**Protecting river quality – getting the sums right**

Tony Warn (8th January 2024)

1. I have worked with brilliant colleagues for more than 40 years on how to calculate and justify what must be done to protect or improve the quality of rivers. We have covered, for example, more than £30 billion of capital investment by the water industry.
2. My software (RQP and SIMCAT) is still used for today’s plans and decisions. It is enhanced when: we face new types of pollution; if we learn new threats or new ways of controlling threats; and if we have access to new types of useful data.
3. A sound calculation will always calculate the confidence that its results are correct and demonstrate that they have not been ruined by poor or limited data. We can then choose a good balance between:

### targeted actions for individual sites on a river;

### national actions such as bans on chemicals, imposing fixed standards on all discharges, and setting national or regional controls on the use of land.

1. How do we justify a need to persuade or insist that farmers, industry and others take fair and needed action? How do we select the best mix of actions? This set of notes covers:

### the type of standards and targets we must use;

### how we identify the current or potential failures of standards (the red on a map);

### how pollution mixes into a river and we can calculate the controls needed to meet the targets for the downstream river;

### doing this for an entire river and, as necessary, summing results across all the rivers in our country, and covering all kinds of pollution.

1. We always need to assess errors. This is the confidence that a standard has been failed, and that action is needed and would not be wasted. We can then control the risk of bad decisions. We can also avoid over-elaborate and lengthy calculations in cases where we demonstrate that improved precision is trivial compared with the errors imposed by things like limited sampling.
2. Good calculations make us effective when we need to:

### examine the importance of issues and establish whether we need to act;

### devise, influence and implement laws and policies;

### set up and maintain helpful monitoring;

### define standards for rivers and set up national targets for compliance;

* choose actions that best meet local standards and national targets;
* set up controls on pollution, and a fair basis for legal action if these are failed;
* review progress over the years – and update our plans.

## Standards

1. Over the years, we may have monitored thousands of sites by taking regular samples. If we look, say, at sets of three years of results for a pollutant at the monitoring points on all of our rivers, and plot a histogram for each site, we usually see that the shapes of the histograms are similar across nearly all sites.
2. This means that if a site has a mean of 10, we have a good idea of the ranges of values that lie above and below 10. The range might, say, exceed a value of 30 for 10-20 days in a year. Such days might be consistent in being more commonly shared in summer or winter. They might tend to take place either in dry or wet weather, and they might be more be likely to occur in sequences than not.
3. This persistence in the shapes and structure of the histograms helps us work out:

### what needs to be done;

### the confidence that decisions are correct and will succeed;

### sites or issues that require more elaborate monitoring or calculations (and perhaps a more complicated standard).

1. We sometimes talk as if a standard is a concentration that must never be exceeded – an absolute limit. But standards deal not only with exceeding a concentration but with how often this can happen, over what period of time, and in how many of a set number of those periods of time.
2. It is easy to declare that a concentration should NEVER be exceeded. But there is a big difference between risks of 1 in a hundred, 1 in a thousand, or 1 in a million. With an absolute limit, each decision may sit at a different and unknown point on this range. This means that a different “standard” is used for every location. Bad decisions will result.
3. With absolute limits, assessments of failure will be biased by sampling rates – the more samples the tighter the standard. Action will be wasted on sites of low risk, and real problems will escape detection. We will get corrupt comparisons of nations and regions – penalising those that take more samples.
4. If we insist on absolute limits, we cannot work out correctly and honestly the conditions required to protect water quality using, say, legal permits for discharges to rivers. Polluters would be justified in objecting to such permits.
5. A minimum requirement is that a standard must be a summary statistic like an annual mean, or an annual percentile such as the 95 or 99-percentile. These standards embody setting particular risks of exceeding specified concentrations.
6. This use of a summary statistic is supported by the observation that a summary statistic for a site is linked to an underlying shape observed for that pollutant at nearly all sites. Achieving an annual mean gives a good idea of the range of values that contribute to it, and an idea of the frequency at which high concentrations might be expected.
7. This standard is the simplest that allows a correct calculation of action to avoid failures. No hidden risks to rivers, nor unfair costs to industry, need stem from our calculations. And the impacts of limited data can be quantified correctly and used to help us take sensible decisions; and to decide where more information is needed.
8. Where there is monitoring, we can assemble data on all sites of excellent quality in terms of its biology, and plot a histogram of, say, the annual mean for a pollutant for all these sites. The results might identify, say, an annual mean for which 90% of sites have excellent biology. This annual mean is a potential standard. The data for these sites, and parallel data for sites with poorer biology, allows the calculation of the risk of not getting excellent biology, even if the standard is met.
9. 40 years ago, Alan Barnden compared estimates of annual 90-percentiles of total ammonia with the results of angling competitions across the whole country. This showed the 90-percentile value that looked to be a good standard for getting good fisheries.
10. This sort of standard, the mean or percentile, is a summary statistic. It will embrace effects like toxicity but it may also be affected by other features. These may include a link to types of land use, or a correlation with other pollutants. There may also be supporting information from toxicology.
11. Some sites may face risks of rare events that have huge concentrations. We might use an annual mean standard to cover such risks by dragging the annual mean standard to a lower value where its underlying shape suggests high values occur at an acceptably low frequency. But we shall still need a background regime for managing river quality that deals with the risks of freak events or illegal activities. Or we may need a complex model for taking decisions.

1. An annual mean or percentile leads indirectly to settings for the features required for what used to be called an Ideal Standard. This includes:

### what is the concentration?

### how often can the concentration be exceeded – 5% of the time?

### over what period of time – 1 year?

### in how many of these periods of time – 1 year in 10?

### what confidence of failure will lead to particular actions? - 95%?

1. For item [e] the action might be legal. 95% confidence of failure accepts a risk of 5% that we face taking wrong action. Such a mistake is usually driven by the errors produced by having limited numbers of samples from monitoring.

1. For a failure that is merely noted, or which needs, say, action like increased monitoring, we might reduce the 95% confidence to 50%. For a deadly threat to public health, we might demand less than 1% confidence of failure (and so respond dramatically to the first glimmer of a risk).

1. The level of sampling, and the resulting errors in estimates of things like the annual mean or 90-percentile, dominate the risk of taking bad decisions. We must calculate these errors.
2. An annual mean from 36 samples may have errors of ±15%. For 12 samples this might be ±25%. Such errors can make it pointless to worry about other errors, or whether we need to opt for very elaborate calculations, or more complex standards.
3. Continuous monitoring may be needed at sites that are precious or face abnormal risks. There may be pollutants where such monitoring is affordable everywhere. We need to take care in using short bits of such data to calculate things like an annual mean – how we merge 10,000 results from a cold wet day in January with the 11 samples taken during the rest of the year.
4. A standard once used for bathing waters was:

### a level of 200 units

### met for 95% of the time

### over one summer, and:

### required at least 50% confidence of failure for action

### required compliance for 19 summers in 20

1. Item (e), 19 summers in 20, means that (b) is actually tightened from 95% towards 99%, or that the level set by (a) is decreased from 200 to 100. The standard is also tightened by (d) – low confidence of failure for action.
2. The severity of the standard is dictated by the total effect of (a) to (e). If an item is not set, it will tend to vary from place to place or from time to time. This is the same as allowing variations to the standard’s concentration.
3. For bathing waters, item (e) stems from the influence of weather, especially sunshine and storms. A better option is to average 3-5 summers – a line that was picked up in later years.
4. Work on toxicity may lead to a concentration that is declared “safe”. If this is to be our standard, we must nail down how often this concentration can be exceeded. The least we can do is to set up something like an annual mean or percentile. In doing this, we need to take into account any degree of precaution that was built into the “safe” value – the “safety factors”.
5. As mentioned above (§15), we usually see that the data used to set up an annual mean standard often confirms an underlying shape to the data used to calculate the annual mean for a particular determinand. This shape applies nearly everywhere. It is often a log-normal distribution.

1. This outcome simplifies the calculations of action to protect river quality and helps assess compliance with a standard. Errors in assuming a log-normal distribution are nearly always trivial compared with those accepted as produced by our rates of sampling (§87).
2. Why choose an annual 90-percentile and not a mean for a river quality standard? Usually this will be where the histograms of determinand concentrations can have a different fatness – cases where the ratio of the standard deviation to the mean varies from place to place. And we feel a need to control such variability, perhaps because high values have toxic effects. But errors from sampling will be much bigger for things like an annual 90-percentile than for the corresponding annual mean.
3. You might prefer a more complex standard – one that specifies how many hours a high value can be allowed to last and how often such events can be allowed to happen over a period of 10 years. You may want to simulate 50 years of minutes, hours and days in the life of a river. In some cases, it may be realistic to include all such details. But we must retain the ability to calculate unbiased values of statistics such as the annual mean and compare them with values recorded for rivers.
4. Computer simulations nearly always show that such embellishments are unnecessary and that a summary statistic works. This is reinforced by the errors we accept from choosing our rates of sampling.

## Confidence of failure

1. We might take 36 samples in 3 years and calculate the summary statistic. We then compare this with the standard. In doing this it is shameful not to calculate the “confidence of failure”.
2. We might decide that 95% confidence is needed to justify action. We might show as RED on a map all the places which have 95% confidence of failure. We then give priority to designing action to improve these places.
3. It is sometimes unpopular to demand as much as 95% confidence before taking expensive action. But the cost of extra monitoring is usually trivial compared with the wasted cost of action on failures that are not real – those that have “failed” because of bad luck with monitoring.
4. We may be told to act wherever we cannot show “no risk of damage”. Such a requirement still comes down to addressing the five numbers for the standard (§21).
5. We may face laws that require that rivers are classified as being in one of five classes: High, Good, Moderate, Poor or Bad. We must estimate the confidence that a site has been placed in the correct class.

1. We might collect 36 samples over 3 years. Even for these, the risk the class is wrong can be 25%. And the risk of declaring wrongly that class has changed is 30%. (Such calculations are easy to do).
2. Rivers have many standards. Some people and newspapers might declare that a site fails if any one of its 20 standards is failed. They may then colour this section of river red on a map. For 20 different standards, the site “fails” if any one of them fails.
3. This leads to a pessimistic bias in the number of sites reported as failed. Suppose a site is truly compliant but has six standards which risk being wrongly reported to fail one year in 10. This site will be reported as “failed” in 5 years in 10. For 20 standards this rises to 9 in 10.

1. This is no basis for national targets. Reported trends will be bad even when, in truth, things are improving. We should plan and report separately for each pollutant.

## Action

1. As noted above, we decide the balance between:
2. national action such as a ban on a chemical, or imposing fixed standards or fixed reductions on all of types of pollution of particular size or type;
3. the use of special controls calculated separately for each site or catchment;
4. Option (a) reduces the effect of errors from sampling in a national target. These are averaged out when we adopt a target of reducing the total number of failed sites. Our target might be to reduce the total of “bad” lengths of rivers – without being worried about which ones actually get better, or whether good quality rivers needed the improvement that was given to them.
5. For nitrate, failure might lead to general constraints on all farmers in the entire catchment. It may be unclear that this will secure compliance. The constraints may then need to be tightened in the future.
6. If a water is declared as “sensitive”, it might be that all sewage works above a certain size are required achieve specific and fixed effluent standards. This leads to improvements, but perhaps not always to rivers in best need of them.
7. We can use SIMCAT to calculate what option (a) might achieve.
8. Option [b] attempts to get a good decision on the action at each site: action to meet a standard at that place. This usually gives us more improvements for our money.

## Dilution and mixing

1. We’ll focus on discharges such as wastewater treatment works. The same logic applies to things like urban runoff, agriculture, mines, and all kinds of uses of chemicals.
2. We might think we can use a selected value for low river flow to calculate the discharge quality needed to achieve a standard in the downstream river. This makes sense only if the discharged load is constant throughout the year, and the river has no pollution from upstream sources.
3. Such decisions must be based on the full range of the values of the flow and quality of rivers and discharges, and how values are to be mixed together. This is also demanded by the need to calculate correctly the values of statistics like the annual mean or percentile (or for any other type of standard). The calculations will then include all the combinations of variations in dilution, upstream pollution, and discharge load.
4. This calculation requires a technique like Monte Carlo Simulation, or a simulation of several years in the daily life of a river. The simulation of a single month or a few days is misleading except where it is part of a simulation that covers several years.
5. The simpler method of Monte Carlo Simulation is sound in nearly all cases. The statistical errors linked to monthly or weekly sampling are larger than any extra precision that might be added by more complicated models. And the saving in the time and cost of making decisions might be enormous.

## Monte Carlo Simulation

1. We’ll look at mixing a single discharge with a river. We use the Mass Balance Equation:

### F is the river flow upstream of the discharge

### C is the concentration in F

### f is the flow added by the discharge

### c is the concentration in f

### T is the concentration downstream of the discharge

1. A single application of the equation cannot calculate the mean or percentile of discharge quality, c, that is needed to achieve a mean or percentile for T.
2. In Monte-Carlo Simulation, a single value for each of F, C, f and c is extracted from the full spread and frequencies of their possible values. The above equation is then used to calculate a value for T from these values of F, C, f and c. This is repeated, say, 2000 or 5000 times.
3. Usually, the thousands of sets of values of F, C, f and c are extracted from distributions that are assumed to be log-normal. But any forms of distribution can be used, including types that are non-parametric – based on the actual shape of the histogram of raw data. We can even select from sets of thousands of values of actual data, if we have them, and feel we need to use them.
4. The flows of rivers and from sewage works are affected by rainfall. To model this, we include, for example, the correlation between river flows and flows from sewage works. Correlation coefficients are easily calculated and can be used to tie together the 2000 pairs of values of F and f, and f and c etc (and/or with 2000 values for things like temperature or pH, etc).
5. We need data that characterise the distributions of F, C, f and c. In most cases, two summary items are enough. We use the most easily available:

river flow (F): mean and 95-percentile low flow

upstream river quality (C): mean and standard deviation

discharge flow (f): mean and standard deviation

discharge quality (c): mean and standard deviation

1. The results of the calculation provide the link between the distributions of c and T and how the mean and percentile values of T vary with the mean and percentile values of flow and quality for the discharge and the upstream river. (Monte Carlo Simulation for individual discharges is provided by RQP and MPER).

## Modelling a catchment

1. SIMCAT does Monte Carlo calculations for an entire catchment. The calculations investigate things like individual discharges, type of diffuse pollution, and plans for industry, economic growth, climate change, new standards, new policies and new laws.
2. SIMCAT works its way down a river, dealing, as necessary, with thousands of kilometres and hundreds of tributaries, abstractions, dis­charges and the sources of 20 types of diffuse pollution. Water quality is cal­culated down the whole length of the rivers.
3. Increasingly, data for SIMCAT are produced and edited by data bases and mapping systems such as SAGIS. SIMCAT does its calculations and provides results for SAGIS. SAGIS can then display the results on maps and clever diagrams. And the results can be combined to provide summaries that cover all the catchments in the country and in individual regions.
4. At points where effluents or diffuse pollution enter a river, SIMCAT uses Monte-Carlo Simulation to mix hundreds or thousands of values of flow and quality for the pollution with the corresponding values of flow and quality for the upstream river.

1. At all points in the catchment, SIMCAT tells us the full breakdown of pollution into the contributions from any and all of the upstream discharges or the zones of particular types of diffuse pollution. This shows us where to act in order to protect water quality.
2. For example, it can show that a discharge contributes 2.1% of the annual mean concentration at a point in the river that is 60 miles downstream. It would add that this 2.1 is, say, plus or minus 0.3.
3. As discussed below (§74), if dischargers do a lot of sampling of their inputs, the calculated figures for their particular contribution to river quality will be accurate at all points in the downstream river.
4. Runs of SIMCAT can also be used to estimate the limits needed at individual discharges to meet local river quality targets throughout the catchment. The resulting improvements proceed downstream.
5. As the river flows downstream, the thousands of values of river quality are adjusted to account for specified effects like natural decay and further diffuse inputs. The results will define the upstream quality for any subsequent inputs or places of interest.
6. At abstractions, the values of flow and loads will be reduced according to the scale and type of abstraction. This will also remove correctly, some of the upstream loads provided by sources of pollution. Such removals are also listed in the reports of downstream apportionments.

## Errors

1. SIMCAT calculates the confidence that any point in any river is worse than its standards.
2. The total effect of sampling is modelled. We include the effects of the sampling rates for all discharges and all rivers, and how these rates combine and affect the estimates of the quality down the whole of the downstream river. We can also add the uncertainties linked to any equations used in some calculations (such as for unionised ammonia and bio-available metals in RQP).
3. The errors make us think about the effort we actually need to devote to details like the in-river processes that affect water quality, or how we define the input distributions. There is little point in a lot work in acquiring lots of extra detail if the effects are much smaller than the errors from sampling.

## Calibration

1. When you assemble your data, you will be lucky if the results of your first run agree with all the measurements of flow and quality recorded by monitoring. To secure a fit you will check for mistakes and look for missing sources and sinks of flow or pollution. This is calibration.
2. SIMCAT can use a process called “gap-filling” to calibrate automatically, or to check your success in calibration. It does this by filling remaining gaps between observed and calculated results for means and percentiles by:

### adding extra river flows, or removing them, so that it reproduces exactly what is measured at things like flow gauges;

### making adjustments that secure agreement with the distributions of quality recorded at the monitoring sta­tions;

### reporting the gaps that have been found by gap-filling (additions or removals) as part of the apportionment of river pollution between all the upstream contributions.

### You might also use such gap-filling calculations to help identify what actually caused them.

## Examples from SIMCAT’s old work

1. It is often a quick and simple matter to examine issues with SIMCAT.
2. I have a map that shows the improvement of rivers achieved from 1980 until 2005 with the help from SIMCAT and RQP.
3. Years ago, 37% of 50,000 km of rivers were reported as failing new phosphate standards. If all sewage works received a fixed level of treatment, failure was calculated to reduce to 31%. This would cost £3.6 billion – something worth knowing early when deciding future policy and plans.
4. We calculated the contributions from diffuse pollution. A 30% reduction would reduce failures from 31 to 27%.
5. Removing phosphorus from certain types of detergents would bring only 0.4% of rivers into compliance with river quality targets, though even this had a calculated benefit of £130 million that exceeded the estimate of the cost of the action.
6. Lawyers proposed a definition of “no deterioration”. SIMCAT calculated that it would lead to costs of £13 billion to achieve this. A policy that avoided the wasted investment caused by sampling would cost only £2 billion. Early modelling can improve bad ideas.
7. New standards for nutrients for special habitats were shown to be unachievable even if all discharges were improved by 90%. We would also need to cut farming.
8. Climate change poses risks. But in early calculations, the changed river flows and higher temperatures and population were shown to produce a downgrade of only 0.4% of kilometres for 2050. But the effects of bigger and more frequent storms are more difficult to predict.
9. I once helped contest a legal action against a permit proposed for a discharge. A main point was the use of a log-normal distribution for river flow rather using the actual values. It claimed a proof that the data were not log-normal.

It is a simple matter to calculate the effect of such assumptions on the calculated river quality downstream of the discharge. SIMCAT and RQP can be run using the actual data. And set up to use 50,000 values.

The difference was trivial and negligible compared with the ±14% imposed from estimates of annual mean quality using samples of the river and the discharge.

Tony Warn

8th January 2024