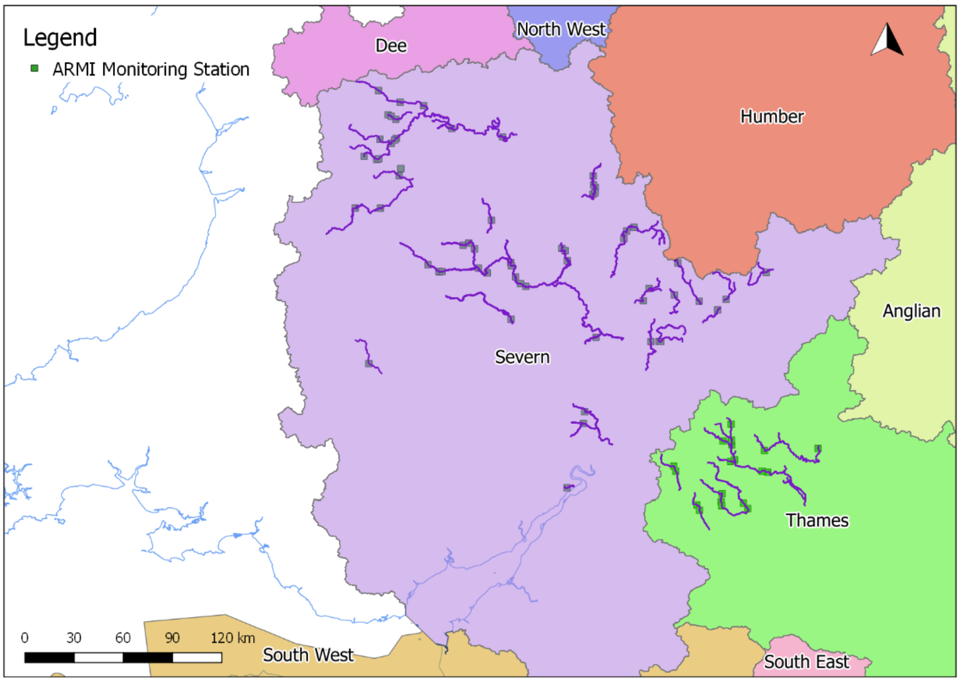
The investigation used secondary data, specifically records collected by the ARMI initiative, in addition of data freely available through statutory bodies' repositories and publications. Secondary data identify "*any data that are examined to answer a research question* *other than the question(s) for which the data were initially collected*" (Vartanian, 2010). In a broader sense, this include any "*analysis of data collected by someone else*" (Boslaugh, 2007). In contrast, primary data are those where the same individual/team of researchers designs, collects, and analyses the data.

The advantage of using secondary data comprehends a series of factor, such as access to huge amount of data that would otherwise take money and time to collect, and use of information that can be of higher quality and/or involve larger samples that are more representative of the target population(s), thus increasing the validity of the analysis. On the other side, some of the disadvantages lie in the fact that the information might need transformation to a great degree, thus requiring long time and/or high computational power to convert the data in a format/layout pertinent to the scope of the research.

3.2 Case Study

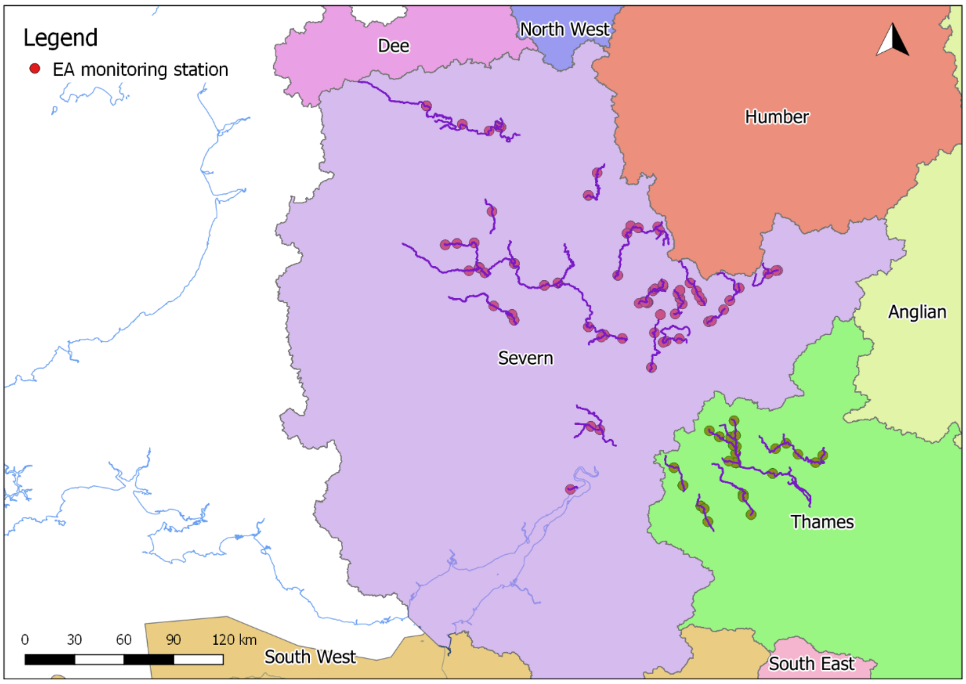
The Riverfly Partnership and the Severn Rivers Trust, which are parts of the ARMI Initiative, have agreed to give access to the database and repository they have instituted for monitoring reasons. For the purpose of the research, it has been established to use the data collected on over twenty waterbodies within the Severn and Thames Basin. This dataset is one of the most data-richest nationally, as each waterbody has one or more monitoring stations with data spanning each month form many waterbodies and available since 2011 for some locations. In total 86 stations for 49 unique stretches (as identified by the EA) were sent.

Figure 14 Area of study – ARMI monitoring station



As for the data collected by statutory bodies, the EA and Defra host rich archives on their websites, which provide large and high-quality databases on water quality data. For the purpose of this research, data collected on ASPT and BMWP scores were utilised.

Figure 15 EA monitoring stations in the area of study



3.3 Data Manipulation

In general, it is frequent that variables are not in the format more useful to the analysis; therefore, it is necessary to operate some manipulation, adjustment or recoding.

The data provided by the Riverfly Partnership and the Severn Trent Trust have been recorded on one single excel-formatted file per location, within a not-standardised number of months; meaning that single location's results are registered on different files, with each file containing information collected over a disparate number of months. To collate all the data together, the method used in investigation has involved the utilisation of a script that builds a single worksheet in Excel with all the information contained in a range of files. Then, all the records have been transformed into a single format, the fields have been cleaned of NULL value and only the pertinent information have been kept.

3.4 GIS development

For the management and viewing of data, a GIS (Geographic Information System) was created, using ArcGIS and QGIS software. The key element of GIS is the database, a collection of associated maps and information in digital form. This comprises of two elements: a spatial database describing the geography (shape and position) and an attribute database that describes the characteristics or qualities of these places.

The use of GIS was needed because with the GIS, the traditional database search capacities is expanded to include the ability to analyse data depending on their location. This was necessary in order to relate data that have different geometries, such as rivers and monitoring stations and obtain cross-references between ARMI and statutory agencies' results.

In details, GIS data employed in the study comprised of:

- base map data that have been obtained from the EA repositories,

- georeferenced geometric entities (managed in a geodatabase) that have been developed from points

The georeferenced entities for monitoring stations were created starting from National Grid Reference information recorded both for ARMI and EA monitoring stations. The NGR is a system of geographic grid references highly used in the UK, which is based on the Ordnance Survey Grid. The NGR system identifies location with alphanumeric number. In details, Great Britain is covered by 100-kilometre grid squares, with each grid square identified by two letters. These squares are further divided into smaller squares by grid lines representing 10-kilometre spacing, each numbered from 0 to 9 from the south-west corner, in an easterly (left to right) and northerly (upwards) direction. From these reference system, a location can be obtained using to numbers: an easting (along the horizontal axis) and a northing (along the vertical axis).

Conversion of NGR location to northing and easting has been conducted using a tool developed by Edina. Subsequently, ARMI stations have been located on a map and spatial joins have been run in order to link the data to the river stretches and obtain unique identification keys for rivers (EA Waterbody ID). The same methodology has been followed for EA monitoring stations. Once both datasets were georeferenced, only waterbodies were both ARMI and EA monitoring stations were present have been selected. This has been done by running queries in SQL (Structured Query Language) on the Waterbody ID attribute and retaining only the pertinent results. All shapefiles created have been also saved in the .csv format, which is the base format used by statistical package used in this project.

3.5 Data Analysis

The data obtained from the GIS built have been then used into R statistical software.

R is a free open-source software tool for statistical analyses and graphics, which is widely used in statistics. The base software offers many tools for the statistical analysis of datasets and high quality graphics, but it can also be expanded with additional packages created by a rich community of developers. The fundamental aspect of R is that is quite flexible and permits to quickly run statistical analysis including linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering and others.

The analysis on the dataset available have been conducted on ARMI results from 2011, this has been done considering the amount of comparable data available from both ARMI and EA datasets.

Statistical analysis has been conducted on different levels, examining several aspects.

Firstly, the study conducted an exploratory analysis on the distribution of single taxa included in the ARMI methodology. Shapiro-Wilks analysis were run for normality. The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05. Passing the normality test allows to state that no significant departure from normality was found. Deal with dataset with normal distribution is useful because statistical test for normal-distributed data are widely used and have high statistical power. Because non-normality was detected, non-parametric analysis has been conducted on most of the datasets.

The second level of analysis has involved examining the relative occurrence of single taxa. The Kruskal-Wallis test was used as an alternative to the ANOVA with no assumption of normality in the taxa distribution to test for independence of sampling and seasonality of taxa distribution. The test is used to test if k samples ($k>2$) come from the same population or populations with identical properties as regards a position parameter (the position parameter is conceptually close to the median, but the Kruskal-Wallis test takes into account more information than just the position given by the median). This test allowed to determine to what extent the results of the ARMI Score corresponded to changes in faunal composition within different sites, months and season, and thus how sensitive the index is in reflecting community structure (Ghani *et al.*, 2016).

Then, the study verified whether correlation within ARMI taxa was present. Correlation analysis has been computed running pairwise Paerson correlation on all the ARMI taxa results. Paerson correlation is a measure of the linear dependence between two variables; for a sample, it is described by the formula:

$$r = \frac{1}{n-1} \sum \frac{(x_i - \overline{X})(y_i - \overline{Y})}{s_x s_y}$$

Where n is the number of observations in the sample, Σ is the summation symbol, X_i is the x value for observation i, $\bar{x}$ is the sample mean of x, Y_i is the y value for observation i, $\bar{y}$ is the sample mean of y, S_x is the sample standard deviation of x, and S_y is the sample standard deviation of y. The value of *r* ranges between +1 and -1:

- r > 0 indicates a positive relationship of X and Y: as one gets larger, the other gets larger.
- r ≤ 0 indicates a negative relationship: as one gets larger, the other gets smaller.
- r = 0 indicates no relationship

Usually, for r > 7 positive correlation is considered high, r > 4 is moderate, anything less low. For negative correlation, it is the other way around, r ≤ -7 negative correlation is considered high, r ≤ -4 is moderate, anything higher low.

There is ambiguity among literature as to whether Paerson correlation test requires normality of the variables. However, most agree that with big enough samples the normality assumption can be broken thanks to the central limit theorem .

Nevertheless, assuming the non-normality of the taxa distribution, the correlation was also investigated with Spearman rank correlation. This is a non-parametric test that does not require any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.

The following formula is used to calculate the Spearman rank correlation:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

rho= Spearman rank correlation

d_i= the difference between the ranks of corresponding values Xi and Yi

n= number of value in each data set

Rho values fall within the same limits than the Paerson coefficient, with similar consideration on low, moderate and high correlation.

The study then proceeded with the examination of correlation amongst the ARMI score and invertebrate index score. Firstly, it was addressed the relative monitoring frequency of ARMI and statutory agencies' monitoring activities. This has been done considering the number of recorded monitoring events per single year and comparing the results by site. Relative monitoring frequency has been use to assess the difference in monitoring events and corroborate the consequent statistical analyses about ARMI and agencies' scores.

Correlation between scores has been done using Spearman correlation running the test on monitoring results where samples were collected within 14-days' timeframe difference (maximum distance of 7 days before or after the ARMI monitoring activity)

Differences among the invertebrates scores were assessed using Wilcoxon, and Kolmogorov-Smirnov tests, which utilise the median and the mean to check whether samples come from the same population, testing the null hypothesis that samples have same distribution. Analysis of variance was tested with Levene's and Bartlett's tests.

Building upon correlation values, to check whether the ARMI method could explain the status of invertebrate communities as described by the BMWP and ASPT scores, multiple linear regression models have been built. Linear regression, or Multiple Linear regression (MLR) when more than one predictor is used, determines the linear relationship between a response (Y/dependent) variable and one or more predictor (X/independent) variables. The least-squares method is used to minimize the vertical distance between the response and the fitted linear line.

The requirements of the test are:

- A dependent response and at least one independent predictor variable, measured on a continuous scale.
- Measurement error in the response variable must be normally distributed and have constant variance, with predictors free of measurement error.

The two assumption were considered satisfied due to the nature of the monitoring and the fact that the sample was big enough to assume normality in the measurement error.

In general, an MLR model that describes a dependent variable y by independent variables $x_1, x_2, \ldots, x_p$ ($p > 1$) is expressed by the following equation, where y is the dependent variable, the numbers a and β_k ($k = 1, 2, \ldots, p$) are the parameters, and ϵ is the error term.

$$y = \alpha + \sum_k \beta_k x_k + \epsilon$$

The p-value for each parameter tests the null hypothesis that the coefficient is equal to zero, meaning that the parameters has no effect on the response variable. A low p-value (≤ 0.05) indicates that the predictor is likely to be significant, meaning that changes in the predictor's value are related to changes in the response variable. A p-value higher than 0.05 means that it is possible with a 95% confidence to state that changes in the predictor are not related to the response.

All the models were built on the dataset were ARMI and agencies' monitoring occurred within a 14-days' timeframe (maximum distance between sampling of 14days). This was done in order to reduce assumptions about the macrobenthic communities when ASPT and BMWP scores were obtained and, consequently, decrease the models' errors. Adjusted R-squared and p-values were examined to check how much variance of the invertebrate scores produced by EA experts could be explained by the taxa included in the ARMI method. Predictors were removed if their p-value was not significant.

In order to improve the variance explained by the model, different parameters and interactions have been investigated. The number of taxa identified within a single monitoring events have been introduced as predictor because it was found statistically significant for all the models. This value consists of the total of the number of different taxa recognised in the monitoring events so that if, for instance, only mayflies and shrimps were sample, the taxa count would be 2. If mayflies, shrimps, caseless caddisflies and olives were sampled in the monitoring activity, the count taxa value would consist of 4. Anova (analysis of variance) tests were run to test if the introduction of new predictors was statistically significant.

Coefficients for all parameters identified in the models were used to obtain the regression line formula. Subsequently, scores were predicted using the linear equation for each model. The values obtained were then transformed into quality categories following the table below:

Table 1 Quality categories for Invertebrates scores

BMWP score	ASPT score	Quality or diversity interpretation
>150	>6	Very good
101-150	>5	good
51-100	>4	moderate
16-50	<4	poor
0-15		Very poor

Prediction power of each model was tested assessing the number of times the forecasted category matched the ones reported by the EA.

The models with best predictive power were than simplified dropping single predictors with no statistical significance (p-value ≤ 0.5). Predictors with highest p-value in each subsequent simplified model were dropped in single steps. Final model was then selected using all remaining variables significant at .95 level.

One final MLR models was built, including selected ARMI taxa and taxa count as independent variables, with the ASPT score as response. This was done because the predictive power of models was consistently better for the ASPT score.

After the identification of the model, because a linear regression model is not always appropriate for the data, residuals were examined. A residual is the difference between the observed value of the dependent variable (y) and the predicted value ($\hat{y}$), so that each data point has one residual. Residuals ar often examined with a residual plot, which a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. If the points in a residual plot are randomly dispersed than a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate. Collinearity was investigated with the VIF package in R.

Finally, in order to validate the model obtained, the dataset was split into a training and test set and cross-validation was run, using the DDAG package from R. Cross-validation allows to resample and check for stability of the model(s). The MLR model was rerun for the training set and new coefficients were obtained. These were then used to predict agencies' scores and river quality categories in the test set. The split of the dataset and the successive process was run on multiple random samples in order to reduce the sampling error and the uncertainty around the model. Results were checked for p-value consistency and mean squared error (MSE), which is used as an estimator that measures the difference between the estimator and what is estimated.

4 Results

4.1 Monitoring Frequency

The first part of the analysis considered the frequency of the ARMI monitoring activity since 2011.

For the 36 waterbodies selected, the Initiative recorded 639 monitoring activities since 2011 with 60 monitoring stations (Table 2), with a total of 102 events in 2011, 86 in 2012, 89 in 2013, 167 in 2014 and 195 in 2015.

Table 2 Rivers monitored by ARMI

ARMI Monitoring Site	N samples	raw %	N stations
Afon Tanat - conf Afon Rhaeadr to conf Afon Vyrnwy	30	4.69	2
Alne - conf Claverdon Bk to conf R Arrow	18	2.82	1
Ampney and Poulton Brooks (Source to Thames)	2	0.31	2
Arrow - source to Spernall Hall Fm, Studley	31	4.85	2
Bow Bk - Shell to conf R Avon	1	0.16	1
Bow Bk - source to Lett's Mill	1	0.16	1
Churn (source to Perrots Brook)	30	4.69	2
Cinderford Bk conf Blackpool Bk to Severn Estuary	14	2.19	1
Clun - conf R Unk to conf R Teme	25	3.91	3
Corve - conf Seifton Bk to conf R Teme	7	1.10	1
Dikler (Source to Wyck Rissington)	11	1.72	1
Dikler (Wyck Rissington to Windrush) and Lower Eye	20	3.13	1
Ell Bk - source to conf R Leadon	13	2.03	1
Evenlode (Bledington to Glyme confluence)	10	1.56	1
Eye (Source to Dikler)	2	0.31	2
Finham Bk - source to conf Canley Bk	32	5.01	1
Glyme (Dorn confluence to Evenlode)	6	0.94	1
Leach (Source to Thames)	27	4.23	2
Leigh-Cradley Bk - conf Suckley Bk to Teme	52	8.14	1
Lugg - conf Norton Bk to conf R Arrow	5	0.78	1
Piddle Bk - source to conf Whitsun Bk	8	1.25	1
Preston Bagot Bk - source to conf R Alne	19	2.97	1
Quinny Bk - source to conf R Onny	5	0.78	1
Rea - conf Farlow Bk to conf R Teme	31	4.85	3
Salwarpe - source to conf Elmbridge Bk	4	0.63	2
Severn - conf Bele Bk to conf Sundorne Bk	10	1.56	1
Sherbourne Brook	5	0.78	1
Stour (Worcs) - conf Smestow Bk to conf R Severn	16	2.50	2
Stour (Worcs) source to conf Smestow Bk	10	1.56	1
Teme - conf R Clun to conf R Onny	14	2.19	2

Teme - conf R Onny to conf R Severn	60	9.39	4
Teme - source to conf Ffwdwen Bk to conf R Clun	30	4.69	3
Whitsunn Bk - source to conf Piddle Bk	8	1.25	1
Windrush (Slade Barn Stream to Dikler)	10	1.56	1
Windrush and tributaries (Little Rissington to Thames)	35	5.48	3
Worfe - conf Wesley Bk to conf R Severn	37	5.79	5
	Total =639		

Looking at the ARMI monitoring activity per year, the frequency of the monitoring events increased, or remained constant, for most of the water bodies (Table 3), with only a low proportion of rivers that showed a sensible decreasing in the activity (i.e. Churn, Windrush (Little Rissington to Thames)).

Table 3 Frequency of ARMI monitoring

River Name	2011	2012	2013	2014	2015
Afon Tanat - conf Afon Rhaeadr to conf Afon Vyrnwy	0	2	2	11	15
Alne - conf Claverdon Bk to conf R Arrow	0	0	0	9	9
Ampney and Poulton Brooks (Source to Thames)	0	0	0	0	2
Arrow - source to Spernall Hall Fm, Studley	6	10	7	8	0
Bow Bk - Shell to conf R Avon	0	0	1	0	0
Bow Bk - source to Lett's Mill	0	0	1	0	0
Churn (source to Perrots Brook)	12	6	12	0	0
Cinderford Bk conf Blackpool Bk to Severn Estu-ary	0	0	0	6	8
Clun - conf R Unk to conf R Teme	3	6	0	5	11
Corve - conf Seifton Bk to conf R Teme	0	0	0	3	4
Dikler (Source to Wyck Rissington)	3	2	2	4	0
Dikler (Wyck Rissington to Windrush) and Lower Eye	5	4	1	7	3
Ell Bk - source to conf R Leadon	0	0	0	7	6
Evenlode (Bledington to Glyme confluence)	4	4	2	0	0
Eye (Source to Dikler)	0	0	0	0	2
Finham Bk - source to conf Canley Bk	5	9	5	7	6
Glyme (Dorn confluence to Evenlode)	0	0	0	6	0
Leach (Source to Thames)	7	13	3	3	1
Leigh-Cradley Bk - conf Suckley Bk to Teme	12	8	10	11	11
Lugg - conf Norton Bk to conf R Arrow	0	0	0	0	5
Piddle Bk - source to conf Whitsun Bk	0	0	2	6	0
Preston Bagot Bk - source to conf R Alne	0	0	0	9	10
Quinny Bk - source to conf R Onny	0	0	0	0	5

Rea - conf Farlow Bk to conf R Teme	0	0	3	10	18
Salwarpe - source to conf Elmbridge Bk	0	0	0	0	4
Severn - conf Bele Bk to conf Sundorne Bk	0	0	0	3	7
Sherbourne Brook	0	0	5	0	0
Stour (Worcs) - conf Smestow Bk to conf R Severn	8	2	0	0	6
Stour (Worcs) source to conf Smestow Bk	8	2	0	0	0
Teme - conf R Clun to conf R Onny	0	0	5	7	2
Teme - conf R Onny to conf R Severn	6	6	16	17	15
Teme - source to conf Ffwdwen Bk to conf R Clun	0	0	0	12	18
Whitsunn Bk - source to conf Piddle Bk	0	0	2	6	0
Windrush (Slade Barn Stream to Dikler)	2	1	3	1	3
Windrush and tributaries (Little Rissington to Thames)	14	7	5	5	4
Worfe - conf Wesley Bk to conf R Severn	7	4	2	4	20

4.2 Exploratory analysis

The distribution of single taxa scores ranged from 0 to 3 for most of the taxa (cased and caseless caddisflies, mayflies, blue winged olives and stoneflies), with olives and shrimps score ranging between 0 and 4 (Figure 16, Table 4). Scores showed similar standard deviation values, ranging from ± 0.8 to ± 1.0. Distribution of single taxa scores was not normal ($p \leq 0.05$ for Shapiro test) for the totality of taxa when considering all monitoring activities; non-normality was also observed when considering taxa scores within single rivers. When looking at single stations, taxa scores were again non-normal for over half of the sites, for all taxa (Table 5).

Table 4 Descriptive statistics for taxa monitored by ARMI

Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
cased c	639	1.2	0.8	0	3
caseless c	639	1.2	0.7	0	3
mayflies	639	0.9	0.8	0	3
bwo	639	0.9	0.9	0	3
flatbodied	639	1.0	1.0	0	3
olives	639	2.0	0.9	0	4

Figure 16 Distribution of single ARMI taxa scores

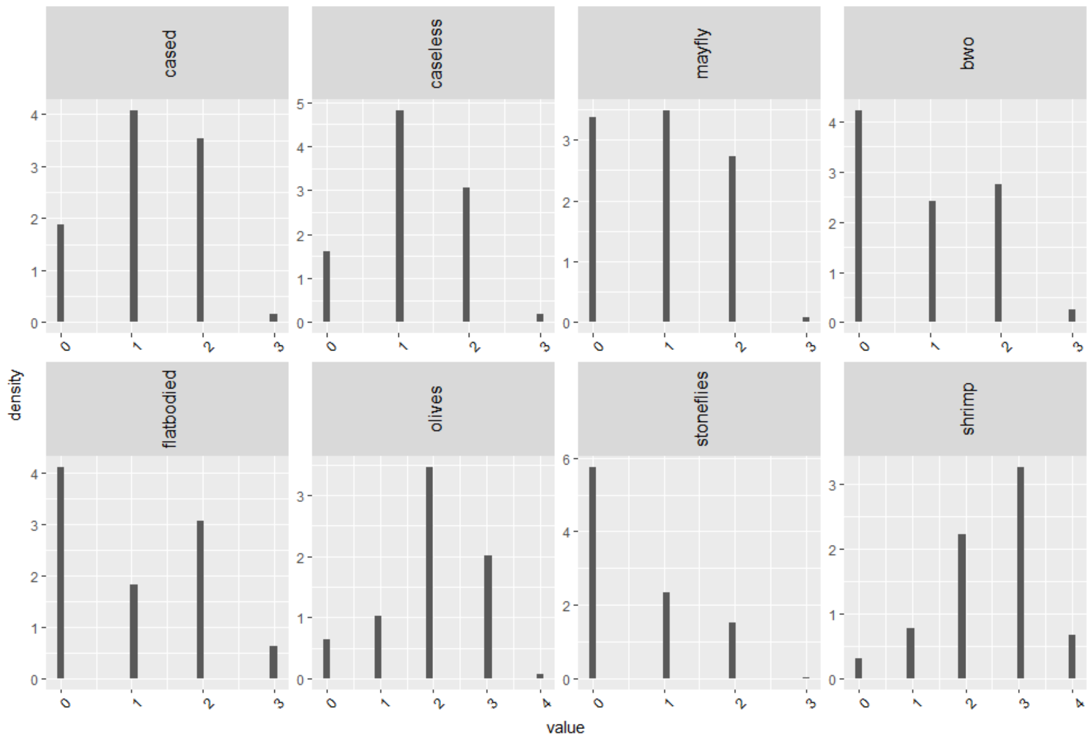


Table 5 % of non-normal distribution of single taxa scores (considering Monitoring site)

taxa	non-normal %
cased	59.18%
caseless	49.02%
mayflies	68.89%
bwo	73.33%
flatbodied	67.50%
olives	54.90%
stoneflies	76.32%
shrimps	50.98%

Concurrently, distribution of individuals per taxa showed that for most of the monitoring activities the sampled individuals for each taxon were relatively low in numbers (mean between 4.8 and 21.2), with the exception of olives and shrimps, which showed a higher mean of sampled individuals (mean of 275.9 for shrimps and 99.9 for olives).

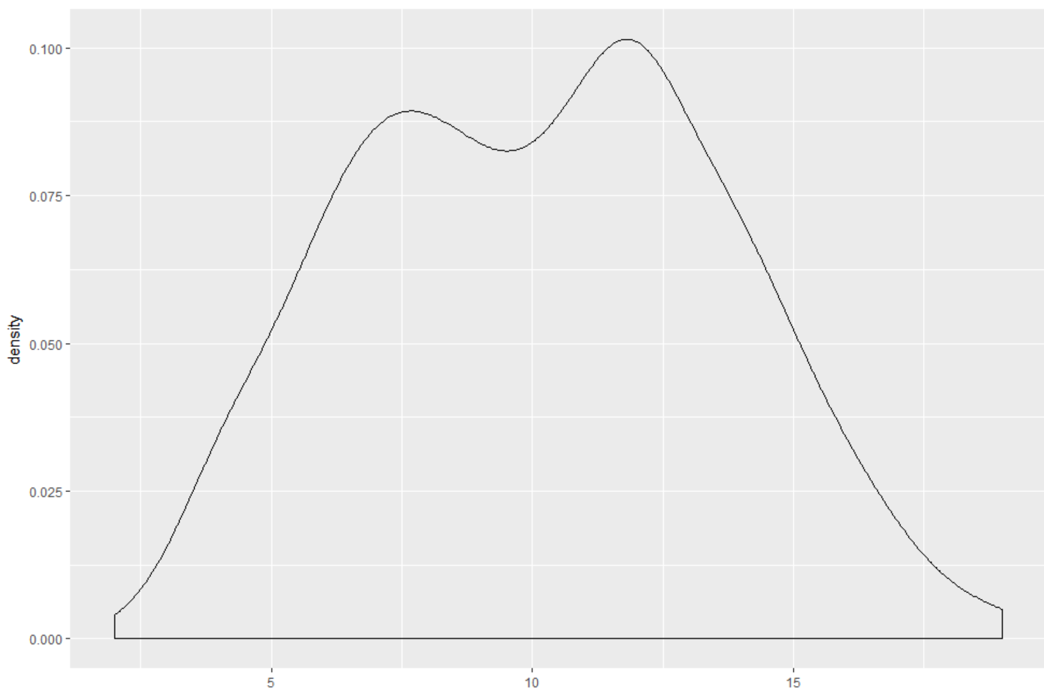
Table 6 Descriptive statistics for ARMI taxa - number of individuals sampled

```
Descriptive statistics
==========================================================
Statistic            N     Mean    St. Dev.   Min     Max
----------------------------------------------------------
cased c             639    13.8      21.6      0      210
caseless c          639    12.8      29.7      0      524
mayflies            639     9.9      16.7      0      100
bwo                 639    13.6      32.0      0      400
flatbodied          639    21.2      42.7      0      310
olives              639    99.9     176.1      0     1.350
```

When analysing the scores obtained with the riverfly methodology, ARMI score showed a bimodal distribution, with two peaks around the value of 12 (modal value) and between the scores of 7 and 8 (Figure 17).

```
ARMI Score
    Shapiro-Wilk normality test
    W = 0.97974, p-value = 9.842e-08
```

Figure 17 Distribution of the ARMI score



Kruskal-Wallis test reported a low p-value (≤0.05) for all the ARMI samples amongst rivers, indicating a statistical significant difference in the scores.

```
Kruskal-Wallis rank sum test
Kruskal-Wallis chi-squared = 36.671, df = 17, p-value = 0.003727
```

Analysing the historical data per river, averaged ARMI scores were quite constant, with year-to-year differences ≤ 2 points for around 65% of the cases and within 3 points for around 80%.

Table 7 Average ARMI Score over the years

River Name	2011	2012	2013	2014	2015
Afon Tanat - conf Afon Rhaeadr to conf Afon Vyrnwy		10	10.50	11.64	11.73
Alne - conf Claverdon Bk to conf R Arrow				13.89	11.22
Ampney and Poulton Brooks (Source to Thames)					9.50
Arrow - source to Spernall Hall Fm, Studley	6.33	5.80	6.00	7.38	
Bow Bk - Shell to conf R Avon			9.00		
Bow Bk - source to Lett's Mill			8.00		
Churn (source to Perrots Brook)	11.50	12.17	11.50		
Cinderford Bk conf Blackpool Bk to Severn Estuary				7.83	7.88
Clun - conf R Unk to conf R Teme	10.67	15.50		13.60	11.73
Corve - conf Seifton Bk to conf R Teme				8.00	6.50
Dikler (Source to Wyck Rissington)	9.33	7.00	8.00	8.75	
Dikler (Wyck Rissington to Windrush) and Lower Eye	9.80	9.75	11.00	11.86	12.00
Ell Bk - source to conf R Leadon				10.71	10.17
Evenlode (Bledington to Glyme confluence)	5.50	8.75	5.50		
Eye (Source to Dikler)					9.50
Finham Bk - source to conf Canley Bk	5.20	6.78	8.20	7.43	5.17
Glyme (Dorn confluence to Evenlode)				7.17	
Leach (Source to Thames)	9.14	11.54	10.33	9.00	11.00
Leigh-Cradley Bk - conf Suckley Bk to Teme	13.75	15.63	14.40	17.18	15.00
Lugg - conf Norton Bk to conf R Arrow					8.60

Piddle Bk - source to conf Whitsun Bk			3.50	5.50	
Preston Bagot Bk - source to conf R Alne				9.22	9.40
Quinny Bk - source to conf R Onny					14.80
Rea - conf Farlow Bk to conf R Teme			7.67	11.20	13.17
Salwarpe - source to conf Elmbridge Bk					7.75
Severn - conf Bele Bk to conf Sundorne Bk				6.33	6.29
Sherbourne Brook			11.40		
Stour (Worcs) - conf Sme-stow Bk to conf R Severn	6.50	6.00			5.33
Stour (Worcs) source to conf Smestow Bk	4.50	4.50			
Teme - conf R Clun to conf R Onny			11.20	10.86	5.50
Teme - conf R Onny to conf R Severn	10.33	10.50	8.69	11.76	13.40
Teme - source to conf Ffwdwen Bk to conf R Clun				9.00	8.11
Whitsunn Bk - source to conf Piddle Bk			4.00	4.50	
Windrush (Slade Barn Stream to Dikler)	12.00	12.00	14.00	16.00	13.67
Windrush and tributaries (Little Rissington to Thames)	10.71	13.86	15.40	14.00	13.50
Worfe - conf Wesley Bk to conf R Severn	8.14	10	11.00	8.25	9.05

Nevertheless, when looking at the spread of the score within the year, results illustrated that many sites had a higher degree of variance, with over 60% reporting a variance of ≥ 2 points within the year, with around 43% showing a variance of over 3 points (Table 8)

Table 8 Intra-annual variance of ARMI score for each river

River me	2011	2012	2013	2014	2015
Afon Tat - conf Afon Rhaeadr to conf Afon Vyrnwy		0.00	0.50	2.85	3.50
Alne - conf Claverdon Bk to conf R Arrow				7.86	8.69
Ampney and Poulton Brooks (Source to Thames)					4.50
Arrow - source to Sperll Hall Fm, Studley	0.27	1.96	1.67	9.41	
Bow Bk - Shell to conf R Avon					
Bow Bk - source to Lett's Mill					
Churn (source to Perrots Brook)	0.82	0.57	0.45		
Cinderford Bk conf Blackpool Bk to Severn Estuary				0.57	1.55
Clun - conf R Unk to conf R Teme	2.33	3.10		1.30	10.62
Corve - conf Seifton Bk to conf R Teme				3.00	11.00

Dikler (Source to Wyck Rissington)	4.33	8.00	2.00	1.58	
Dikler (Wyck Rissington to Windrush) and Lower Eye	0.70	0.92		0.81	3.00
Ell Bk - source to conf R Leadon				4.90	2.97
Evenlode (Bledington to Glyme confluence)	3.00	6.25	0.50		
Eye (Source to Dikler)					4.50
Finham Bk - source to conf Canley Bk	1.20	4.94	2.20	1.62	2.57
Glyme (Dorn confluence to Evenlode)				4.17	
Leach (Source to Thames)	7.48	4.60	1.33	1.00	
Leigh-Cradley Bk - conf Suckley Bk to Teme	3.30	4.27	2.93	1.56	0.60
Lugg - conf Norton Bk to conf R Arrow					7.30
Piddle Bk - source to conf Whitsun Bk			0.50	1.90	
Preston Bagot Bk - source to conf R Alne				2.94	1.38
Quinny Bk - source to conf R Onny					5.70
Rea - conf Farlow Bk to conf R Teme			2.33	4.40	4.62
Salwarpe - source to conf Elmbridge Bk					2.92
Severn - conf Bele Bk to conf Sundorne Bk				6.33	4.90
Sherbourne Brook			1.30		
Stour (Worcs) - conf Smestow Bk to conf R Severn	0.29	0.00			1.07
Stour (Worcs) source to conf Smestow Bk	0.29	0.50			
Teme - conf R Clun to conf R Onny			4.70	3.14	4.50
Teme - conf R Onny to conf R Severn	12.67	9.90	3.70	4.07	2.40
Teme - source to conf Ffwdwen Bk to conf R Clun				7.45	3.05
Whitsunn Bk - source to conf Piddle Bk			2.00	2.70	
Windrush (Slade Barn Stream to Dikler)	0.00		3.00		2.33
Windrush and tributaries (Little Rissington to Thames)	6.84	3.14	4.30	4.50	1.67
Worfe - conf Wesley Bk to conf R Severn	2.14	2.00	2.00	18.92	11.73

Looking at seasonality, Kruskal-Wallis test results indicated that ARMI score distributions were significantly different (p-value $\leq$ 0.05) when considering scores obtained in different months, and highly significant different (p-value $\leq$ 0.01) when considering records achieved in warm (April to September) and cold (October to March) months:

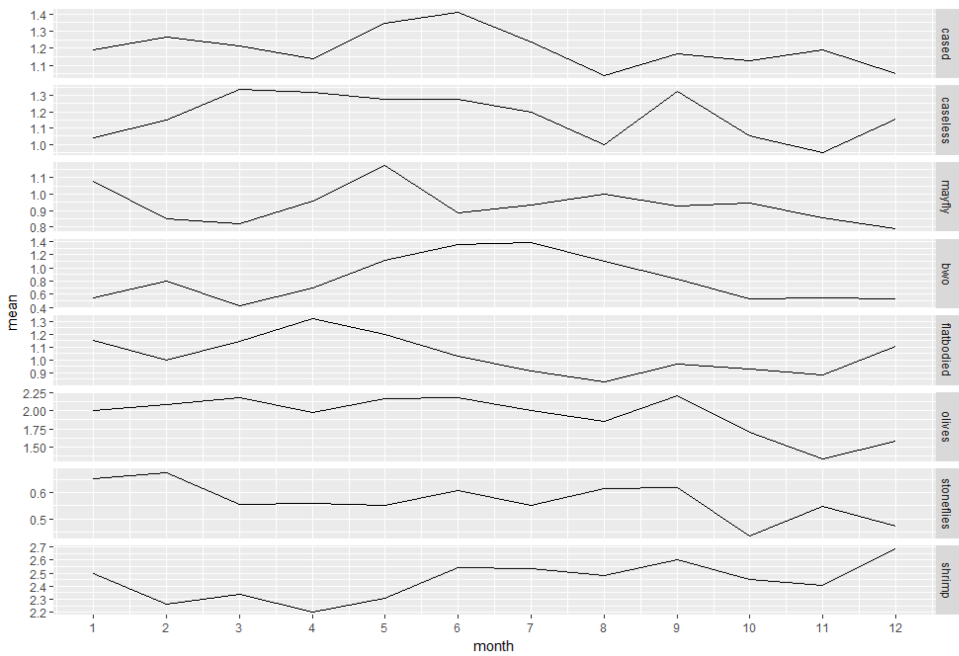
Table 9 Kruskal-Wallis test

```
Month - Test
        Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 21.994, df = 11, p-value = 0.02442
===============================================================




Warm/Cold - Test
        Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 10.99, df = 1, p-value = 0.0009159
================================================================
```

When running the Kruskal-Wallis test for single rivers, the analysis was not statistically significant for almost the entirety of the dataset with the test unable to run for a proportion of the dataset due to the low number of samples in each sub-group. Same results were obtained when running the test for single monitoring sites.

However, analysis of average taxa score per month confirmed that seasonality variation was present in the samples collected by the volunteers for almost all taxa. In particular, the plotted results show high differences on monthly basis.

Figure 18 Average ARMI Taxa score per month



4.3 ARMI monitoring taxa correlation

Analysis of pairwise correlation plots of ARMI taxa showed that no taxa were highly correlated with the others. All pairwise correlation results returned values lower than 0.3 for both Spearman and Paerson correlation tests, with only mayflies and cased caddisflies showing a moderate correlation with a value around 0.35 for both coefficients.

Figure 19 Pairwise correlation plot - ARMI taxa (individuals)

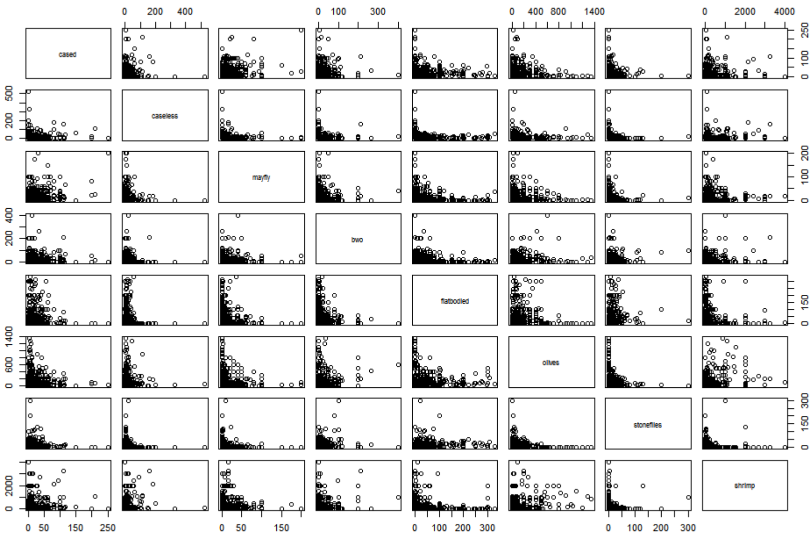
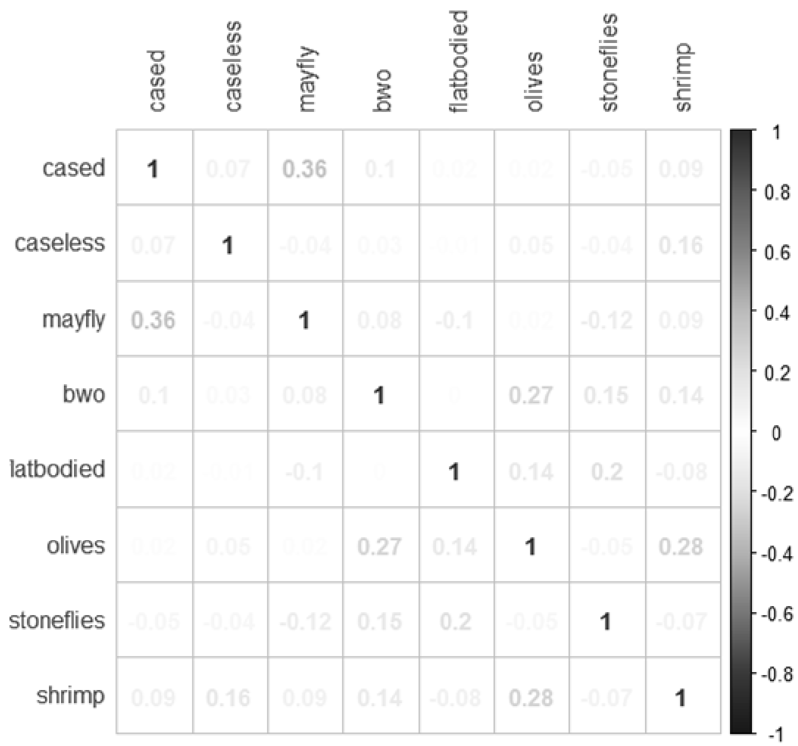


Figure 20 Correlation matrix - ARMI taxa



4.4 Comparison of monitoring results

The EA monitored the 36 waterbodies with a frequency that was different over the years from the ARMI volunteers. In total, from 2011, 397 sampling events have been run by the EA in contrast with the 639 run by the ARMI volunteers.

Table 10 Rivers monitored by EA

Rivers	N samples	raw %	cumulative %
Afon Tanat - conf Afon Rhaeadr to conf Afon Vyrnwy	12	3.02	3.02
Alne - conf Claverdon Bk to conf R Arrow	8	2.02	5.04
Ampney and Poulton Brooks (Source to Thames)	15	3.78	8.82
Arrow - source to Spernall Hall Fm, Studley	7	1.76	10.58
Bow Bk - Shell to conf R Avon	12	3.02	13.60
Bow Bk - source to Lett's Mill	36	9.07	22.67
Churn (source to Perrots Brook)	12	3.02	25.69
Cinderford Bk conf Blackpool Bk to Severn Estuary	2	0.50	26.20
Clun - conf R Unk to conf R Teme	22	5.54	31.74
Corve - conf Seifton Bk to conf R Teme	10	2.52	34.26
Dikler (Source to Wyck Rissington)	8	2.02	36.27
Dikler (Wyck Rissington to Windrush) and Lower Eye	3	0.76	37.03
Ell Bk - source to conf R Leadon	15	3.78	40.81
Evenlode (Bledington to Glyme confluence)	8	2.02	42.82
Eye (Source to Dikler)	3	0.76	43.58
Finham Bk - source to conf Canley Bk	15	3.78	47.36
Glyme (Dorn confluence to Evenlode)	3	0.76	48.11
Leach (Source to Thames)	20	5.04	53.15
Leigh-Cradley Bk - conf Suckley Bk to Teme	11	2.77	55.92
Lugg - conf Norton Bk to conf R Arrow	10	2.52	58.44
Piddle Bk - source to conf Whitsun Bk	3	0.76	59.19
Preston Bagot Bk - source to conf R Alne	5	1.26	60.45

Quinny Bk - source to conf R Onny	2	0.50	60.96
Rea - conf Farlow Bk to conf R Teme	10	2.52	63.48
Salwarpe - source to conf Elmbridge Bk	22	5.54	69.02
Severn - conf Bele Bk to conf Sundorne Bk	17	4.28	73.30
Sherbourne Brook	6	1.51	74.81
Stour (Worcs) - conf Smestow Bk to conf R Severn	12	3.02	77.83
Stour (Worcs) source to conf Smestow Bk	4	1.01	78.84
Teme - conf R Clun to conf R Onny	12	3.02	81.86
Teme - conf R Onny to conf R Severn	24	6.05	87.91
Teme - source to conf Ffwdwen Bk to conf R Clun	2	0.50	88.41
Whitsunn Bk - source to conf Piddle Bk	6	1.51	89.92
Windrush (Slade Barn Stream to Dikler)	17	4.28	94.21
Windrush and tributaries (Little Rissington to Thames)	11	2.77	96.98
Worfe - conf Wesley Bk to conf R Severn	12	3.02	100.00

The frequency of monitoring events run by the ARMI volunteers per year was higher for 72% of the rivers, with an average increase in monitoring activity for these 26 waterbodies of 142%. Considering all waterbodies, including the 10 where EA had run more sampling, the average increase for ARMI was still around 89%.

Figure 21 Average Monitoring activities per year

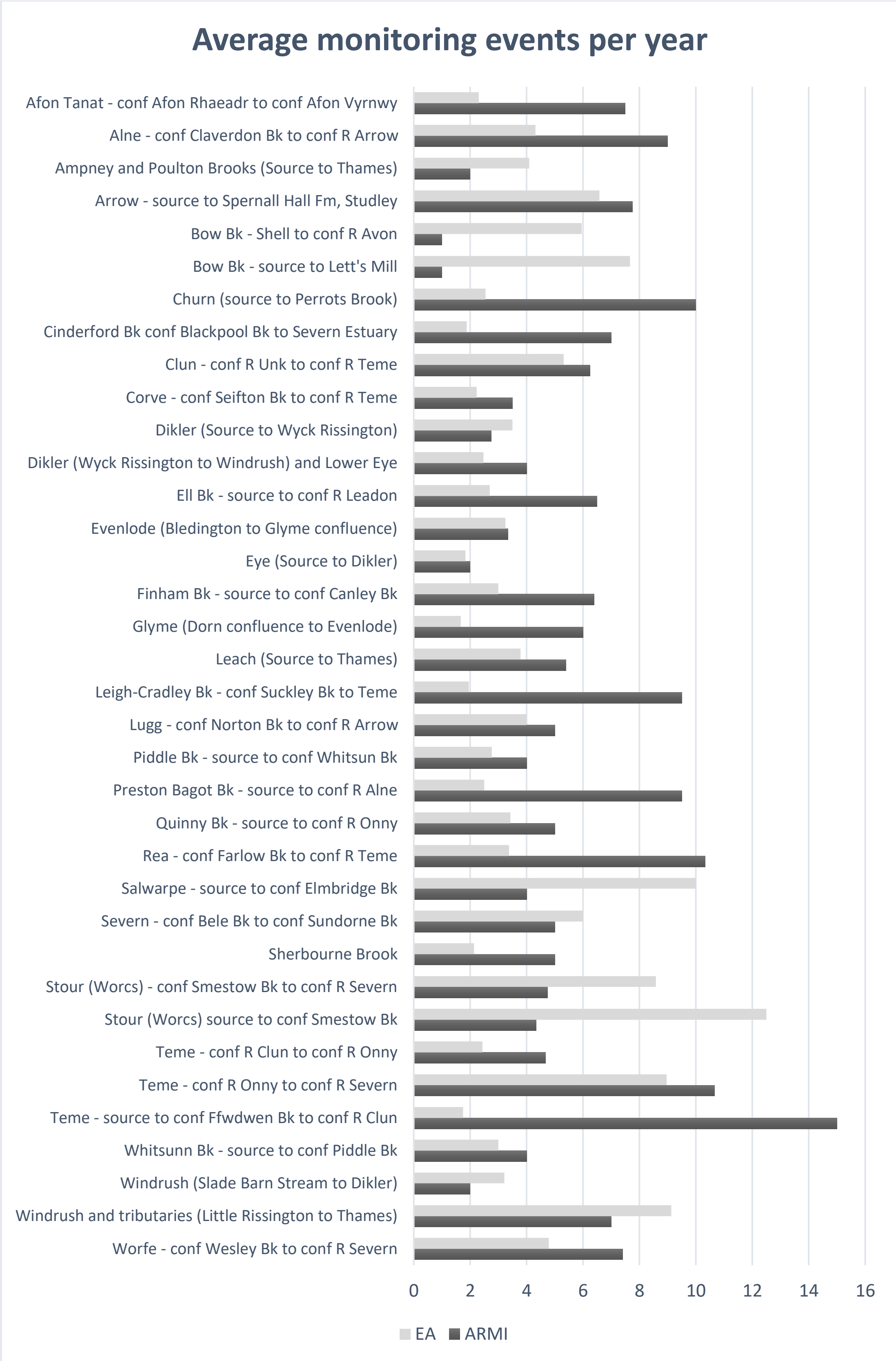
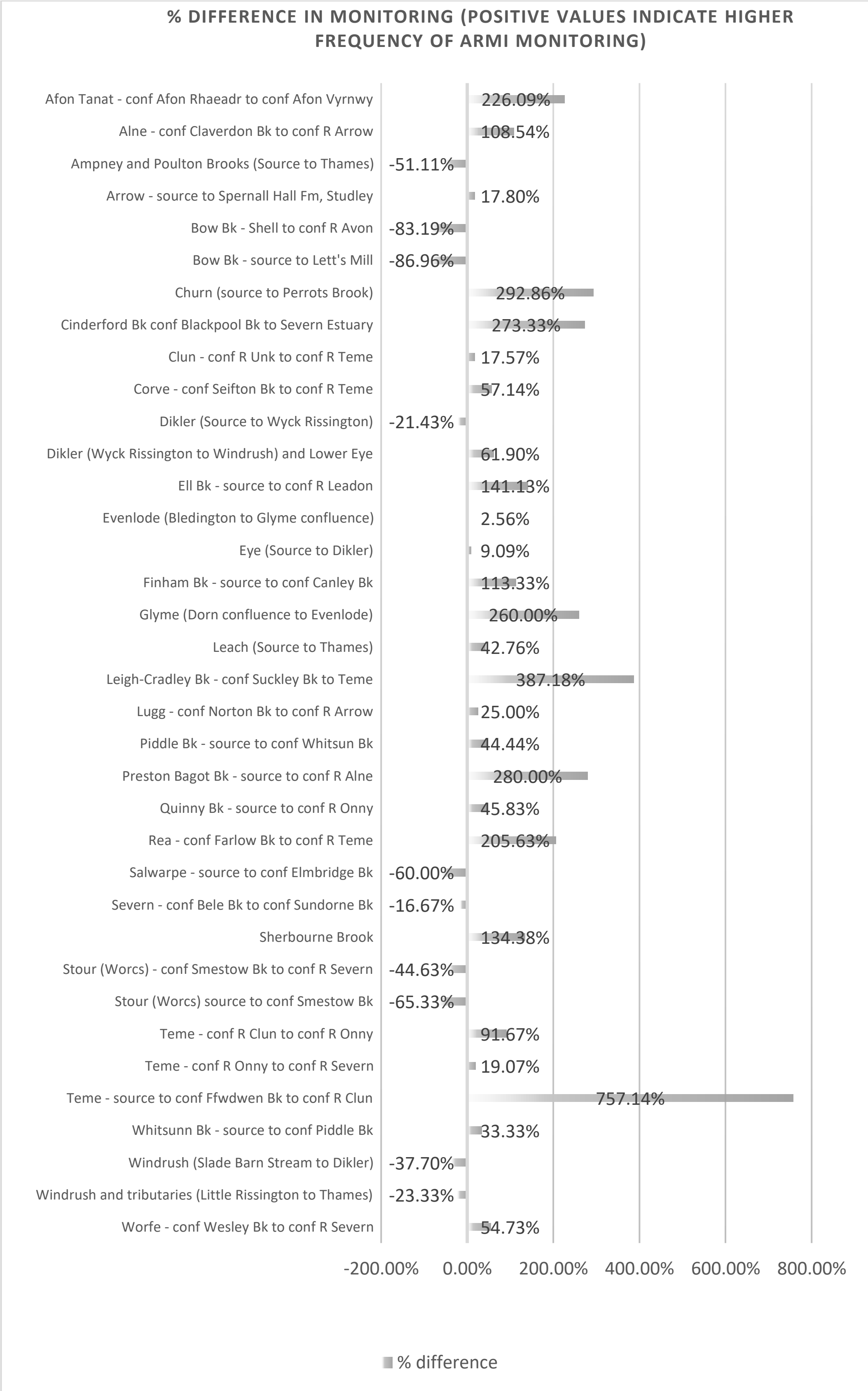


Figure 22 % Difference in monitoring activity between ARMI and EA

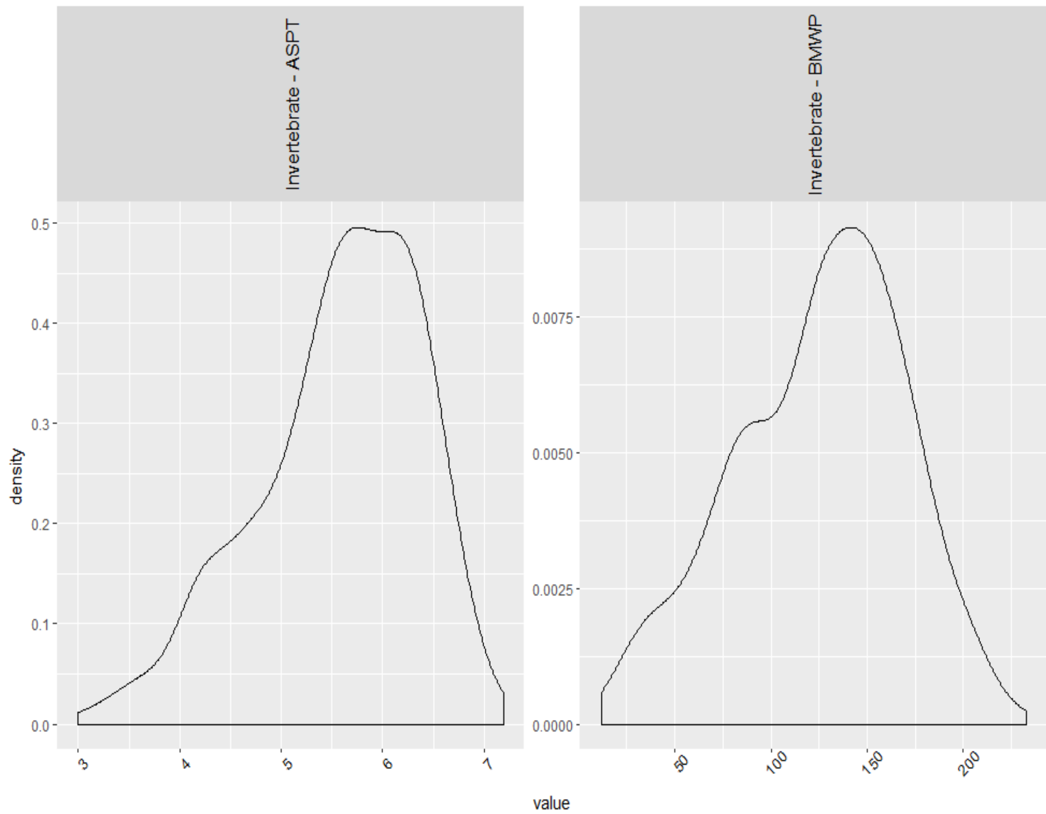


Distribution of ASPT and BMWP scores followed a non-normal distribution (p-value ≤ 0.05 for Shapiro test) for both parameters, with density plots showing a slightly right-skewed distribution for both scores.

Table 11 Normality tests for ASPT and BMWP scores

```
ASPT
    Shapiro-Wilk normality test
     W = 0.96473, p-value = 3.499e-08

BMWP
    Shapiro-Wilk normality test
    W = 0.98396, p-value = 0.0002191
```

Figure 23 ASPT and BMWP scores distributions



Analysing average invertebrates scores by months, ARMI scores were higher between May and July and in September. ASPT were higher between June and August, while BMWP in April, and between July and September.

Figure 24 Average ARMI score by month

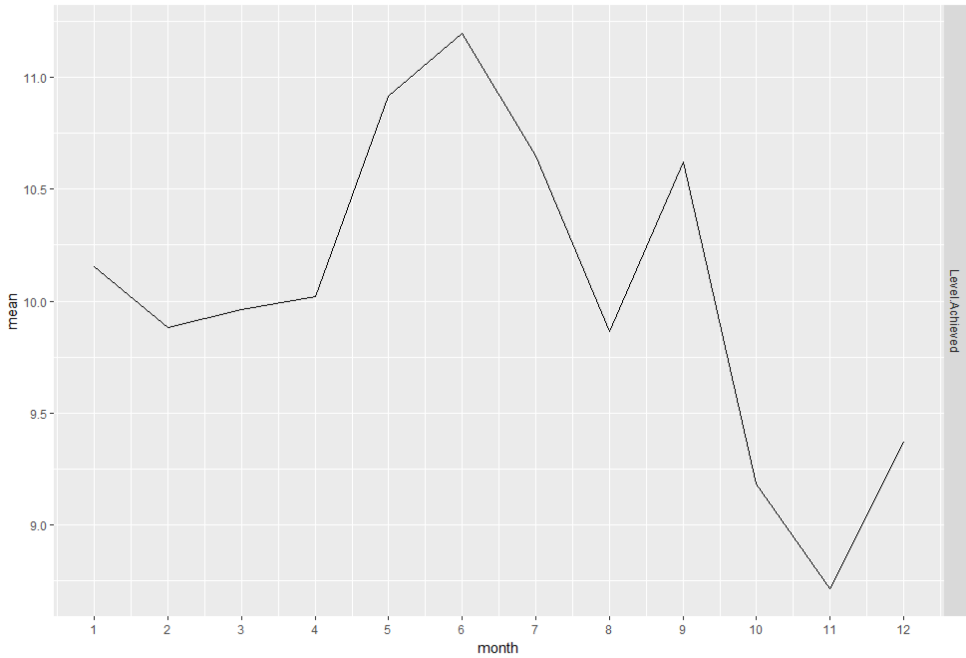
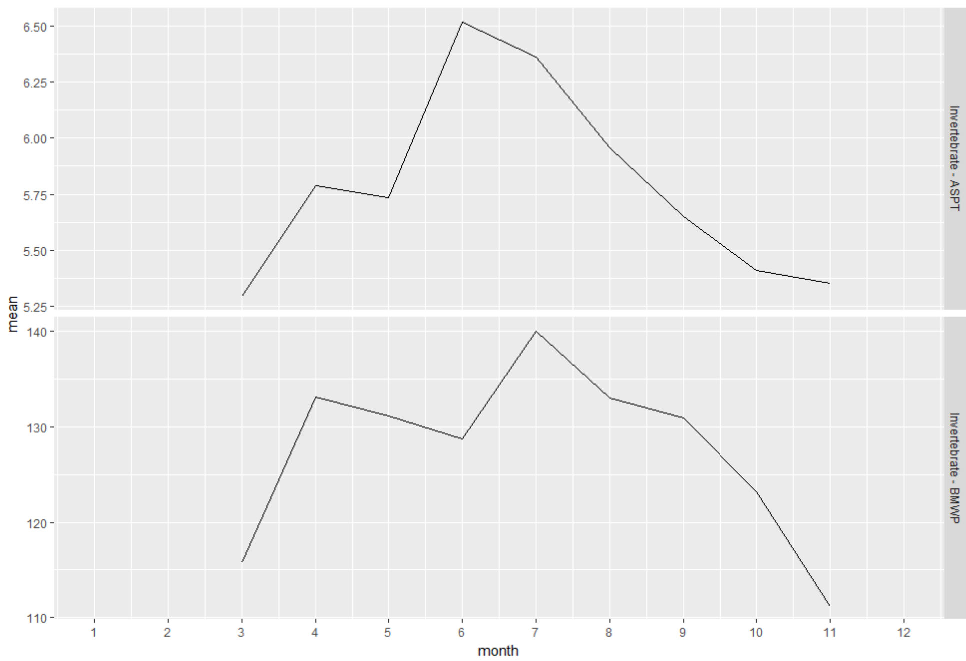


Figure 25 Average Invertebrate scores by month



4.5 Comparison of data collected within a 14-days' time frame

The dataset obtained combining monitoring events from ARMI and statutory agencies occurred within a 14-days' time frame, consisted of 73 observations for 22 water bodies. The highest proportion of the score reported a good or very good river quality category, with only a small number of samples reporting moderate or poor results.

In details,

ASPT river category scores for this dataset were:

```
good   moderate   poor very good
 36          4    1         32
```

BMWP were:

```
 good  moderate    poor   very good
 34        8        1         30
```

Distribution of scores in this combined dataset is showed in the following figures.

Figure 26 Boxplots of invertebrates scores for matching dataset

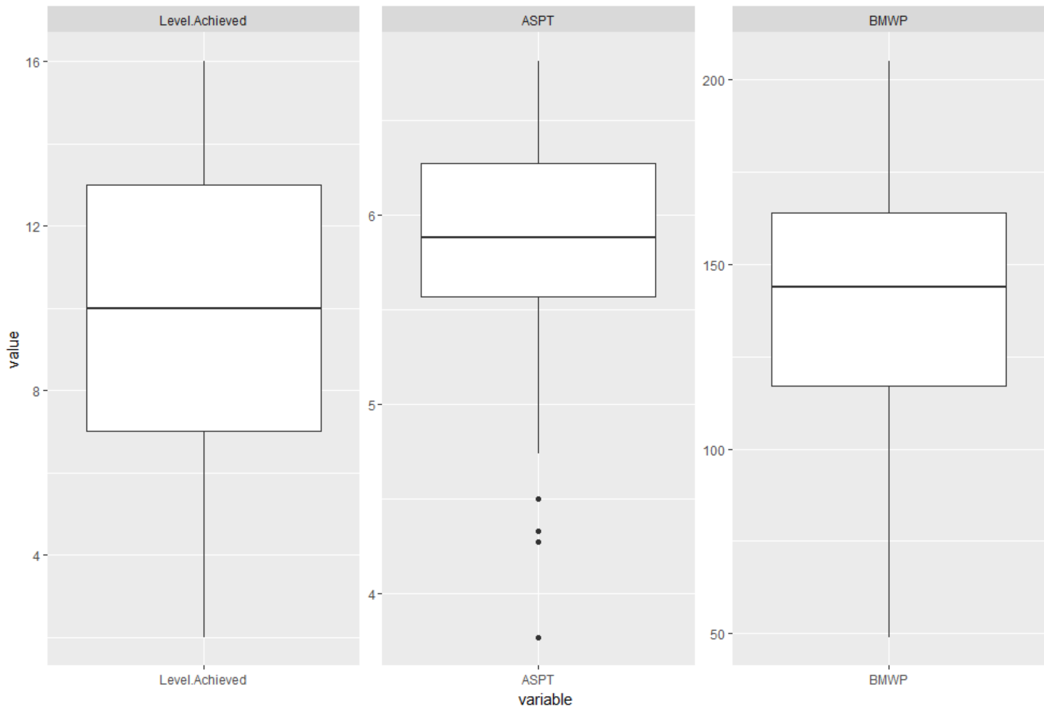
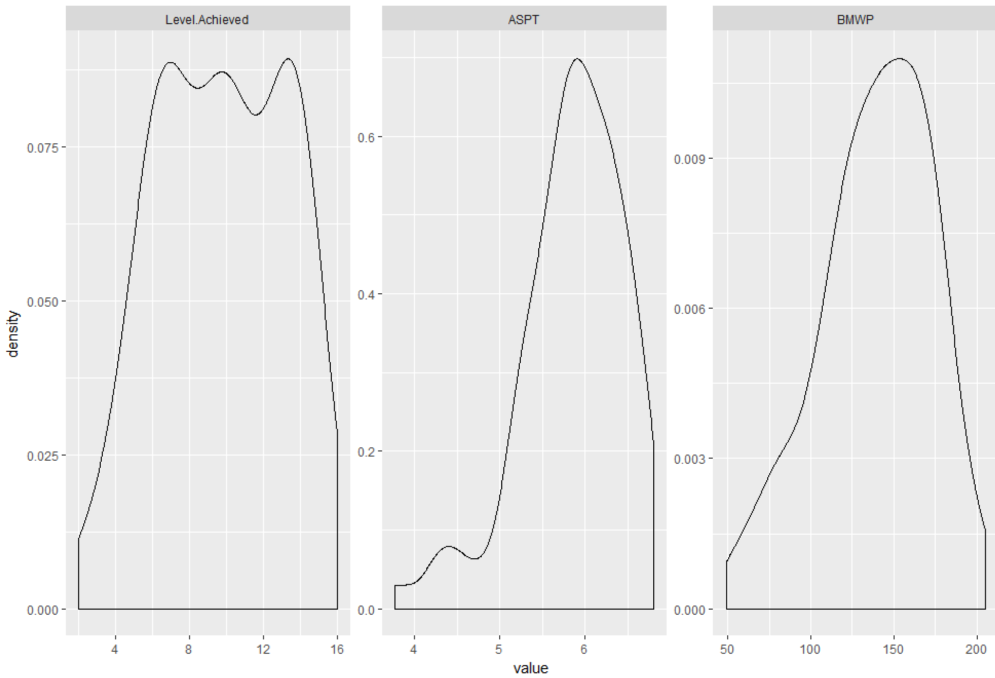


Figure 27 Density plot of invertebrates scores in matching dataset



Again, distribution of professionals' scores showed evident skewedness, with ARMI scores instead displaying 3 peaks with a multimodal distribution. Distributions appeared diverse amongst the biotic indices, with Wilcoxon test and Kolmogorov-Smirnov test confirming that all the scores distributions were significantly different among them (p-value ≤ 0.05):

Table 12 Distribution tests

```
ASPT, ARMI Score
  # Wilcoxon rank sum test with continuity correction
  # W = 0, p-value < 2.2e-16

BMWP, ARMI Score
  # Wilcoxon rank sum test with continuity correction
  # W = 0, p-value < 2.2e-16

  ASPT, BMWP
# Wilcoxon rank sum test with continuity correction
  # W = 0, p-value < 2.2e-16
======================================================
      ASPT, ARMI Score
Two-sample Kolmogorov-Smirnov test
```

```
D = 0.83212, p-value < 2.2e-16
alternative hypothesis: two-sided

      BMWP, ARMI Score

Two-sample Kolmogorov-Smirnov test

D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided


   ASPT, BMWP

Two-sample Kolmogorov-Smirnov test

D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Boxplots suggested higher variance for ARMI score. Analysis of variance with Levene's and Bartlett's tests reported extremely low p-value, hence confirming sufficient evidence to claim that the variances were not equal.

Table 13 Homogeneity of Variance tests

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value    Pr(>F)
group   2   424.1 < 2.2e-16 ***
      408
============================================================

Bartlett test of homogeneity of variances

Bartlett's K-squared = 1280, df = 2, p-value < 2.2e-16
```

4.6 Correlation among scores

Considering the dataset of data with a 14-days' timeframe, correlation analysis among ARMI score and ASPT and BMWP scores resulted in low positive Spearman correlation values for ARMI and BMWP scores, with a moderate positive correlation between ARMI and ASPT scores (Figure 28 and Figure 29). Correlation between ASPT and BMWP was a lot higher (~ 0.5), as expected.

Table 14 Correlation among scores

```
Spearman's rank correlation rho

ASPT, ARMI Scores


S = 48597, p-value = 0.03269

alternative hypothesis: true rho is not equal to 0

sample estimates:

      rho

0.2503173



        Spearman's rank correlation rho

BMWP, ARMI Scores

S = 57005, p-value = 0.3094

alternative hypothesis: true rho is not equal to 0

sample estimates:

     rho

0.12062



        Spearman's rank correlation rho

BMWP, ASPT Score


S = 31540, p-value = 3.394e-06
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.5134587
```

Figure 28 Plot of ARMI and ASPT score

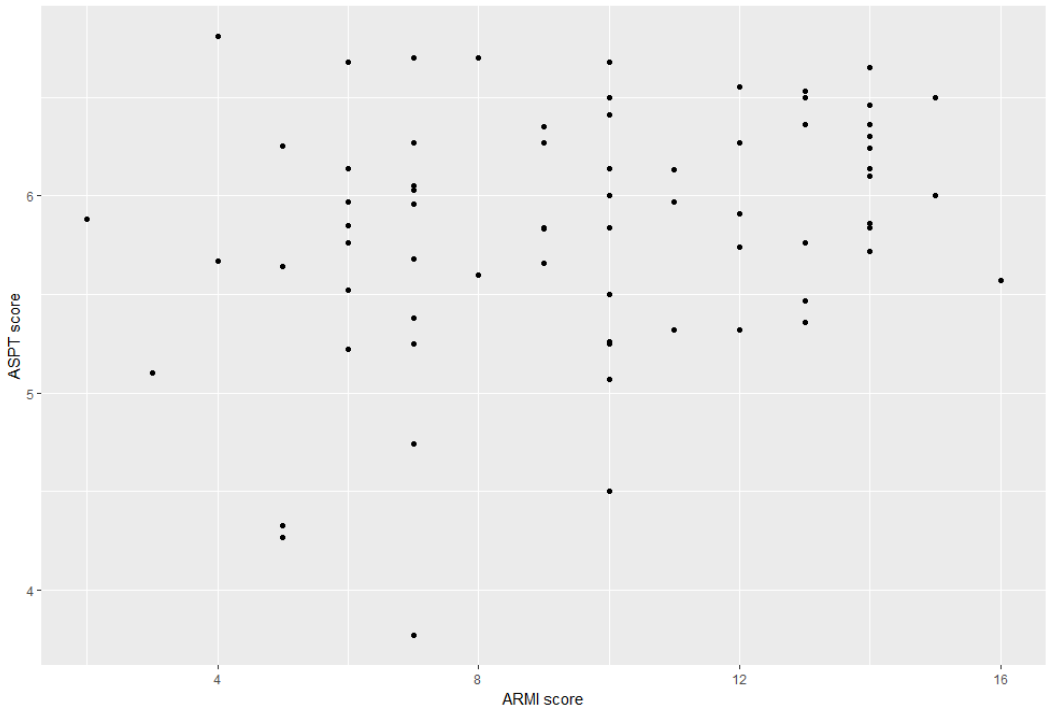
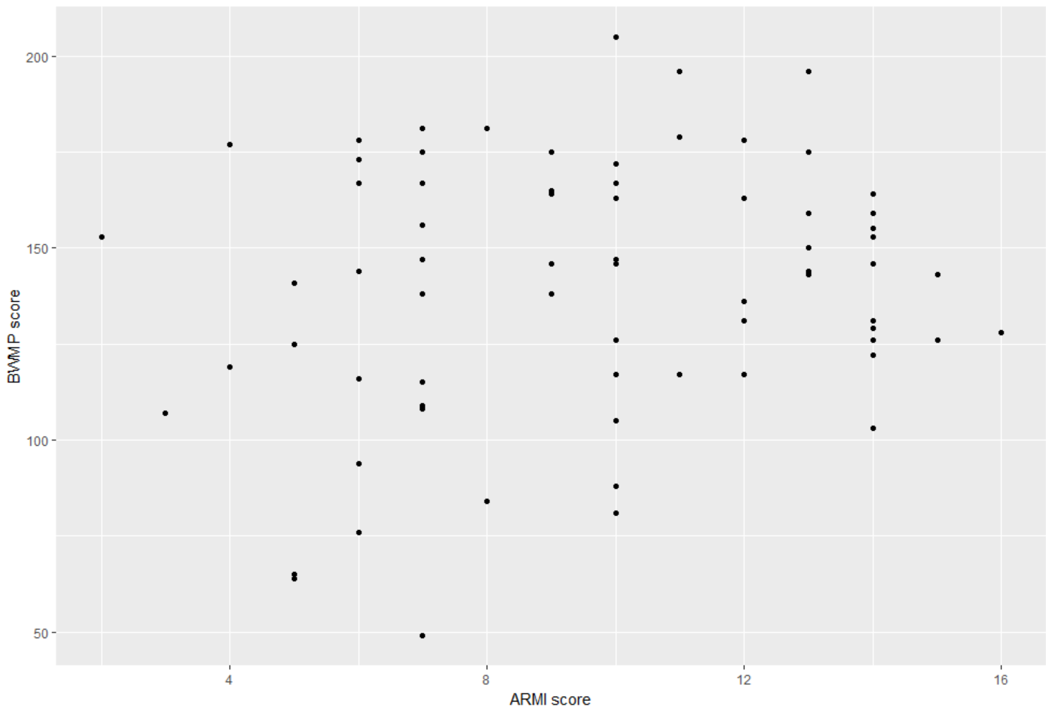


Figure 29 Plot of Armi and BMWP scores



4.7 Multiple Linear Regression Model

4.7.1 Model development using ARMI score as basic predictor

Multiple linear regression models built on ARMI score as predictor showed statistical significance (p-value ≤ 0.5); however, the models presented low predicting power, explaining just 6% of the variance for ASPT and 2% for the BMWP score.

Table 15 ASPT basic model

```
lm(formula = `Invertebrate - ASPT` ~ Level.Achieved, data = combined
)

Residuals:
     Min       1Q   Median       3Q      Max
-1.94860 -0.33418  0.05964  0.39788  1.23522

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.38302    0.20695  26.012   <2e-16 ***
Level.Achieved  0.04794    0.02010   2.385   0.0197 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.587 on 71 degrees of freedom
Multiple R-squared:  0.07418, Adjusted R-squared:  0.06114
F-statistic: 5.689 on 1 and 71 DF,  p-value: 0.01974
```

Table 16 BMWP basic model

```
lm(formula = `Invertebrate - BMWP` ~ Level.Achieved, data = combined
)

Residuals:
    Min      1Q  Median      3Q     Max
-85.199 -22.736   3.801  27.109  65.264

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     121.279     11.751  10.320 8.99e-16 ***
Level.Achieved    1.846      1.141   1.617     0.11
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.33 on 71 degrees of freedom
Multiple R-squared:  0.03553, Adjusted R-squared:  0.02194
F-statistic: 2.615 on 1 and 71 DF,  p-value: 0.1103
```

Introducing the number of sampled taxa for each monitoring events in the MLR model resulted in an increasing in the variance explained up to 22% for the ASPT and

16% for the BMWP scores respectively. Anova confirmed that the introduction of number of taxa as predictor was statistically significant (p-value ≤ 0.05 for both models).

Table 17 ASPT model (ARMI score + taxa number)

```
lm(formula = `Invertebrate - ASPT` ~ Level.Achieved + count_taxa,
    data = combined)

Residuals:
     Min       1Q   Median       3Q      Max
-1.82687 -0.31442  0.09375  0.30852  1.30898

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     4.94583    0.21801  22.686  < 2e-16 ***
Level.Achieved -0.05353    0.03139  -1.705 0.092630 .
count_taxa      0.25643    0.06450   3.976 0.000169 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.534 on 70 degrees of freedom
Multiple R-squared:  0.2447,   Adjusted R-squared:  0.2232
F-statistic: 11.34 on 2 and 70 DF,  p-value: 5.415e-05
```

```
Analysis of Variance Table

Model 1: `Invertebrate - ASPT` ~ Level.Achieved
Model 2: `Invertebrate - ASPT` ~ Level.Achieved + count_taxa
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     71 24.465
2     70 19.958  1    4.5067 15.807 0.0001686 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 18 BMWP model (ARMI score + taxa number)

```
lm(formula = `Invertebrate - BMWP` ~ Level.Achieved + count_taxa,
    data = combined)

Residuals:
    Min      1Q  Median      3Q     Max
-78.904 -19.904  -0.688  20.574  68.441

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      98.671     12.616   7.821  3.9e-11 ***
Level.Achieved   -3.402      1.817  -1.872 0.065337 .
count_taxa       13.261      3.732   3.553 0.000687 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.9 on 70 degrees of freedom
Multiple R-squared:  0.1829,   Adjusted R-squared:  0.1595
F-statistic: 7.833 on 2 and 70 DF,  p-value: 0.0008512
```

```
Analysis of Variance Table

Model 1: `Invertebrate - BMWP` ~ Level.Achieved
Model 2: `Invertebrate - BMWP` ~ Level.Achieved + count_taxa
  Res.Df   RSS Df Sum of Sq      F   Pr(>F)
1     71 78886
2     70 66834  1     12052 12.623 0.000687 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4.7.2 Model development using single taxa scores as basic predictors

When using the scores of each single ARMI taxa, the overall prediction power of the models showed an evident increase for the ASPT model, with an adjusted R-squared value of around 0.30. BMWP model adjusted R-squared of 0.05 was only slightly higher than the basic model built on the ARMI score.

Table 19 ASPT taxa model

```
lm(formula = `Invertebrate - ASPT` ~ cased + caseless + mayfly + 
    stoneflies + olives + bwo + flatbodied + shrimp, data = combined
)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.56619 -0.26187  0.09081  0.33816  0.87312 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.09384    0.24619  24.753  < 2e-16 ***
cased        0.11211    0.10171   1.102  0.27449    
caseless    -0.12874    0.10505  -1.226  0.22486    
mayfly       0.01102    0.09747   0.113  0.91036    
stoneflies   0.18518    0.08577   2.159  0.03459 *  
olives      -0.20880    0.08016  -2.605  0.01142 *  
bwo          0.12278    0.07844   1.565  0.12243    
flatbodied   0.22732    0.07126   3.190  0.00221 ** 
shrimp      -0.09734    0.07440  -1.308  0.19545    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5069 on 64 degrees of freedom
Multiple R-squared:  0.3777,   Adjusted R-squared:  0.2999 
F-statistic: 4.855 on 8 and 64 DF,  p-value: 0.0001051
```

Table 20 BMWP taxa model

```
lm(formula = `Invertebrate - BMWP` ~ cased + caseless + mayfly + 
    stoneflies + olives + bwo + flatbodied + shrimp, data = combined
)

Residuals:
    Min      1Q  Median      3Q     Max 
```

```
-76.595 -22.175   4.538  25.171  59.507

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 142.5433    15.9951   8.912 8.07e-13 ***
cased         6.8205     6.6081   1.032   0.3059    
caseless     -3.9644     6.8251  -0.581   0.5634    
mayfly       -0.3909     6.3329  -0.062   0.9510    
stoneflies   -4.5106     5.5724  -0.809   0.4213    
olives       -3.9328     5.2080  -0.755   0.4529    
bwo           5.5511     5.0961   1.089   0.2801    
flatbodied   10.7644     4.6301   2.325   0.0233 *  
shrimp       -4.3933     4.8340  -0.909   0.3668    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.93 on 64 degrees of freedom
Multiple R-squared:  0.1513,	Adjusted R-squared:  0.0452 
F-statistic: 1.426 on 8 and 64 DF,  p-value: 0.203
```

Replacing the ARMI score with the single taxa scores increased the model power by a sensible degree. ASPT models adjusted R-squared increased to 0.36, while BMWP to 0.11. Anova analysis confirmed the significance of the taxa count predictor in the developed models.

Table 21 ASPT advanced taxa model (taxa + taxa number)

```
lm(formula = `Invertebrate - ASPT` ~ cased + caseless + mayfly + 
    stoneflies + olives + bwo + flatbodied + shrimp + count_taxa, 
    data = combined)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.56638 -0.17556  0.03321  0.30746  0.90433 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.51233    0.32069  17.189  < 2e-16 ***
cased        0.05277    0.09968   0.529  0.59844    
caseless    -0.21064    0.10495  -2.007  0.04904 *  
mayfly      -0.10560    0.10288  -1.026  0.30859    
stoneflies   0.04487    0.09737   0.461  0.64649    
olives      -0.25555    0.07856  -3.253  0.00184 ** 
bwo         -0.04302    0.09736  -0.442  0.66009    
flatbodied   0.12966    0.07730   1.677  0.09844 .  
shrimp      -0.08107    0.07134  -1.136  0.26011    
count_taxa   0.21645    0.08114   2.668  0.00970 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4843 on 63 degrees of freedom
Multiple R-squared:  0.4408,	Adjusted R-squared:  0.361 
F-statistic: 5.519 on 9 and 63 DF,  p-value: 1.398e-05
```

```
Analysis of Variance Table

Model 1: `Invertebrate - ASPT` ~ cased + caseless + mayfly + stonefl
ies +
    olives + bwo + flatbodied + shrimp
Model 2: `Invertebrate - ASPT` ~ cased + caseless + mayfly + stonefl
ies +
    olives + bwo + flatbodied + shrimp + count_taxa
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     64 16.445
2     63 14.776  1    1.6689 7.1156 0.009702 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 22 BMWP advanced taxa model (taxa + taxa number)

```
lm(formula = `Invertebrate - BMWP` ~ cased + caseless + mayfly +
    stoneflies + olives + bwo + flatbodied + shrimp + count_taxa,
    data = combined)

Residuals:
    Min      1Q  Median      3Q     Max
-75.799 -18.663  -0.024  21.114  66.368

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   94.158     20.069   4.692  1.5e-05 ***
cased          1.883      6.238   0.302 0.763748
caseless     -10.779      6.568  -1.641 0.105741
mayfly       -10.094      6.438  -1.568 0.121907
stoneflies   -16.185      6.093  -2.656 0.009998 **
olives        -7.823      4.916  -1.591 0.116587
bwo           -8.245      6.093  -1.353 0.180829
flatbodied     2.639      4.838   0.545 0.587350
shrimp        -3.040      4.465  -0.681 0.498455
count_taxa    18.010      5.078   3.547 0.000742 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.31 on 63 degrees of freedom
Multiple R-squared:  0.2925,   Adjusted R-squared:  0.1915
F-statistic: 2.895 on 9 and 63 DF,  p-value: 0.00628
```

```
Analysis of Variance Table

Model 1: `Invertebrate - BMWP` ~ cased + caseless + mayfly + stonefl
ies +
    olives + bwo + flatbodied + shrimp
Model 2: `Invertebrate - BMWP` ~ cased + caseless + mayfly + stonefl
ies +
    olives + bwo + flatbodied + shrimp + count_taxa
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    128 151021
2    127 135835  1     15186 14.198 0.0002507 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis of influence of single taxa suggested that for high number of caseless caddisflies and shrimps, both invertebrate scores reported a slightly lower value, with a more pronounced effect in the ASPT score. Other taxa seemed to have a slight positive effect on the scores.

Figure 30 ARMI Taxa and ASPT scores plots

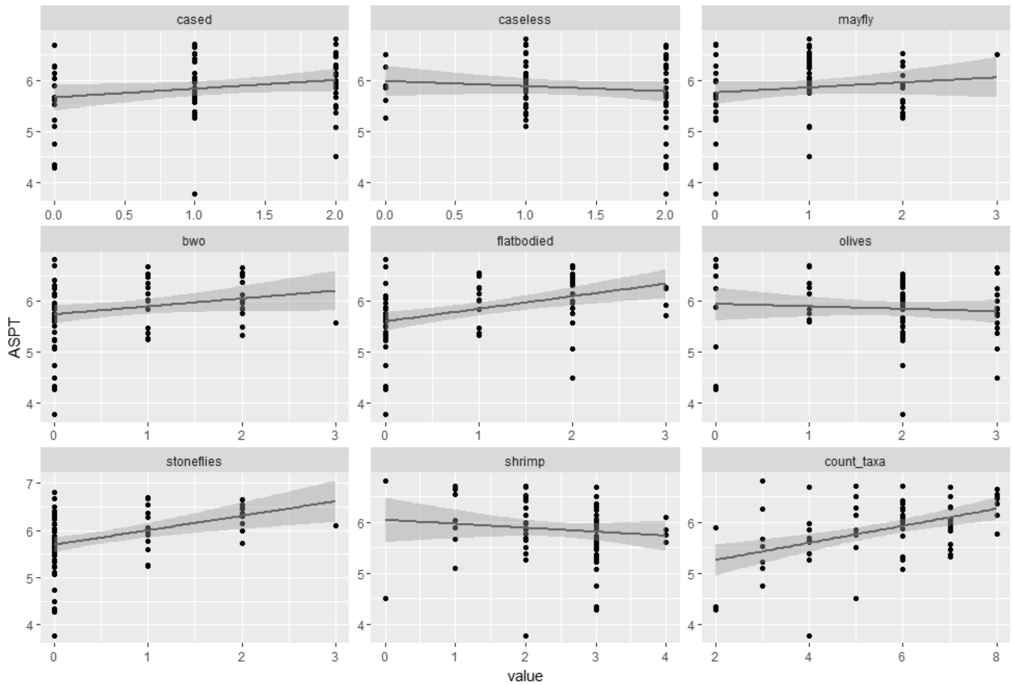
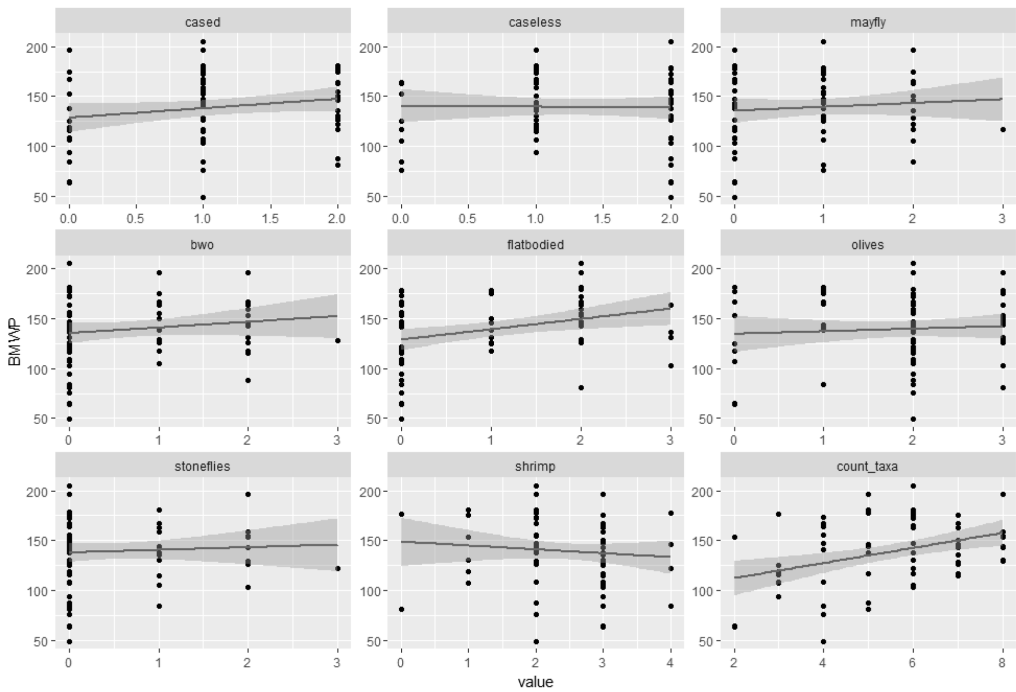


Figure 31 ARMI Taxa and BMWP scores plots



Prediction Power

Prediction of ASPT and BMWP categories using all the models developed showed the forecast was sensibly better for models developed using single taxa scores, with introduction of count taxa predictor that further improved the forecast.

Table 23 Outcome of prediction for each MLR model

Predictors	Variable to predict	Non-Correct prediction	Correct prediction	Success rate
ARMI score	ASPT	32	41	56%
Taxa scores	ASPT	25	48	65%
Taxa scores + count taxa	ASPT	21	52	67%
ARMI score	BMWP	42	31	42%
Taxa score	BMWP	38	35	48%
Taxa scores + count taxa	BMWP	29	44	60%

ASPT prediction was consistently better than the BMWP one, with a success rate ranging from 56% for the simple model to 67% for the one with all predictors. The BMWP models were correctly predicting the BMWP category between 42% and 60% of the time.

4.7.3 Final model and interactions

Predictors with p-value > 0.5, where removed, so that mayflies, bwo, shrimps, stoneflies, caseless and cased caddis were dropped as explanatory variables.

Final minimum model showed an adjusted R-squared value of 0.36, with all predictors significant at .95 level. Coefficients from the model were used to predict the ASPT scores and obtain river quality categories. The model successfully estimated the category 69% of the time.

Table 24 Minimum Model (ASPT)

```
lm(formula = `Invertebrate - ASPT` ~ olives + flatbodied + count_tax
a,
    data = combined)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.58558 -0.17445  0.06807  0.31829  1.05377 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.17268    0.21507  24.052  < 2e-16 ***
olives      -0.29758    0.07317  -4.067 0.000125 ***
flatbodied   0.14914    0.07235   2.062 0.043024 *  
count_taxa   0.19452    0.04754   4.092 0.000114 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4849 on 69 degrees of freedom
Multiple R-squared:  0.3861,	Adjusted R-squared:  0.3594 
F-statistic: 14.47 on 3 and 69 DF,  p-value: 2.081e-07
```

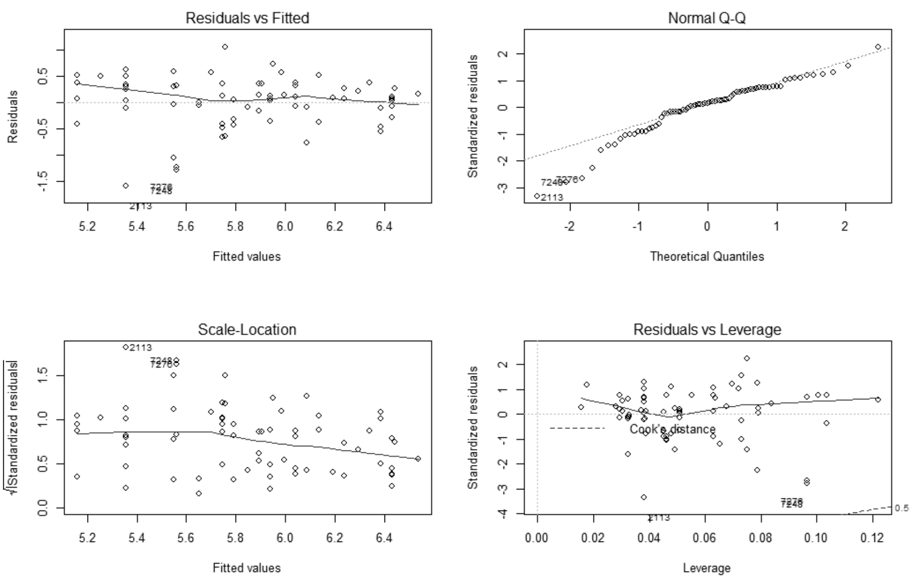
Model Coefficients

```
(Intercept)      olives  flatbodied  count_taxa 
  5.1726783  -0.2975837   0.1491415   0.1945170 
```

Analysis of multicollinearity reported low VIF values, confirming that none was present among the predictors.

```
   olives   flatbodied count_taxa 
  1.380826   1.669317   1.942215 
```

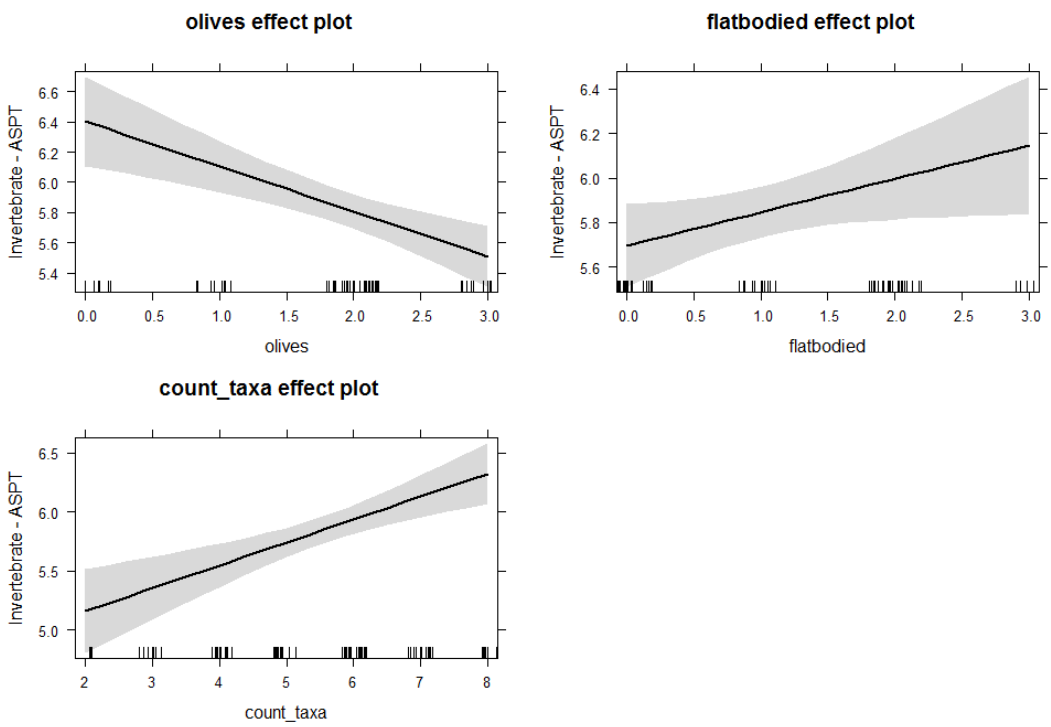
Figure 32 Final Model Residuals' Plots



Analysis of residuals suggested that no heteroskedasticity is present; however, for low ASPT score the model was underpredicting, while really low values were not captured at all by the model.

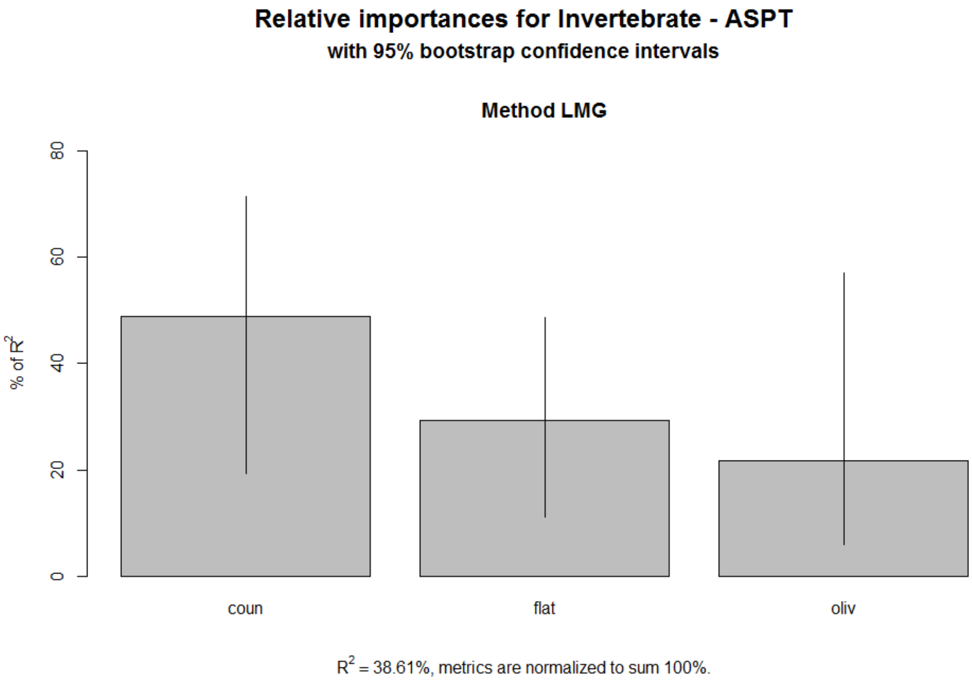
The model further indicated that higher number of olives would decrease the ASPT score, while flatbodied number and number of taxa were positively correlated to an increased score. Error bands around the regression lines were consistently larger around extreme values, with much of the variance still included in the intercept, which was highly significant. Introduction of interaction and quadratic terms was tested, however model performance remained the same.

Figure 33 Influence of Taxa score on ASPT score



Relative importance of predictors showed that the majority of variance explained by the model was due the number of taxa predictor, with the other variables having less impact.

Figure 34 Relative importance of predictors for ASPT score



Model's error and Cross-Validation

The plot of the predicted ASPT versus the actual ASPT score, displays the spread of the prediction, confirming that the model is unable to forecast really low ASPT values. Figure 36 shows the density plot of the error % for each prediction, with the model mostly underpredicting by a small percentage but with an evident right tail where the error percentage is high. This tail included observations were the model overpredicted low ASPT value that corresponded to poor and moderate river quality categories.

Figure 35 Predicted ASPT versus Actual ASPT score

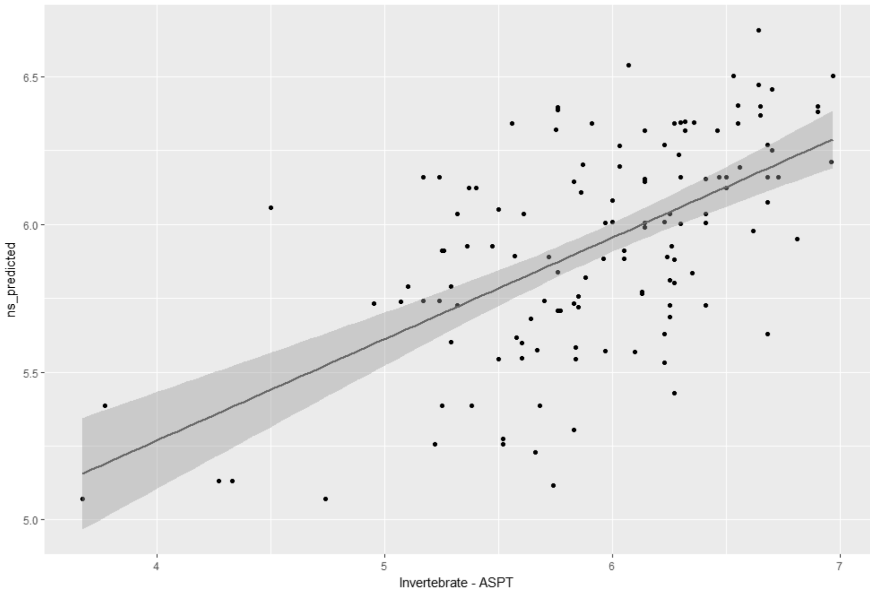
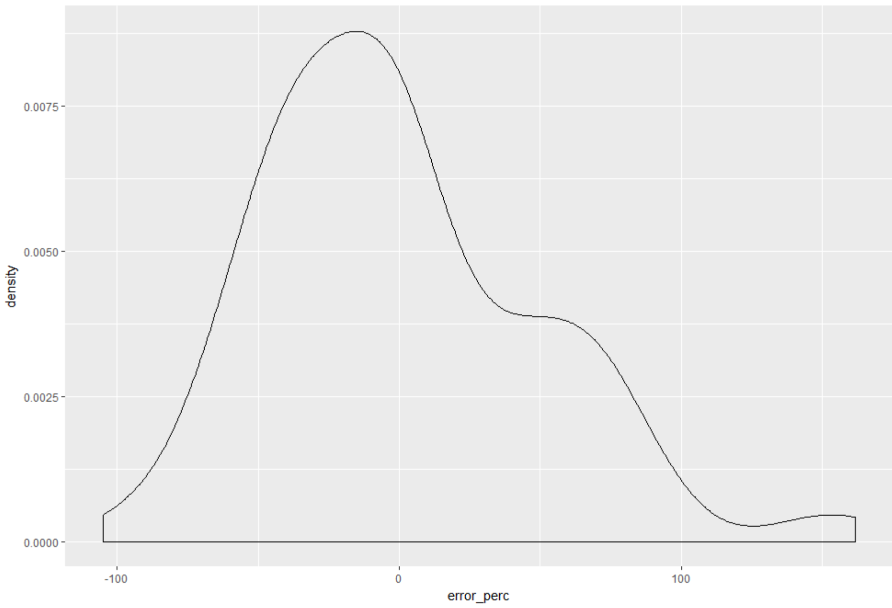


Figure 36 Density plot of model error percentage



Finally, cross validation with 5 and 3 folds reported a mean squared error of 0.242 and 0.247 respectively. Error percentage remained within 2-3% of difference among the different training sets and consistent with the error of the overall model.

Figure 37 Cross-validation Plot - 5 folds

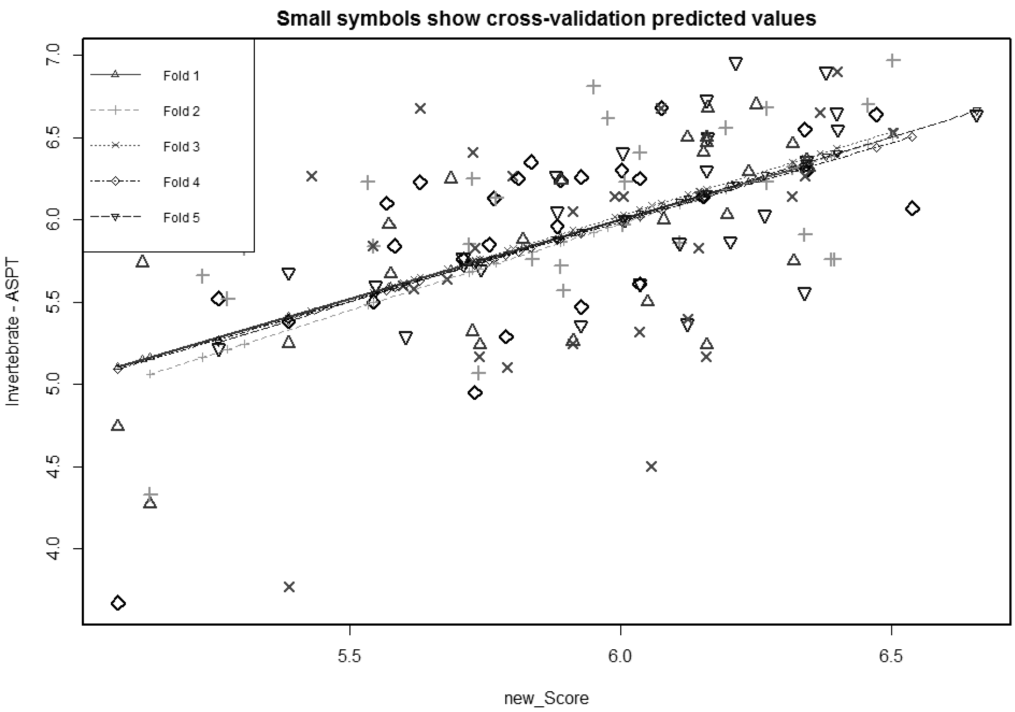
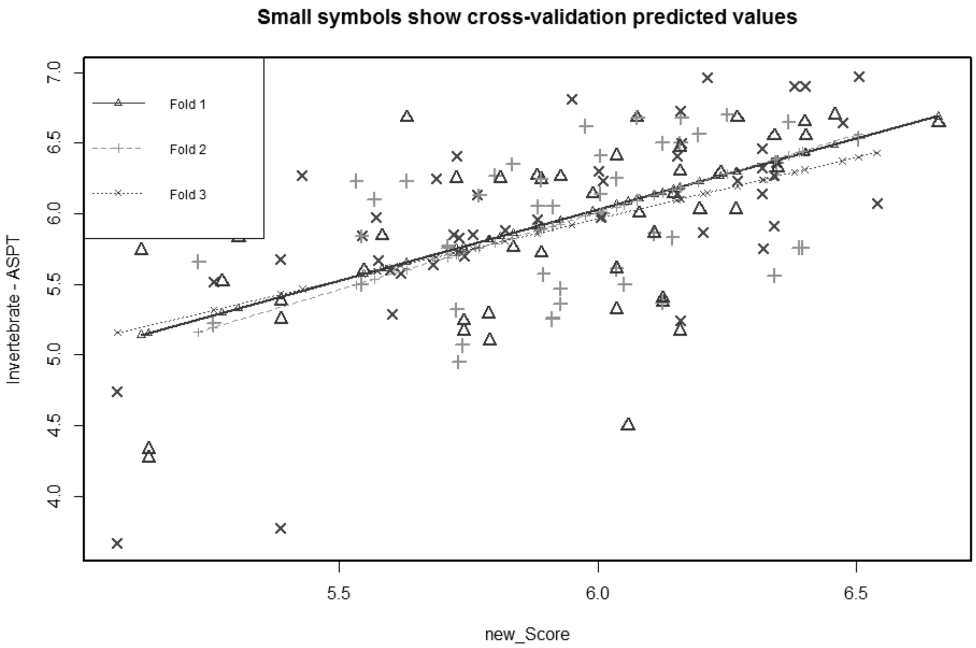


Figure 38 Cross-validation plot - 3 folds



5 Discussion

5.1 Theoretical validation of the ARMI method

Much of the literature confirmed the applicability of biological indices and citizen monitoring in assessing the quality of riverine ecosystems. Biological indices have been amply employed in evaluating river quality since their development. Publication of the WFD has then increased the creation of different indices to judge several aspects of the riverine ecosystem.

As for citizen science, the literature highlighted that the feasibility of different methods is direct consequence of the methodological basis and ultimate goal(s) of the monitoring activities/networks. On a second level, the academic world concurs that much of the level of success of citizen science initiative depends on a correct interaction between the academic world and the amateurs, where proven methodologies feed into volunteer activities with simplified but effective procedures. If this interaction works, volunteer initiatives have been proven to be effective in many studies because they nurture the fundamental enthusiasm of citizens to protect their local environment. On a further level, volunteer work produces back data that can be at a coarser resolution due to sampling techniques but with a much more evident capability to cover more locations more often, thus increasing temporal and spatial resolution of the studies.

The study demonstrated that the Anglers' Riverfly Monitoring Initiative follows these points because it builds upon clear theoretical frameworks shaped by decades of work on the elaboration of functional indices in Europe and in UK.

In particular, the ARMI method has developed from the amply used BMWP technique both in the theoretical and practical framework. The 3-minute kick sampling technique is thoroughly employed in obtaining other freshwater indices all around Europe, due to that fact that it is easy to perform and effective in sampling species individuals. Further, the close interaction between the development of a functional method that could be simple enough to be employed by amateurs and EA professionals has constructed a selection of taxa that are abundantly employed all over Europe. Especially, it is amply demonstrated that the eight taxa selected are markedly sensible to organic pollution, hence able to express, up to a certain degree, the quality status of water bodies. Further, the fact that they are abundant throughout the whole year permit amateurs to become familiar with a low number of taxa, avoiding time-consuming identification processes and thus reducing the number of mistakes. Finally, the history of the development of the ARMI initiative has been built on intertwined communications between amateurs and professional so that the network is composed by a close collaboration of many stakeholders from both worlds. Therefore, data are fed into the professional world in an easy and effective way, where the importance of the volunteers' results is amply understood. The ARMI data collected by volunteers is therefore used as a valuable baseline on which a more detailed picture of the ecological quality of rivers is built.

In fact, although with a simplified method, ARMI volunteers have successfully identified pollution events in many cases, which have been subsequently assessed by the EA and other involved parties.

5.2 Spatial and temporal variability of ARMI monitoring

Results of monitoring frequency confirmed the ability of ARMI volunteers to pick up differences in the faunal composition of macroinvertebrates in different rivers throughout the year.

Distribution of individuals per taxa was adequately diverse between monitoring activities with most taxa occurring with relatively low number of individuals (mean between 4.8 and 21.2), with the exception of olives and shrimps, which showed a higher mean of sampled individuals (mean of 275.9 for shrimps and 99.9 for olives).

When looking at variance, it was evident that most of the average ARMI score by year were pretty much stable, with differences year-on-year that were within a 2-point spread on average. Nevertheless, analysing the intra-annual variance confirmed that monthly variance was sensibly higher, with some scores fluctuating well over 2-3 points difference. This was confirmed by the plot of single taxa score distribution by month, which showed an evident trend in the presence and occurrence of different taxa during the year. This was expected as communities' composition vary during the year with some taxa more predominant that others.

The statistical tests performed confirmed the statistically significant difference between samples recorded during warm and cold months. Nevertheless, due the not sufficient number of samples per each site or river over multiple months and years, it was not possible to check the presence of seasonality when analysis was run at a finer resolution (river or sites). Analysis of correlation amongst ARMI taxa resulted in low or moderate-low values confirming the variable nature of the communities' composition.

5.3 Concurrence of volunteers' and professionals' results

Considering the whole results from 2011, the ARMI monitoring activities occurred almost twice more often than the EA ones.

The frequency of monitoring events run by the ARMI volunteers per year was higher for 72% of the rivers, with an average increase in monitoring activity for 26 waterbodies of 142%. Considering all waterbodies, including the 10 where EA had run more sampling, the average increase for ARMI was still around 89%. This was expected and confirmed the advantages of citizen science in covering more locations more often.

Analysis of scores' distributions showed that ASPT and BMWP scores reported good and very good river quality for most of the samples, with right-skewed distributions. On the other hand, ARMI score were less skewed, with a bimodal distribution, seeming to pick up more of the variability, likely due to the higher number of samples obtained in each year.

Analysis of comparable results, with scores related when samples were recorded within a 14-days' timeframe, resulted in a big part of the data being omitted from the analysis. Comparable results were only 73; however, this selection was necessary in order to not compare samples obtained too far apart in time. Failing to do so would have introduced higher error in the model(s) due to the increased possibility that pollution or any other events that could alter the macrobenthic composition could have happened over the time between the two monitoring activities.

Analysis of this dataset showed that distributions were more different than the ones considering the whole data. In particular, ARMI score was a lot less skewed; further, its variance was wider than the other two scores, suggesting again that more variability was picked up. Statistical tests confirmed that distributions and their variance were significantly different amongst the three indices.

Nevertheless, analysis of correlation indicated that a moderate positive correlation existed between ARMI and the other scores. This confirmed the hypothesis that ARMI methodology was not too distant from the trend picked up by professionals.

5.4 MLR model development

MLR models using only ARMI score as predictor confirmed that much of the variance in ASPT and BMWP scores was not picked up. In fact, using only the score as predictor resulted in really low adjusted R-squared values of 0.06 for the ASPT and 0.02 for BMWP. Estimation of the right river quality category using these models reported a success rate of 56% for ASPT and 42% for BMWP.

The study discovered that using single taxa as predictors and introducing the number of taxa sampled in total during the activity seemed to explain a lot more variability of professionals' score. This was confirmed by running Anova(s), which reported statistical significant improvements in model efficacy with the new predictor.

In particular, using all taxa scores and the number of taxa sampled to build new models increased the adjusted R-square values to 0.36 for ASPT and 0.21 for BMWP. Concurrently, forecast success rate increased to 67% and 60% for ASPT and BMWP models respectively.

Relationship between single taxa score and ASPT and BMWP scores suggests that higher number of individuals contributed to higher ASPT and BMWP scores for most taxa. On the contrary, olives and shrimps were negatively correlated, with higher number corresponding to lower ASPT and BMWP scores. A minimum model was developed, using flatbodied, olives and number of taxa as predictors for ASPT. This final model resulted having the same prediction power as the one including all predictors, but with all variables being significant. The model was then validated with Cross-validation technique, which also confirmed the error percentage of the model.

Overall, all models resulted in R-squared values never higher than 0.4. Prediction success was never higher that 67% for the ASPT. These results seem to highlight the importance of setting a local "trigger level" following professional judgment and historic trend. Therefore, selection of the thresholds becomes the key that gives validity to the assessment of the river quality in the scoring system.

Nevertheless, the development of the MLR models also appears to indicate that the ARMI methodology is indeed able to assess a part of the variability of the indices employed by professionals.

5.5 Limitations and further studies

Much of the limitation in the study were due to low number of monitoring results comparable and the fact that ASPT and BMWP results report just the score and not the actual number and composition of taxa identified.

Selection of a not-too-wide time frame comported in a small sample on which it was possible to run the analysis; further, selection of the 14-days' threshold also introduced a bias. Subsequently, the possibility that the sample does not represent the population has increased greatly. This means that much of the prospect of improving the index is subjected to obtaining more synchronised data from both professionals and though the ARMI technique. The ideal situation would require repeated monitoring activities were samples are obtained at the same time in the same site/river.

In particular, further studies on the introduction of the number of taxa sampled as new variable could open the way to a new score system, were the ARMI methodology could work on two levels. A first one where the method stays more or less the same and condition are assessed on the local trigger level. A second one, where the score could be less dependent from the ARMI trigger and more comparable to the indices employed by the professionals.

6 Conclusion

As citizen science becomes more and more important as a support for the scientific community, it is important to promote initiatives that employ techniques that are scientifically valid and can improve the amount and quality of data available for analysis and encourage participation at the same time.

The study found that the Angler's Riverfly Monitoring Initiative is instituted on a solid background. Much of the framework is based on techniques employed by other common biological indices. Selection of taxa, identification methodology and assessment of the river quality are all derived from the BMWP index, which has been used extensively in UK for the last few decades. Further, the network is highly interconnected with the academic and professional world, which provides both support and feedback and, in exchange, obtains an increasing amount of valid data to analyse.

Samples collected by the Initiative since 2011 in the Severn and Thames River Basin District areas indicate that seasonality, variability and difference in macrobenthic communities are successfully picked up by the ARMI volunteers. Further, number of monitoring activities is a lot higher than the one provided by the EA, with an overall increase in the same area of 89%.

When comparing results obtained by the initiative and invertebrate scores produced by professionals, the study found that there is a moderate positive correlation between the ARMI score and the ASPT and BMWP ones; however much of the variability is not successfully identified. This is likely due to the fact that the ARMI methodology is a simplified version of the one employed by the EA. Subsequently, the use of the trigger level is stressed as paramount for a correct identification of local problems.

Nevertheless, the study found that using single taxa scores obtained by ARMI volunteer and introducing the number of taxa sampled as a further predictor explains more of the variability of the ASPT and BMWP scores. In particular, using all taxa scores and the number of taxa sampled to build new models increased the adjusted R-square values to 0.36 for ASPT and 0.21 for BMWP. Concurrently, forecast success rate increased to 67% and 60% for ASPT and BMWP models respectively. Low number of comparable samples limited the investigation of more powerful models. Therefore, the study suggests that the institution of a project where samples are collected concurrently by volunteer and professional with the different techniques could help the investigation. This would allow to examine the possibility of an improved index, which could be less reliant on the alarm trigger level.

References

Abbasi, T. & Abbasi, S.A. (2012) *Water Quality Indices*. Elsevier.

Aceves-Bueno, E., Adeleye, A.S., Bradley, D., Tyler Brandt, W., Callery, P., Feraud, M., Garner, K.L., Gentry, R., Huang, Y., McCullough, I., Pearlman, I., Sutherland, S.A., Wilkinson, W., Yang, Y., et al. (2015) Citizen Science as an Approach for Overcoming Insufficient Monitoring and Inadequate Stakeholder Buy-in in Adaptive Management: Criteria and Evidence. *Ecosystems*. 18 (3), pp. 493–506.

Armitage, P.D., Moss, D., Wright, J.F. & Furse, M.T. (1983) The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*. 17 (3), pp. 333–347.

Bailey, R., Norris, R. & Reynoldson, T. (2004) Bioassessment of Freshwater Ecosystems. In: *Bioassessment of Freshwater Ecosystems*. Boston, MA: Springer US. p. doi:10.1007/978-1-4419-8885-0_1.

BART (2015) *Rivers Trust, education, land and river management advice | Bristol Avon Rivers Trust*. Available from: http://www.bristolavonriverstrust.org/ [Accessed 21 February 2016].

Bartram, J. & Ballance, R. (1996) *Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes*. E & FN Spon.

Blossom, T. (2012) *Fishing for Data : Potential for Citizen Science to Conserve Freshwater Ecosystems Table of Contents*.

Boslaugh, S. (2007) *Secondary data sources for public health: a practical guide*.

Buytaert, W., Dewulf, A., De Bièvre, B., Clark, J. & Hannah, D.M. (2016) Citizen Science for Water Resources Management: Toward Polycentric Monitoring and Governance? *Journal of Water Resources Planning and Management*. pp. 1816002.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T.C., Bastiaensen, J., De BiÃ¨vre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D.M., Hergarten, C., Isaeva, A., et al. (2014) Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*. 2 (October), pp. 1–21.

Calow, P. & Petts, G.E. (1984) *Rivers Handbook: The Science and Management of River Environments. Volume 2: Hydrological and Ecological Principles*. Wiley.

Chapman, D. & Jackson, J. (1996) Biological Monitoring. In: *Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes*. p. pp. 35.

Chave, P.A. (2001) *The EU Water Framework Directive: An Introduction*. IWA Publishing.

Clarke, R, T. & Davy-Bowker, J. (2014) *UKTAG River assessment method: Benthic invertebrate fauna*. pp. 1–92.

Cohn, J.P. (2008) Citizen Science : Can Volunteers Do Real Research ? *BioScience*. 58 (3), pp. 192–197.

Conti, M.E. (2008) *Biological Monitoring: Theory & Applications, Bioindicators and Biomarkers for Environnemental Quality and Human Exposure Assesment*. WIT.

Cummins, K.W. (1974) Structure and Function of Stream Ecosystems. *Bioscience*. 24 (11), pp. 631–641.

Davis, W.S. & Simon, T.P. (1995) *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*.

Dunlap, R.E. & Michelson, W.M. 1940- (2002) *Handbook of environmental sociology*. Greenwood Press.

Environment Agency (2016) *DATA.GOV.UK - Datasets*. Available from: https://data.gov.uk/data [Accessed 21 February 2016].

Everard, M. (2008) Selection of taxa as indicators of river and freshwater wetland quality in the UK. *Aquatic Conservation: Marine and Freshwater Ecosystems*. 18 (6), pp. 1052–1061.

Everard, M., Fletcher, M.S., Powell, A. & Dobson, M.K. (2011) The Feasibility of

Developing Multi-Taxa Indicators for Landscape Scale Assessment of Freshwater Systems. *Freshwater Reviews*. 4 (1), pp. 1–19.

Di Fiore, D. & Fitch, B. (2016) The riverfly monitoring initiative: structured community data gathering informing statutory response. *Environmental Scientist*. (August), pp. 36–41.

Forman, R.T.T. (1995) *Land mosaics: the ecology of landscapes and regions*. Cambridge University Press.

Frake, A., Hayes, P. & Region, E.A.S.W. (2001) *Report on the millennium chalk streams fly trends study : a survey carried out in 2000 among 365 fly fishermen, fishery owners, club secretaries and river keepers*.

Ghani, W.M.H.W.A., Rawi, C.S.M., Hamid, S.A. & Al-Shami, S.A. (2016) Efficiency of Different Sampling Tools for Aquatic Macroinvertebrate Collections in Malaysian Streams. *Tropical life sciences research*. 27 (1), pp. 115–133.

Giller, P.S. & Malmqvist, B. (1998) *The Biology of Streams and Rivers*. Oxford University Press.

Hand, E. (2010) Citizen science: People power. *Nature News*. 466 (7307), pp. 685–687.

Hannigan, J.A. (2012) *Environmental sociology*. 2nd edition. Political Science.

Hawkes, H. (1997) Origin and development of the biological monitoring working party score system. *Water Research*. 32pp. 964–968.

Holling, C.S. (1978) *Adaptive environmental assessment and management*.

Karr, J.R. (1999) Defining and measuring river health. *Freshwater Biology*. 41 (2), pp. 221–234.

Korycińska, M. & Królak, E. (2006) The use of various biotic indices for evaluation of water quality in the lowland rivers of Poland (exemplified by the Liwiec River). *Polish Journal of Environmental Studies*. 15 (3), pp. 419–428.

Laws, E.A. (2000) *Aquatic pollution : an introductory text*. Wiley.

Moffett, E. & Neale, M. (2015) Volunteer and professional macroinvertebrate monitoring provide concordant assessments of stream health. *New Zealand Journal of Marine and Freshwater Research*. 8330 (May 2015), pp. 1–10.

Nerbonne, J.F. & Nelson, K.C. (2008) Volunteer macroinvertebrate monitoring: Tensions among group goals, data quality, and outcomes. *Environmental Management*. 42 (3), pp. 470–479.

Paisley, M.F., Trigg, D.J. & Walley, W.J. (2014) Revision of the biological monitoring working party (BMWP) score system: derivation of present-only and abundance-related scores from field data. *River Research and Applications*. 30 (7), pp. 887–904.

Peacock, B. (2008) *The Riverfly partnership*.

Priestley, S. (2015) *CBP 7246: Water Framework Directive : achieving good status of water bodies*. (July), pp. 1–25.

Rech, S., Macaya-Caquilpàn, V., Pantoja, J.F., Rivadeneira, M.M., Campodònico, C.K. & Thiel, M. (2015) Sampling of riverine litter with citizen scientists ??? findings and recommendations. *Environmental Monitoring and Assessment*. 187 (6), .

Rosenberg, D. & Resh, V.H. (1993) *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman & Hall.

Roy, H.E., Rorke, S.L., Beckmann, B., Booy, O., Botham, M.S., Brown, P.M.J., Harrower, C., Noble, D., Sewell, J. & Walker, K. (2015) The contribution of volunteer recorders to our understanding of biological invasions. *Biological Journal of the Linnean Society*. 115pp. 678–689.

Ruaro, R. & Gubiani, É.A. (2013) A scientometric assessment of 30 years of the Index of Biotic Integrity in aquatic ecosystems: Applications and main flaws. *Ecological Indicators*. 29pp. 105–110.

Scaglia, P. (2009) *Il monitoraggio biologico negli ambienti fluviali : applicazione di metodi tradizionali e metodi conformi alla Direttiva 2000 / 60 / CE basati sullo studio delle comunità dei macroinvertebrati* . Università degli studii di Pisa.

The Riverfly Partnership (2015) *The Riverfly Partnership - Press Release*.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. & Cushing, C.E. (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*. 37 (1), pp. 130–137.

Vartanian, T.P. (2010) *Secondary data analysis*. Oxford University Press.

Wallace, J.B. & Webster, J.R. (1996) The Role of Macroinvertebrates in Stream Ecosystem Function. *Annual Review of Entomology*. 41 (1), pp. 115–139.

Wright, J.F. (2000) An introduction to RIVPACS. In: *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques*. pp. pp. 1–24.

Wu, J. & Hobbs, R. (2007) Key Topics in Landscape Ecology Jianguo Wu & Richard J. Hobbs (eds.). *Landscape Ecology*. (1999), pp. 355–365.

Ziglio, G., Siligardi, M. & Flaim, G. (2006) *Biological Monitoring of Rivers: Applications and Perspectives*. John Wiley & Sons.