

QA/QC Automated Data Evaluation Procedure Version 2.0

By Chris Herrington, Analyst
Environmental Resource Management Division,
Watershed Protection & Development Review Department (WPDRD)
City of Austin, Texas

In order to standardize quality assurance and control procedures for the Water Resource Evaluation (WRE) Field Sampling Data Base (FSDB), an automated method of evaluation was programmed. The following report provides a summary of these procedures including logic and potential uses in screening data for analysis and presentation. A two stage data entry method is used to reduce transcription errors and allow field personnel to review laboratory and field results as they are entered into the database. Flags attributing descriptors of data quality are used to characterize QA/QC results for individual samples and parameters. Accuracy, precision, and laboratory blank data are used to generate the final flags for each data point. Additional modifications for the future are presented including ion balances, post calibration checks, and holding time verification in hours.

Overview of Function and Deployment

- **What is the Data Approval and Flagging Process?**

The Data Approval and Flagging Process is an automated evaluation of the Quality Assurance/Quality Control (QA/QC) data associated with a given sample result which, according to pre-determined control limits, results in the generation of a single-character flag enabling both Water Resource Evaluation (WRE) staff as well as members of the general public who request data to have an instant understanding of the general accuracy and precision of that data point. QA/QC data include, but are not limited to, laboratory control standards, laboratory and field replicate/split samples, matrix spikes, field and lab blank samples and post-calibration of field instrumentation. QA/QC data can be large in quantity (a single data point may have many QA/QC data points) and complex, making assessment of data quality difficult. Thus, with the flag generated by the data approval process the data user may have an instant understanding of the validity of any given data point.

The flagging process consists of two modules, preliminary and final review, that increase in complexity and contain multiple procedures that are activated in a pre-determined sequence in an attempt to minimize database computation time and simplify organization of the process code itself. Both modules have been programmed in PL/SQL, and the code for the packages are available upon request from Field Sample Database Staff. These modules are controlled and activated by a central program unit named QC_FLAG, which is the primary program unit and the unit called by the forms to activate the process.

Guidelines used to determine the acceptability of QA/QC data are stored in the FSDB_QC_LIMITS table within the database. These guidelines may be specified by project,

lab, medium, parameter, unit, method and filter fraction. Additionally, guidelines may change over time (based on effective date) so that acceptable limits may differ from one period of time to another to reflect changing data quality needs for any given project. For some tests, default guidelines are assumed when no guidelines have been specified.

Flow charts describing the function of each program unit are available from FSDB staff.

- **When does the process flag the data?**

The process will “fire,” or flag the data, whenever a record in the sample result table is inserted or updated.

Database staff recommend that QA/QC data be entered prior to entering sample results (Figure 1). Thus, a QC Batch ID will be generated and may be entered with the sample result data.

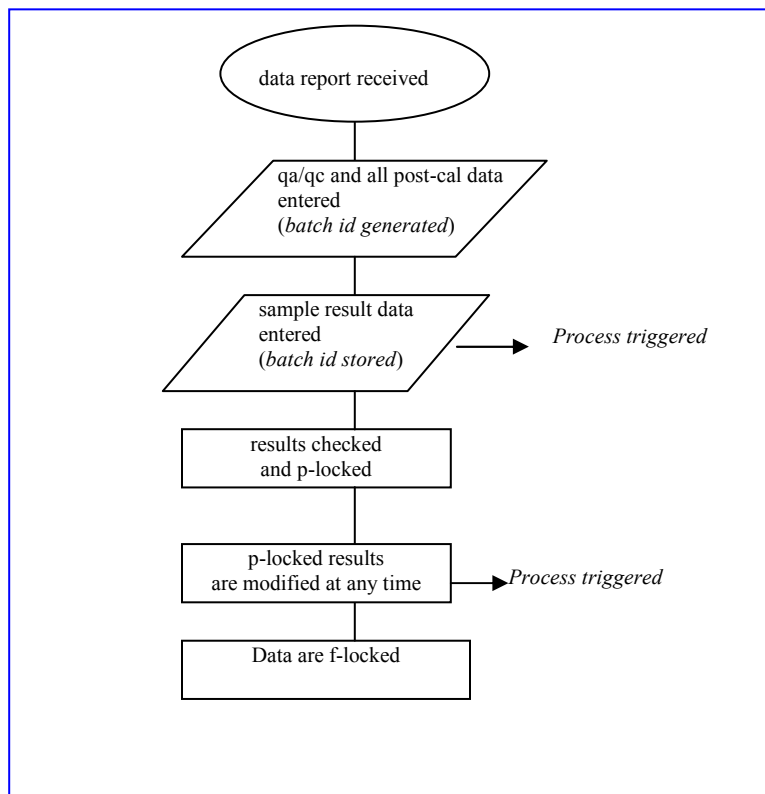


Figure 1. Recommended data entry flow diagram.

- **What is different in version 2.0 of this process?**

Analysis of post-calibration data has been added to the process. Post-calibration results are only examined if specified control limits exist in the FSDB_QC_LIMITS table. Failure of post-calibration will result in rejection of all associated sample results.

Analysis of field blank data has been added to the process. If field blank results are greater than sample results, data are flagged as estimated.

Logical tests have been expanded so that any dissolved, suspended or filtered result must be less than or equal to any total result for the same parameter, unit, and method within a given field sample. Conversely, all total results must be less greater than any dissolved, suspended or filtered result for the same parameter, unit and method within a given field sample. Failure of a logical test will result in rejection of all associated sample results.

Reliance on the project parameters table has been removed. A new FSDB_QC_LIMITS table has been added to the FSDB. If a result combination does not exist in the FSDB_QC_LIMITS table, that record is automatically added by the process and flagged for review by database staff. Additionally, default guidelines are now assumed for several QC tests if no guidelines are specified. The default guidelines are:

QC Test	Default Guidelines
Field Splits, Lab Splits	10% COV
Field Replicates	20% COV
Lab Standards	80-120 % Recovery
Lab Matrix Spikes	70-130 % Recovery

Field replicate and field split procedures have been revised using a PL/SQL construct known as a table array, improving the efficiency of these procedures.

More logical methods of dealing with censored data points were employed in this version of the process. Refer to the sample code for more information.

The list of possible final flags has been reduced, so that the only possible final flags are: R-rejected, U-usable, J-estimated, S-out of range but not rejectable.

- **What are the additional data quality considerations not considered by the process?**

The automated flagging process does not remove project staff from all responsibility to ensure the quality of data being collected. Full data review procedures are specified in the current Water Resource Evaluation Section Standard Operating Procedures Manual. Project staff must also be aware that there are additional considerations which should be addressed that are not handled at the present time by the flagging process.

A review of data documentation should be conducted by project staff to verify that the chain of custody has been maintained, sample identification is consistent, and all deviations from planned project guidelines are identified in the database for further consideration. This document review should be performed by the person responsible for officially receiving laboratory reports for a given project. Please note that the preliminary review will not be performed by the database.

Important items to be considered in the document review include:

1. Comparing chain of custody forms, field logbooks, data entry logbooks and laboratory data summary to verify the accuracy of all sample identification names or numbers, sampling dates and project personnel from the time of sample collection to production of final sample report by the primary laboratory.
2. Checking that all samples submitted for analysis have a value for all requested parameters.
3. Note any comments from field data sheets, field logbooks, or laboratory packets, that indicate possible data integrity corruption such as missing samples or incorrect analysis methods.

Data Flags and Test Guidelines

The automated data evaluation process is a multi-step procedure in which individual sample results are qualified according to the acceptability of associated QA/QC data by the database. Laboratory data will be automatically assessed to determine whether data should be rejected or qualified as usable to some degree by assessing adherence to data quality objectives related to precision, accuracy and distinction from background noise. For the purposes of this document, accuracy will refer to the difference between the reported value of the sample and the true value. Precision will refer to difference between repeated values for the same sample. Accuracy of data is evaluated by comparing laboratory control standards, matrix spikes and post-calibration results to test guidelines. Precision of data is evaluated by comparing split and replicate samples generated in both the field and laboratory to test guidelines. Difference from noise is established by comparing blank results to sample results.

Please note that the data qualification process is performed by the database. The procedure consists primarily of two modules, preliminary and final review, that work in conjunction to generate a single-character flag (Table 1) stored in the SAMPLE_RESULT table with the associated result.

Table 1. List of the potential final flags generated by the data evaluation process and stored with the sample result in the SAMPLE_RESULT table.

QC Flag	Description of Flag
U	Results are completely usable
R	Results are completely unusable
J	Results are to be considered an estimated value
S	Result was out of standard range but not rejectable
?	No sample result data exists

In the current version, the data evaluation process considers eight types of QA/QC analyses. The guidelines for these tests originate from information distributed by contract laboratories or published in guidance documents such as Standard Methods. Where no information could be found for a given parameter and QA/QC test, guidelines were assumed by database staff using conventional assessment limits (Table 2).

Table 2. Summary of general guidelines for QA/QC tests considered by the Data Evaluation Process. Note that a “split” identifies two or more samples generated from the same larger container whereas a “replicate” identifies two or more samples collected from the same location at approximately the same time in different sample containers.

QA/QC Test	General Limits
Standard Range Check	The 5 th and 95 th percentiles of all data were used as the lower and upper limits of the standard range check for each specific project-medium-lab-parameter combination.
Laboratory Splits	Percent coefficient of variation must be less than or equal to 10% between laboratory split(s) and sample.
Field Splits	Percent coefficient of variation must be less than or equal to 10% between field split(s) and the sample.
Field Replicates	Percent coefficient of variation must be less than or equal to 20% between field replicate(s) and the sample.
Logical Test	Generally, no dissolved, suspended, or filtered result may be greater than a total result for a given project parameter combination. Additional logical tests apply for nitrogen results (NH ₃ , Organic N must be less than TKN) and phosphorus results (orthophosphorus must be less than total phosphorus).
Lab and Field Blank Samples	Sample results must be greater than blank results
Laboratory Control Standards	Percent recovery of lab control standards spikes must lie between 80 and 120%.
Laboratory Matrix Spikes	Percent recovery of lab matrix spikes must lie between 80 and 120%.
Post-calibration Check	Post-calibration results must be within specified limits.
Holding Times	Current holding times range from 0 (analysis must be performed on the same day) to 180 days, and result from information listed in Standard Methods or lab documents.

Preliminary Review Module

The first sub-part of the data evaluation process is contained within the package “QC_PRELIMINARY” and contains the code for the Range Check, Logical Test, Field Split, Field Replicate and Field Blank evaluations. Note that the entire process is controlled by the QC_FLAG package. Though labeled “preliminary,” this designation generally arises from an organizational need to compartmentalize the code for the application. However, these program units have been separated into the “preliminary” package because they may be performed on any result, even if the QC Batch ID is null.

The program units for the QC_PRELIMINARY package include the following tests:

1. A **logical test** to compare parameter results to results for related parameters should be performed wherever applicable. An example of a logical test may be that if a parameter X is defined as the sum of the values of parameter Y and parameter Z, then the values for parameters X and Y may not individually be greater than the value for parameter X. Failure of the logical test results in the data point being rejected. If multiple logical tests exist, then a single failure will terminate the procedure and result in a rejected value.
2. A **range check** to compare results to a standard set of expected minimum and maximum values should be performed. The purpose of a range check is to identify any unusual results that should be examined in greater detail by the project personnel. Although results of the range check are evident in the “description” field of the sample result table, the QC flag will only be “S” if the sample is both out of range and does not fail any other tests.
3. A **field split** check to compare the percent coefficient of variation between sample results and split samples generated in the field. Field splits are a test of the precision of the data, and are two samples generated from the same sample container by collection staff in the field. Other than being generated in the field, a field split is functionally no different from a lab split. A default %COV value of 10 will be applied if no guidelines are specified. Failure of a field split %COV to be within control limits will result in the rejection of associated results. *Future modifications to the process may be made to dynamically set the precision criteria based on historically observed natural variation for a given analyte.*
4. A **field replicate** check to compare the %COV between sample results and replicate samples generated in the field. Field replicate samples are samples collected in individual bottles at approximately the same place and time, thereby yielding some estimate of the potential heterogeneity present at the time of sample collection. Field split samples are similar to field replicate samples, although field splits are generated from the same sample container separated into multiple smaller containers and thus should exhibit a minimum of natural variation. A default %COV value of 20 will be applied for field replicates if no guidelines are specified. Failure of a field replicate sample %COV to be within control limits will result in the flagging of the sample result as “estimated.”

Future modifications to the process may be made to dynamically set the precision criteria based on historically observed natural variation for a given analyte.

5. A **field blank** to verify that sampling procedures are not adding contaminants to sample containers. Sample results must be greater than or equal to field blank results. If blank results are greater than sample results, the sample will be flagged as “estimated.”

Final Review Module

The final module contains those qa/qc tests generally performed by the contract laboratory and include the Lab Blank, Lab Split, Matrix Spike and the Lab Standard procedures in addition to the Holding Time and Post-Calibration evaluations.

In general, the final review process uses the “worst-case scenario” when multiple results for a given test are present. Thus, a single failure will result in the termination of that process and a failure flag as the outcome of that process. Conversely, all tests present must be within control limits to obtain a positive test outcome.

The final review module contains the following procedures described below:

1. Results from **laboratory blank** analyses are evaluated to demonstrate that target analytes have not been introduced to the sample as a result of the sample handling procedures used in the laboratory, sample transportation process, or sample collection process, and that detected results are distinct from background interference. Lab blank analyses will be compared to sample result values to determine if sample results are greater than the blank value or if sample result values are less than or equal to the blank value. If more than one blank has been analyzed, then a single failure will terminate the procedure and result in a failed blank flag. If the blank value is not less than or equal to the sample result, the result must be flagged as an estimated value pending outcome of other tests. *Future modifications to the flagging process may add a multiplier to the blank value based on a relative standard deviation for multiple blank values, when present, for a given analyte.*
2. Results from **laboratory splits** will be evaluated by the same standards as field splits to test the **precision**, or reproducibility, of the data evaluation methods. A default value of 10% coefficient of variation will be used if no guidelines are listed. If multiple lab splits are present, then a single failure will result in the rejection of associated sample results. *Future modifications to the process may be made to dynamically set the precision criteria based on historically observed natural variation for a given analyte.*
3. Results from **lab control standards** will be evaluated to assess the accuracy of the data. Laboratory control standards, obtained from an independent source of the calibration standards, are used to determine if the laboratory is operating within control limits in order to demonstrate accuracy in the sample results. A default value

of 80-120% recovery will be used if no guidelines are present. If multiple lab control standards are present, failure of any standard will result in the rejection of associated sample results.

4. Results from **laboratory matrix spikes**, known additions of analyte to known sample in order to indicate what effects the sample matrix might be exerting on the accuracy of the sample analysis methods, will also be used to evaluate the accuracy of analysis methods. Default values of 70-130% recovery will be used if no guidelines are present. If multiple matrix spike results exist, failure of any spike will result in the flagging of associated sample results as estimated, pending outcome of other tests.
5. Sample **holding time** will be calculated as the time in days between sample collection and analysis. Holding times will be compared to specified guidelines. If no guidelines are present, the holding time is not evaluated. If the sample was analyzed outside of acceptable holding times, the associated sample results are rejected.
6. **Post-Calibration** of field instrumentation will be compared to specified control limits. Failure of the instrument to post-calibrate within acceptable limits will result in rejection of the result as unusable. No post-calibration check will be performed if control limits are not specified for the result in question in the FSDB_QC_LIMITS table, even if post-calibration data are present.

Calculation of Flag Calculation

The full and final qualification of the data depends on the results of each of the preliminary and final tests performed, and is handled by the QC_FLAG package which generates the single-character final flag that may be applied to the data to provide data users with a rapid assessment of the potential usability of the data. The final flag is stored with the sample result in the QC_FLAG column of the SAMPLE_RESULT table.

The final flag is determined by a function named FLAG_CONDITION located within the QC_FLAG package. The logic contained within that function is briefly described below (Table 3).

Table 3. Description of final flag calculation logic.

Condition	Final Flag
Failure of one or more major test (Logical Test, Lab Standard, Lab Split, Field Split, Post-calibration, Holding Time)	R (Rejection)
Outside of standard range (but no other test failures)	S (Outside Standard Range)
Failure of one or more minor test (Field Blank, Lab Blank, Field Replicate, Lab Matrix Spike)	J (Results Considered Estimate)
All tests with guidelines specified were within acceptable limits	U (Usable)
Sample result data value is null	? (No Sample Result Data)

Suggestions for Additional Modifications

Described below are suggestions for additional improvements that could be made to the existing Data Evaluation Process pending future staff availability and need.

- Holding times are currently calculated in days as most laboratories do not report the time of analysis, only the analysis day. However, new methods for samples such as bacteria require holding times that must be measured in hours. If contract labs could report to Field Sampling Database staff the actual time at which the sample was analyzed, the process could be modified to calculate holding times in hours.
- Groundwater staff have expressed an interest in including ion balance calculations in the Data Evaluation Process. Including the balance is not programmatically difficult assuming all specified member ions were analyzed for every groundwater sample, although inclusion of ion balances into the determination of the final flag may be difficult. How should a bad ion balance affect the usability of the data?