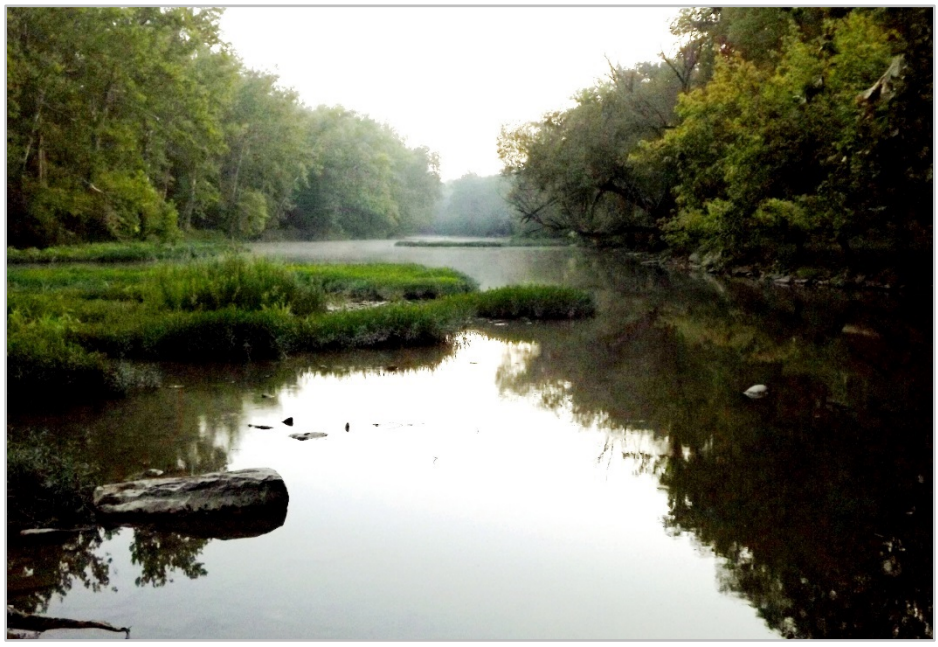
Ohio Environmental Protection Agency

Surface Water Field Sampling Manual - Appendix IV

Data Management



Division of Surface Water
May 19, 2021
Next Revision Due April 22, 2023

Revision History

This table shows changes to this controlled document over time. The most recent version is presented in the top row of the table. Previous versions are maintained by the Ohio EPA Division of Surface Water Modeling, Assessment and TMDL Section Manager.

History	Effective Date
Ohio EPA Surface Water Quality Sampling Manual version 8.0 Appendix IV: Data Management Table of Contents: Updated Section A: Updated and revised. Table of sample result value interpretation added. Paired parameter section updated.	May 22, 2021
Ohio EPA Surface Water Quality Sampling Manual version 7.0 Appendix IV: Data Management Section B: Removed text; section will be reserved for future use. Older version available upon request. Section C: Removed text; section will be reserved for future use. Older version available upon request. Section D: Removed text; section will be reserved for future use. Older version available upon request.	April 22, 2019

Table of Contents

Section A. Data Management

A1. Data Validation Guidelines for QC and Field Samples

For most DSW chemical water quality data, data validation is generally confined to evaluation of Blank results, Duplicate results, paired parameter results (defined below) and confirming that samples were properly preserved/prepared (including filtration, etc. - if indicated by the method). Standards for evaluation of analytical results of those QC sample types and general field samples are described below.

Data can be qualified using the standard qualifiers available as defined by DES (in their field handbook) such as "J" for an estimated concentration or "R" for rejected result as well as one additional qualifier, "Trend." Some results may be too uncertain for some data uses but potentially useful for more general data trend applications.

Data qualifiers should be added by samplers to EA3 as part of their data review process. This will ensure the qualifier remains with the sample result.

Data Qualifiers

All sample results have some amount of uncertainty surrounding the quantification of analyte in a given sample. Data qualifiers are used to indicate that extra uncertainty is present surrounding a given result (e.g., "J" for estimated or "Trend" to indicate more uncertainty). The data qualifier "R", Rejected, is used to indicate that too much uncertainty is present to consider the result quantitatively (for most data applications). "Trend" is a qualifier used by DSW to indicate when data is considered to have less quantitative significance but enough for assessing data trends.

Blanks

Blank contamination can result in qualification of other results that were in the same field batch as that blank. In some cases, these other results may still be useable and other times the sample results should not be considered valid, largely depending on the concentration in the sample vs. the concentration in the blank.

Laboratories often use a factor of three to differentiate a detected compound from background "noise" present in the system (analytical instrument, etc.). When a result exceeds three times the background noise, it is considered to be positively identified in the sample. We can consider blank contamination as extra "noise" in the system, since we don't know the source of the contamination, and use this factor of three to help us assess our data. To do so, the sample concentration must be at least three times the blank concentration for us to be confident that analyte is truly present in the sample.

For the purpose of DSW result qualification, blank contamination is defined as a blank result above the parameter's Reporting Limit (RL). At times, an associated sample result may be between the Method Detection Limit (MDL) and RL, therefore the analyte is found to be present, but the value is highly uncertain. In these cases, the result will not be rejected, and the most severe qualification would be "J."

"B" qualifiers are assigned by the lab for any result exceeding QC rules for blank contamination according to their guidance, independent of DSW procedures.

Sample Result Value	Interpretation
MDL < Sample ≤ RL and Sample ≤ 10x Blank	Indicate an extra level of uncertainty that the sample result is above MDL ("J" qualify)
RL < Sample ≤ 3x Blank	Reject sample results in this range as insufficiently different from blank results
3x Blank < Sample ≤ 5x Blank	Likely indication that the analyte is present beyond contamination, but poor confidence in the numerical result - generally limit data use to data "trend" applications
< 5x Blank < Sample ≤ 10x Blank	Consider the sample result to be an estimated concentration (qualified "J") but still suitable for most data uses
10x Blank < Sample	Do not qualify data (blank contamination does not significantly change the result within the uncertainty of the value reported)

Blank qualification examples:

Blank Result	Sample Result	Method Detection Limit	Reporting Limit	Qualifier	Reason
8	<2	2	5	No qualifier	Result ≤ MDL
8	4	2	5	"J"	System uncertainty
8	7	2	5	"R"	Result ≤ 3x Blank
8	16	2	5	"R"	Result ≤ 3x Blank
8	29	2	5	"Trend"	< 3x Result ≤ 5x Blank
8	79	2	5	"J"	< 5x Result ≤10x Blank
8	81	2	5	No qualifier	Result > 10x Blank

Note: If Data Quality Objectives for a parameter are well above (>5x) blank contamination and field sample results use of only a "J" qualifier, instead of "Trend" or "R", may be warranted.

Field Duplicates

Laboratories analyze and evaluate duplicates for their own internal procedures but DSW staff collect field duplicates to evaluate variability regarding sampling precision for field QC. Duplicates must be submitted "blind" to the laboratory in order to properly assess precision. The duplicate sample results are compared using a statistic called Relative Percent Difference (RPD).

RPD - Relative Percent Difference: $$\% \, Diff. = \left| \frac{x_1 - x_2}{\left[\frac{(x_1 + x_2)}{2}\right]} \right| x \; 100$$

In the %RPD example below one sample result/concentration is substituted in the equation for x_1 (6) and the other for x_2 (10 - it doesn't matter which is which in this equation - but traditionally the duplicate will be x_2).

Example RPD calculation:

$$\frac{|(6 - 10)|}{|((6+10)/2)|} \; x \; 100 = \frac{|-4|}{|8|} x \; 100 = 0.5 \; x \; 100, \text{ (positive since it's an absolute value)}$$

RPD = 50%

We allow a higher %RPD at lower concentrations and a lower %RPD at higher concentrations, since there is a greater percent uncertainty closer to the detection level and %RPD at lower concentrations corresponds to a smaller absolute difference. To account for this varying acceptable %RPD, we assess our duplicate samples using a curved line.

By starting with three points based on the ratio of the sample concentration to the detection limit and the %RPD we are willing to accept, we can generate the equation of a line. The three points used were:

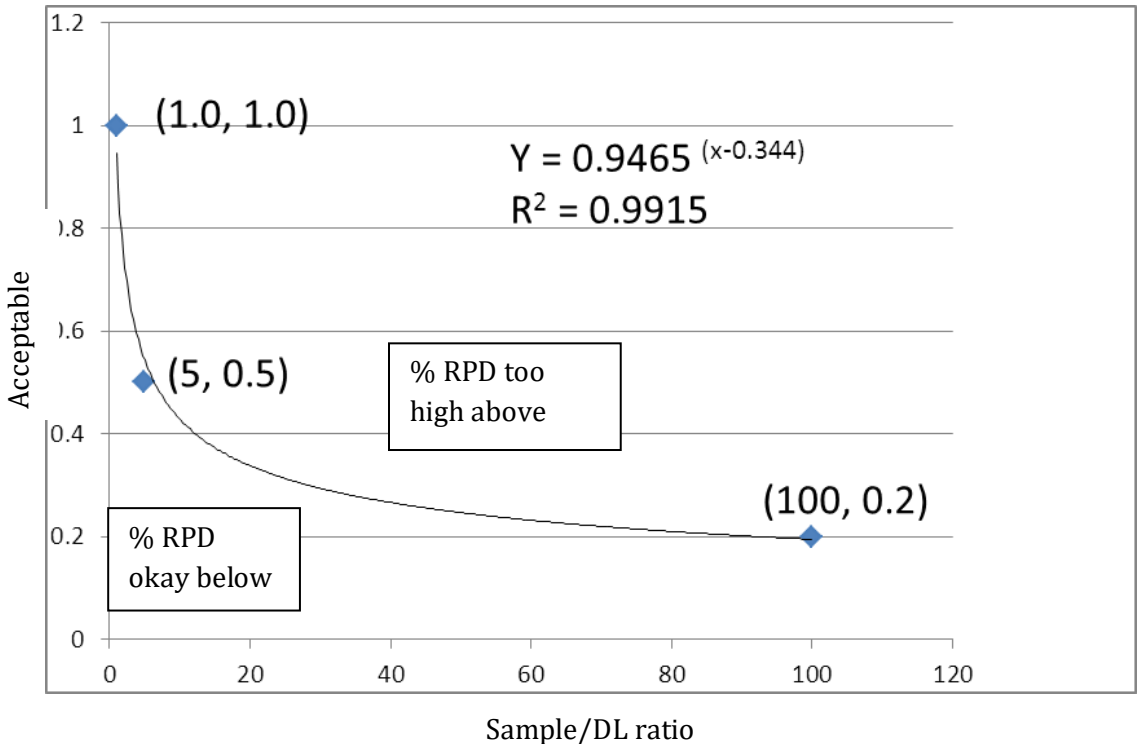
(1, 1.0) – At the minimum detection limit, we are willing to accept approximately 100% RPD

(5, 0.5) – at 5x the detection limit (often near the RL), we are willing to accept approximately 50% RPD

(100, 0.2) – at 100x the detection limit, we are willing to accept approximately 20% RPD

The graph (taken from Excel, using the "Power" option from the "Trendline" function) shown below illustrates the curve of best fit for these three points. The resulting R^2 value confirms a good fit of our line to our points.

Figure 1. Duplicate Maximum Acceptable %RPD Using Sample Concentration/ DL Ratio



Using "Trendline" in Excel, we can generate an equation with a very good fit to these three points. With additional tweaking of the equation (adding 5% to each result,) we get a result that gives us almost exactly 100% RPD when the sample concentration equals the detection limit and puts us back up above 10% RPD for high concentration samples (see the table below).

The resulting final equation is $Y = [(0.9465x^{-0.344}) * 100] + 5$

where x = Sample/DL ratio and y = acceptable %RPD

Using the above equation, we get acceptable %RPDs at the following levels:

Determine Maximum Acceptable %RPD (based on sample concentration to DL ratio)

Sample* Conc./DL (x)	"Trendline" equation from Excel Y = $(0.9465x^{-0.344})*100$	y' = $[(0.9465x^{-0.344})*100]+5$ (add 5% to baseline eqn.)
1	94.65	99.65
2	74.57	79.57
5	54.41	59.41
10	42.87	47.87
50	24.64	29.64
100	20.41	24.41
200	15.30	20.30
1000	8.79	13.79

**Note the duplicate sample concentration. For sample results below the minimum detection limit (and the duplicate is above the MDL), use the MDL in the Duplicate Maximum %RPD calculations (otherwise there is insufficient latitude for variability at low concentrations).*

This leaves us with a two-tiered system for duplicates. If our %RPD is below the values from our equation (i.e., below the curve), we accept the data as valid. If the %RPD exceeds the %RPD from the equation, we don't know which value to believe is correct, the sample or the duplicate value, so we must reject ("R" qualify) the data. At that point, particularly if multiple duplicate pairs have been rejected, the sampler(s) should look into possible causes for the disagreement and work to minimize those causes for future sampling.

Paired Parameters

DSW evaluates two types of paired parameters—subset parameters, where one is a fraction of the other, and a comparison of lab and field measurements of the same parameter.

Subset Parameters

There are some parameter pairings that DES evaluates (using %RPD) in tandem, since they are related. We can make use of these assessments too. Some parameters are fractions or subsets of others, such as nitrate being part of nitrate/nitrite, so that the one parameter should, in theory, never have a higher concentration than the other parameter.

Examples of subset parameters include:

TOC ≥ DOC
Nitrate/Nitrite ≥ Nitrate
Total P ≥ orthophosphate (or dissolved reactive phosphorus)
Total Cr ≥ Hexavalent Cr
TKN ≥ Ammonia
BOD ≥ Dissolved BOD (or other dissolved parameter pairings)

It's theoretically possible that the subset analyte could be 100% of the total (or larger) analyte, but any result where that compound exceeds the total (or larger compound) should be considered an estimated concentration (qualified with a "J"). Results that are quite close may be essentially the same number and valid for most data uses. Similar to how we evaluated duplicate samples above, we will use the same equation to determine the acceptable %RPD for "Paired Parameters" analytical results within the same sample.

For "Paired Parameters" with a %RPD less than the equation amount (using an average Detection Limit this time, since they may be different), we will simply acknowledge the difference with a "J" qualifier, leaving both data points as useable for most applications. However, when the %RPD exceeds the amount from the equation, we will generally not use the two data points and reject (qualify with an "R") the results. In this situation we don't know which result to believe and they are too different for us to be comfortable with the variability present. This all applies only when the subset parameter has a higher concentration than the expected larger/parent parameter. If the subset parameter has a lower concentration, then no evaluation/qualifiers are needed.

Example data for "Paired Parameters" assessed using the maximum %RPD equation:

$Y = [(0.9465x^{-0.344}) * 100] + 5$ (where x is the "parent" sample concentration/MDL and Y is the max. %RPD).

Subset parameter – example concentration	Parent (larger) parameter – example concentrations	Subset MDL (DES webpage*)	Parent MDL (DES webpage*)	Average MDL	%RPD (Parent and Subset)	Max. Allowed %RPD (from the eqn.)	Data Qualifier
Cr^{+6} – 3.6	Tot. Cr – 3.5	3.4 ug/L	0.28 ug/L	1.8 ug/L	2.82	75.87	"J"
Cr^{+6} – 7.5	Tot. Cr – 3.5	3.4 ug/L	0.28 ug/L	1.8 ug/L	72.73	75.87	"J"
Cr^{+6} – 7.8	Tot. Cr – 3.5	3.4 ug/L	0.28 ug/L	1.8 ug/L	76.11	75.87	"R"
Cr^{+6} – 24	Tot. Cr – 16	3.4 ug/L	0.28 ug/L	1.8 ug/L	40.0	44.98	"J"
Cr^{+6} – 26	Tot. Cr – 16	3.4 ug/L	0.28 ug/L	1.8 ug/L	47.62	44.98	"R"
Cr^{+6} – 16	Tot. Cr - 26	3.4 ug/L	0.28 ug/L	1.8 ug/L	47.62	38.06	None (par>sub)

** Detections limits may change – make sure that you are using the MDL associated with your data from that day's DES analysis. For results below the detection limit, the minimum detection limit in the Paired Parameter Maximum %RPD calculations.*

Lab vs. Field Comparison

There are some parameters that are measured in the field and in the laboratory using different equipment and different methods. For these results, a "Paired Parameter" evaluation involves a comparison of both %RPD and absolute difference. While this method could be adjusted for other sets of lab/field parameters, Ohio EPA currently uses it for the only common pairing – conductance.

The threshold for qualification of field conductance requires the result meeting two criteria:

> 10% RPD between field and lab results

and

> 50 µmho/cm absolute difference between field and lab results

If both these criteria are met, the field results for specific conductance (corrected conductivity) and uncorrected conductivity are rejected. Lab conductance, considered the more accurate and robust measurement, is not qualified. If just one or neither of the above criteria are met, there is no qualification of results.

A2. Reserved for other Data Management Topics in the future.

Section B. Reserved - EA3 Station Module Manual

Older version available upon request.

Section C. Reserved - EA3 Manual

Older version available upon request.

Section D. Reserved – Sample Master® Instruction Manual

Older version available upon request.