Glass half full or half empty?

Why 2009 Water Framework Directive classification results are over-optimistic about the state of rivers despite the One-Out, All-Out rule.

Rob Cunningham
April 2012

Summary

Although the One-Out, *All-Out* principle introduces the theoretical risk of pessimism bias in Water Framework Directive classification, in reality there are a host of factors which are likely to lead the 2009 results as being overly-optimistic, particularly for biology. Even if pessimism were to creep into results there are specific safeguards built into the Water Framework Directive that allow Member States to discount action where significant uncertainty exists. In summary:

- The Water Framework Directives classification system uses biological, chemical, and hydro-morphological "quality elements" to describe ecosystem health. The *One-Out, All-Out* rule embodies the precautionary principle in the face of uncertainty about how the complex web of interactions and inter-dependencies operate.

- Concerns about the possibility of pessimism in the application of the *One-Out, All-Out* rule are founded in statistical theory. However, a study undertaken on behalf of the Environment Agency found that this was cancelled out by misclassification of quality elements reported at *Good* or *High* status.

- Optimism bias is introduced in the 2009 results by to a number of factors that include;

    - Monitoring is limited, particularly for biology and there is an implicit assumption that the status of any unmonitored quality element is at least as good as those which are being monitored. Something that extrapolation of data from the surveillance sites shows to be unsafe.
    - A biological monitoring programme that does not reflect the scale of eutrophication risks identified.
    - Chemical standards which means that, in the absence of biological monitoring, there is a significant risk of a chemical "pass" while the true status of biology is a "fail".
    - The failure to include some significant pressures on the aquatic environment in the monitoring programme, most notably silt.

- The resulting optimism bias has been observed in the field and through data analysis.

- The theoretical risk of unnecessary and expensive investment / regulation due to pessimistic classification is wholly mitigated by the *weight of evidence* approach adopted by the Environment Agency. This requires classification to be corroborated across quality elements and/or additional investigations before any targeted action or expenditure is undertaken.

- The disproportionate cost / technical feasibility tests in the WFD provide the final safeguard against wasted investment.

- Ongoing investigations and the roll out of a more robust monitoring will inevitably involve reporting a significant number of "paper" deteriorations. Such are the inevitable consequence of redressing the optimism inherent in the 2009 classification results, a move that takes us closer to understanding the true status of our water.

Contents

1. Introduction

This paper looks back at how monitoring and classification was structured in the run up to the publication of the 2009 River Basin Management Plans. Since then a lot has happened, not least the Judicial Review by Angling Trust and WWF which was settled with the publication of a "Statement of Principles"[1] by Defra. This commits to an accelerated investigative monitoring programme to reduce uncertainty of classification and a catchment approach to planning which encourages third party participation.

The Environment Agency also plans to invest more money in WFD monitoring, creating a fixed network of physico-chemical monitoring sites and shorter cycle of biological monitoring so that no site is classified on data that is more than 3 years old. This should improve statistical confidence in classification results and trends over time. All of this means that we should hopefully enter the second round of River Basin Management Planning with a better understanding of the true status of water bodies and what pressures are driving failures.

However the *One-Out, All-Out* remains a cause celebre for a number of organisations who see it as being overly onerous and a source of pessimism which results in our waters being reported at a lower quality than they really are. This position was most recently aired in oral evidence given to the House of Lords inquiry into European Freshwater Policy[2] in 2011.

This paper is a response to those concerns, setting out the case for the *One Out All Out Rule*, exploring evidence for bias in the 2009 classification results and the factors contributing towards potential misclassification.

2. WFD Classification and the *One-Out, All-Out* rule

The Water Framework Directive is a hugely complex and wide ranging piece of legislation. However, at its heart lies the concept of protecting and enhancing aquatic ecosystems and it is the *ecosystem* that the classification system attempts to characterise. These ecosystems are made up of complex, interconnected and interdependent relationships between species and physical processes.

So, in describing Good status, the WFD requires Member States to assess key indicators of ecosystem health which, for rivers, includes fish, plant life and, invertebrates as well as the physical, chemical and hydrological conditions required to support them.

Many of the interactions and inter dependencies are poorly understood and may take years to express themselves e.g. macrophyte (plant) response to nutrient enrichment and subsequently invertebrate response to macrophytes. So it is no accident that the European legislature chose to adopt the one-out, all-out rule whereby overall classification is defined by the lowest observed

[1] http://assets.wwf.org.uk/downloads/defra_sop_2011.pdf
[2] Transcript of oral evidence to House of lords Inquiry into European Freshwater Policy www.parliament.uk/documents/lords-committees/eu-sub-com-d/EUFreshwaterPolicy/Freshwater-Policy-Evidence.pdf

individual quality element. Rather, as *Recital 11* of the WFD makes clear, the Directive adopts the *precautionary principle* a central tenet of EU environmental law[3] that is enshrined in the 1992 Rio Declaration[4].

> *As set out in Article 174 of the Treaty, the Community policy on the environment is to contribute to pursuit of the objectives of preserving, protecting and improving the quality of the environment, in prudent and rational utilisation of natural resources, and to be based on the precautionary principle and on the principles that preventive action should be taken,.....*
>
> **Recital 11 of the Water Framework Directive**

The precautionary principle is also adopted for a number of "Specific Pollutants" because of their inherent toxicity, persistence etc. The standards set for these chemicals are more rigorous than those required to support Good Biological Status reflecting the fact that they pose a significant long-term risk to aquatic ecosystems and human health. Furthermore, a precautionary approach makes economic sense in many circumstances as pollution prevention is often significantly more cost effective than remediation, pollution of groundwater being a prime example.

3. The theory of pessimism in the *One-Out, All-Out* Rule

A number of organisations across the EU have raised concerns that the *one-out, all-out* rule employed by the Water Framework Directive (WFD) as they fear it introduces a systematic risk of erroneous downgrading of water body classification. In essence the argument is:

- ⇓ Under WFD rules, the status of waterbodies are classified on the basis of a number of biological and chemical quality elements, with the overall result based on the lowest score.
- ⇓ Every piece of data derived from a monitoring programmes has error associated with it.
  - ? If that error results in an individual quality element being reported as a higher quality than reality and so higher than other monitored elements, the overall reported status will be unaffected because the poorest defines classification.
  - ? If that error results in an individual element being reported as a lower quality than reality, and so lower than the other quality elements, the overall reported status will be dragged down by that error.
- ⇒ With over thirty quality elements feeding into classification, there is significant risk of over-pessimistic water body classification which, in turn, will drive unnecessary spending by government and industry.

[3] EU COM (2000)1 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2000:0001:FIN:EN:PDF
[4] Principle 15 of the Rio Declaration 1992

This paper does not question this statistical phenomenon but it is important to note it is not the only one at play. Indeed a statistical analysis of 2009 classification results commissioned by the Environment Agency[5] concluded that

> *.....potential pessimistic bias in the national WFD classification results caused by the 1OAO* [One Out- All Out] ***rule is cancelled out by an even greater optimistic bias*** *that arises from not taking into account the possible mis-classification of QEs* [Quality Elements] *reported as being of Good and High status.*
>
> [emphasis added]

This paper goes further to explain how the approach to monitoring and classification adopted by Defra and the Environment Agency introduces optimism in 2009 classification results and what the consequences are.

4. The story so far

Concerns about the 'One Out All Out' are not new. The issue was widely debated in the run up to the publication of the 2009 River Basin Management Plans because the headline percentage of water bodies being reported as Good Ecological Status was significantly lower than the figures for Biological Status. This was highlighted in the draft River Basin Management Plans, the inference being that biology was satisfactory but that headline figure were being dragged down by the *One-Out*, *All-Out* rule because chemical failures, particularly for phosphate, were not biologically significant.

This perception was criticised by the RSPB and others when it was pointed out that the apparent mismatch was simply an artefact of the monitoring regime. Biological sampling was simply happening in far fewer places than chemical monitoring and therefore fewer failures were being detected. To illustrate this, in the case of the Anglian region, Environment Agency monitored phosphate in 573 water bodies and macrophytes and/or diatoms (the biological elements most sensitive to eutrophication) in just 92[6].

The Environment Agency accepted this and the final River Basin Management Plans contain statements like that reproduced from the Anglian plan below which acknowledge that gaps in monitoring are leading to optimistic classification of biological status. .

[5] Davey A. Bewes, V.(2011) *Statistical review of the one-out, all out rule for classifying water bodies* A report to the Environment Agency by WRC

[6] Page 35 of Anglian River Basin Management Plan http://publications.environment-agency.gov.uk/PDF/GEAN0910BSPM-E-E.pdf

..... from the chemical monitoring the Environment Agency ***is now clear that there is a link between high levels of phosphate in surface waters and biological failures in the main river type*** *(lowland alkaline rivers). The assessment of reasons for failure that we have started to undertake shows that across England and Wales 22 per cent of river water bodies are failing to achieve good status/potential because of excessive levels of phosphate.*

*In this river basin district phosphate results show that i****t is likely that the percentage of water bodies at good or better biological status will reduce from 33 to 27 per cent when additional water bodies are assessed for diatoms and/or macrophytes****. This same analysis points to discharges from sewage treatment works and releases from agriculture being responsible for the majority of this. Rather than wait for the results of more biological assessments, we need to ensure corrective action is started in the first plan cycle.*[7]

[emphasis added]

5. Optimism bias

The flip side to the potential for pessimism bias in the *One-Out, All-Out* rule is that, where quality elements are not monitored, they play no part in classification and so are implicitly assumed to be at least as good as the worst observed result for that water body.

In an ideal world, where all quality elements are monitored or where the standards for those quality elements reflect overall ecological status, this shouldn't matter. However, this is not the case. Instead, there is a real risk that gaps in monitoring and uncertainty around the setting of chemical standards can lead to optimistic misclassification.

The scale of that risk is related to (i) how comprehensive the monitoring network is, (ii) how well monitoring is targeted on those bodies at risk of failing, (iii) how well the most widely-monitored chemical quality elements describe ecological status, and (iv) whether all significant threats are captured by the monitoring programme.

5.1. How comprehensive is WFD monitoring?

Concerns about pessimism in the One-Out, All-Out rule are often accompanied by claims about the large number of quality elements used to classify water body status. This is perhaps not surprising given that the Environment Agency's website states:

.. [the WFD] *is based on a far wider range of assessments than GQA classification. It reports on over 30 measures, grouped into ecological status (including biology and 'elements' such as phosphorus and pH) and chemical status ('priority substances')*[8]

[7] Page 17 of the Anglian River Basin Management Plan
[8] http://www.environment-agency.gov.uk/research/planning/34383.aspx March 2012

What the Environment Agency's website fails to highlight is that, with the exception of around 600 "surveillance sites", very few waterbodies are classified on anything like that number of quality elements. This is because the WFD allows Member States to adopt a risk-based approach to targeting monitoring effort.

In England this has meant that, of the water bodies classified in the 2009 River Basin Management Plans, only 209 were monitored for 4 biological quality elements, 411 for 3, 1662 for 2 and 1988 for 1, while 2835 were classified with no biological data at all.[9] This means that in nearly 40% (39.9%) of waterbodies it is assumed that all biological elements are reaching the status of the poorest chemical element being monitored.

These figures are further put into context by information presented in the South West River Basin Management Plans (Table 1) which shows levels of diatom and macrophyte monitoring lagging far behind the "traditional" chemical and invertebrate monitoring adopted as part of the General Quality Assessment (GQA) over the previous twenty years or more. This is indicative of the approach adopted across England and Wales.

Table 1 Quality elements monitored for 2009 classification results, data derived from South West River Basin Management Plan[10]

Biological Quality Elements	Number monitored	% Total	Chemical Quality Elements	Number monitored	% Total
Diatoms	86	9%	Dissolved Oxygen	596	64%
Macrophytes	78	8%	Ammonia	596	62%
Invertebrates	427	46%	pH	583	64%
Fish	385	41%	Phosphate	596	64%
			Temp	596	64%
			Specific Pollutants	597	64%
			Priority Hazardous Substances	38	4%

[9] Analysis undertaken by RSPB using 2009 classification data supplied by Environment Agency

[10] Number of river, canal and water transfer water bodies (page 10) and monitoring figures (page 33) from South West River Basin Management Plan http://publications.environment-agency.gov.uk/PDF/GESW0910BSTP-E-E.pdf

5.2. Is monitoring effort well targeted?

The targeting of monitoring effort is entirely consistent with the provisions of the WFD and should maintain robust outcomes, assuming enough monitoring is put in place to cover the areas at risk of failing.

This can be tested by simply comparing the scale of risk with the number of water bodies monitored for quality elements that are likely to be impacted. For example Page 11 of Annex G of the SW plan states that:

> *"It is estimated that 59% of the total length of river water bodies are at risk or probably at risk from diffuse phosphorus from agricultural pollution*."

Figures from Table 1 above show that the level of phosphorous monitoring is comparable to the risk presented. However only a fraction of water bodies are monitored for the macrophyte and / or diatom biological quality elements that are most sensitive to eutrophication driven by phosphate pollution. As a consequence, for the overall classification to be accurate, we have to rely on phosphate standards being set at a level that accurately reflects biological response, something examined in see Section 5.3.

5.3. Do chemical standards reflect biological status?

The science behind setting ecologically meaningful physico-chemical standards is complex and fraught with uncertainty because the biological response to the physico-chemical environment is often dependent on a range of incompletely understood factors. As a result, scientists tasked with setting standards face the challenge of turning a broad scatter of observed data into a simple set of classification standards for High, Good, Medium, Poor and Bad Status.

For phosphate, a widespread pollutant that drives eutrophication in freshwater systems, UKTAG set a standard for calcareous lowland rivers of 120μg/l P[11]. This gives a high degree of certainty that biology will be impacted before the P standard is breached and so reduces the chance of pessimistic miss-classification in the absence of corroborating diatom or macrophyte data.

However, there is considerable uncertainty in the relationship between phosphate concentration and diatom response and at this concentration a significant number of water bodies will pass the phosphate standard, but fail for diatoms if they were monitored. Indeed research undertaken for Defra[12] found that, by the time phosphate reaches 100 μg/l in streams, the biological response is already complete. In other words, phosphate levels that affect biology are significantly lower than the standard adopted by UKTAG.

This optimism in the phosphate results were demonstrated by RSPB's analysis of the limited number of Surveillance Monitoring sites where all or most quality elements were monitored.

[11] ww.wfduk.org/sites/default/files/Media/Environmental%20standards/Environmental%20standards%20phase%201_Finalv2_010408.pdf

[12] ADAS (2008) *Linking agricultural land use and practices with a high risk of phosphorus loss to chemical and ecological impacts in rivers* report to DEFRA

The results are reproduced in Table 3 and show that where water bodies were monitored for phosphorus, diatoms and macrophytes, a significant number fail the biological standard but pass for phosphorous. These findings were corroborated by a more comprehensive study undertaken by Reading University for the Environment Agency[13].

A significant mis-match can also be observed between headline *biological status* and *physico-chemical status* at surveillance sites[14] with a significant number of water bodies having biological classification worse than physico-chemical status would suggest (Table 2).

Table 2 Analysis of agreement and mismatch between biological and physic-chemcial classification for surveillance sites (2009 data)

	Number	**% Total**
Number of water bodies where physico - chemical classification is Moderate or Worse but biology Good or Better	37	9.7%
Number of water bodies where physico - chemical classification is Good or Better but biology Moderate or Worse	**150**	**39.3%**
Number of water bodies where biological and physico chemical classification match	195	51.0%

It should be noted that this simple analysis was undertaken some years ago with data available at the time. However the findings should come as no surprise to freshwater ecologists. When UKTAG consulted on standards, a number of external experts (including Natural England) suggested that rather than being too tight, the standards and associated compliance regimes for some river types would allow significant deterioration before triggering a GES failure.

[13] Barahona, C. (2009) *Analysis of outcomes for tests of Phosphorous EQS, Diatoms EQR and Macrophytes EQR*, a report to the Environment Agency

[14] *Do results from the Water Framework Directive "Surveillance Sites" support concerns about the one out all out rule?* RSPB publication, November 2009

Table 3 Analysis of agreement and mismatches between phosphate and biological data

Adapted from RSPB briefing *Does WFD monitoring show a biological response to elevated phosphate levels in rivers?*

	All biology		Diatoms only		Macrophytes only	
	WBs	Length (km)	WBs	Length (km)	WBs	Length (km)
Fail both phosphorus **and** biology	252	4836	240	4615	38	888
Fail phosphorus but good for biology	68	1363	31	629	40	832
Fail biology but good for phosphorus	**189**	**3527**	**182**	**3415**	**18**	**380**

- Total diatom sampling WBs = 607 (including grouped water bodies so actual number monitored is lower)
- The above figures do not take into account confidence of classification
- The above figures are based on the first year of the macrophyte and diatom monitoring programme. It's a rolling programme so it only represent a third of the data that available by the end of 2009
- Macrophytes are only assessed in surveillance water bodies. Diatoms are assessed in water bodies at risk of nutrient enrichment
- Some water bodies have diatoms and macrophyte assessments, so diatom only and macrophyte only results add up to more than the total.

5.4. Are all significant threats captured by the monitoring programme?

There are always likely to be some water bodies where site specific factors are impacting ecology and so aren't captured by the physico-chemical regime. As a result, unless the sensitive biological quality element is monitored, these impacts will go unnoticed and so classification will be erroneously high.

It is, of course, impossible to quantify just how many 'unknown unknowns' exist. However there are likely to be systematic problems, for example alien species are known to be a major issue in some water bodies but their presence can only cause a water body to be reclassified from high to good.

But in the run up to the publication of the 2009 River Basin Management Plans the most controversial omission NGOS was the lack of suspended sediment monitoring in WFD classification. Sediment, primarily from agriculture, is a serious pollutant and causes acute problems in many fisheries, blanketing spawning beds and suffocating eggs.

This should be reflected in findings from biological monitoring but, as figures from South West illustrate, despite a risk assessment which found 47% of water bodies are at risk of failing due to sediment inputs, only 41% were monitored for fish, the quality element most at risk. The apparent gap between risk and monitoring effort is likely to be larger given the uncertainty associated with risk assessment methodologies.

6. The consequences of Misclassification

The question of whether WFD classification, and the approach adopted in England and Wales, is inherently optimistic or pessimistic might be seen as academic to all but a few devotees. The vitally important consideration is what consequences stem from those errors?

6.1. The consequences of pessimistic misclassification

It is clear that those who express concerns about the potential pessimism in the *One-Out, All-Out* rule are worried that monitoring error will drive up costs for taxpayers, water customers, industry and farmers without delivering real improvements. This seems a reasonable concern, not even the most ardent environmental campaigner is going to argue that we should be spending time and effort fixing problems that don't exist. However, as outlined above, the potential for misclassification is mitigated in a number of ways.

But if we assume some pessimistic classification errors will occur, it is important to note that regulation and investment decisions are not made on classification results alone. Instead the Environment Agency and Defra have adopted a stringent *weight of evidence* approach to taking

action[15]. In essence this means attributing a measure of *certainty* to each classification result as well as the *reason for failure* attributed to it. This might mean, for example, only attributing high certainty to classification of less than Good Status where it is corroborated by biological evidence, and undertaking an investigation in order to ascertain why that failure has occurred.

Indeed the sheer weight of evidence required by the Environment Agency to trigger action was a key part of the legal challenge made by WWF and the Angling Trust because the monitoring put in place in the run up to production of the 2009 River Basin Management Plans could not meet the standards of evidence being set. This was reflected in an a Environment Agency paper on combining quality elements[16] that included the statement

> *With the relatively modest monitoring programmes generally envisaged, there may be cases where it is not actually possible to demonstrate with sufficiently high confidence that a site was NotGood, however poor the true quality was.*

In settling the challenge Defra issued a statement of principals which includes commitments to use the investigations programme to resolve issues of uncertainty acting as a barrier to inaction. However, this came too late to influence the 2009 classification results.

Even if this, arguably overly precautious barrier to action were relaxed to allow a more nuanced balance between available evidence and investment to be struck in future planning cycles, the risk of wasted investment is ultimately mitigated by derogations which allow action to be discounted or deferred on grounds of *disproportionate cost* and *technical feasibility*.

The widespread application of these derogations in the 2009 River Basin Management Plans was also raised in the grounds for Judicial Review, but used appropriately they offer a legitimate safeguard against wasted or disproportionate investment.

6.2. The consequences of optimistic misclassification

The 2009 classification results have a legal significance because it is these which have been reported to Europe and against which improvements and deterioration will be judged. If there is significant optimism bias in these results, this should be revealed by the expansion of monitoring and investigation programmes ahead of the next round of River Basin Management Plans to be published in 2015. This means the UK may be faced with the problem of formally reporting widespread deterioration in headline status, even if nothing has changed on the ground.

While the Environment Agency have stated such “deterioration should not lead to infraction proceedings in Europe” it is unclear how such changes will be untangled from other factors that

[15] For example *Weight of Evidence rules for combining macrophyte, diatom and phosphate in river classification* page 91, Annex D of South East River Basin Management Plan http://publications.environment-agency.gov.uk/PDF/GESO0910BSTF-E-E.pdf

[16] *Combining Multiple Quality Elements and Defining Spatial Rules for WFD Classification,* Environment Agency science paper (2007)

may have resulted in real change or what view the Commission will take of any systematic failures in monitoring regime/standard setting that are likely to be revealed.

And, of course the challenge of reporting change to the wider public remains. For example, the Anglian River Basin District water customers will have paid for significant investment in phosphate stripping at sewage treatment works between 2010 and 2015 but reported failures linked to phosphate are likely to rise (See Section 4).

7. Conclusions

While this paper does not challenge the fact that concerns about the theoretical pessimism in the application of the *One-Out, All-Out* rule are well-founded, this is unlikely to be a significant issue in terms of headline of classification or potential wasted investment, not-least because there are so many confounding factors which could lead to over-optimism in results.

In the unlikely event that pessimism bias creeps into headline classification, the impact on decisions about costly investment or regulation will be mitigated by application of the *weight of evidence* approach. If this weight of evidence approach were modified to allow more nuanced approach to balancing available evidence against scale of action, the disproportionate cost test remains as a firm safeguard against wasted investment.

Evidence presented by the Environment Agency itself suggests the biggest risk is that biological failure has been underplayed in the 2009 classification results. This will be revealed as of "paper" deterioration in status as improved monitoring is rolled out across water bodies.

This could be seen as confirming problems with the *One-Out, All-Out* approach to classification as it will come about because more monitoring is undertaken. However, the evidence presented here and elsewhere points to such changes being the inevitable consequence of redressing the optimism inherent in the 2009 classification results, a move that takes us closer to understanding the true status of our water environment.