5. Data Management

This chapter describes the GAW data management process for precipitation chemistry measurements.

Data management encompasses the entire process of collecting, reporting, assuring, archiving, and distributing the values of the observed parameters *and their supporting metadata*. While it is the observed values that are of interest to the data user community, the supporting metadata plays an important role to help establish a level of confidence in the observed values. To maintain a high quality precipitation chemistry monitoring system, it is imperative that all aspects of the field handling, transportation, and laboratory processing for each collected sample be documented. Whenever a deviation from accepted protocols or a compromising event impacts a sample, it must be reported. This metadata about the sample must be stored with the observed values and carried along to provide a history and explanation of the sample.

5.1 Data Collection

Data collection begins at the site at the time that the sample is removed from the collector (see Field Operations chapter for details). All sample-related information is collected and reported on a *Sample History Form.* National Meteorological and Hydrological Services (NMHSs) that do not currently use a comprehensive sample history form are encouraged to do so. Examples of two existing sample history forms that have proven effective are shown in Figures 3.1 and 3.2. NMHSs that currently do not use sample history forms, or are not collecting detailed information on the forms, are encouraged to develop a suitable form from the examples given**.** To ensure that all critical field-related information is collected, the sample history form should have a section for field reporting and another for laboratory operations (or there might be separate forms for the field and laboratory sections). The style and content of sample history forms, by definition, will vary depending on the nature of the sampling protocols used by the individual NMHS (e.g., daily and weekly sample history forms will differ); thus, sample history forms must be custom designed to suit the specific sampling protocols of individual NMHSs. For this reason, no single sample history form is recommended for use in GAW; NMHSs are encouraged to develop their own. When designing a sample history form, it is critical that a set of mandatory data/information be collected for use in quality controlling the sample chemistry data. The critical information is as follows:

(a) Sample ID and Site ID
(b) Site Name
(c) Site operator
(d) Sampling start time and date
(e) Sampling end time and date
(f) Amount of sample collected (from the sample collector)
(g) Precipitation amount (from the standard rain gauge)
(h) Precipitation type.
(i) Condition of sample including sources of contamination.
(j) Condition and functionality of the site's equipment.
(k) Changes to site procedures and deviations from the standard operating procedures.

(l) Check-off box for “No Precipitation ” and “Trace Amount” to report dry or nearly dry conditions during the collection period.

5.2 Field Data Reporting

The field report section should be filled out at the GAW precipitation collection station for every sample. The field information should be written legibly, accurately and without omissions. The sample history forms should be reviewed upon receipt at the laboratory for problems and omissions. Corrective action should be taken immediately to fix any problems which have been identified on the field forms, (e.g., information which was inadvertently omitted should be filled in immediately by contacting the site operator).

The information collected on the field forms must be used to assess the quality of the sample data, i.e., to identify correct or incorrect operation of the sampling instrument, adherence or non-adherence to the standard operating procedures, and suitable quality of the precipitation sample. Site operators must be instructed to fill in all information that could identify improper instrument operation, poor sample quality, improper sample collection and handling methods, and unusual occurrences which could affect sample quality. Also included should be important site changes that could affect sample quality, (e.g., construction of a new building at the site).

> **NOTE: Sample history forms must indicate whether or not precipitation occurred during each sampling period.** A clear indication must be given with every sample if precipitation did or did not occur, e.g., if precipitation did not occur and no sample was obtained, this must be stated. This type of documentation is important because it confirms to the World Data Centre for Precipitation Chemistry that samples were not missing, but instead were not available.

5.3 Laboratory Data Reporting

After the sample and its associated history form arrive at the laboratory, the laboratory personnel are responsible for adding information to the sample history form. Critical laboratory information includes:

(a) Obvious damage to the sample (e.g., leaks, discoloration, exposure to extreme heat or cold.)
(b) Problems with laboratory equipment.
(c) Changes to laboratory procedures and deviations from the standard operating procedures.
(d) Indications of contamination.
(e) Method of analysis and equipment used.

For a more complete description of the information to include on the sample history form please consult the *Data Reporting Manual of The WMO World Data Centre for Precipitation Chemistry*.

5.4 Data Merging

When the chemical analysis of samples is complete, the NMHS must merge the chemical data with its associated information/data on the sample history form. The combined information should be retained in a local archive for use in quality controlling the data and for eventual reporting to the GAW World Data Centre for Precipitation Chemistry. Depending on the monitoring and research needs of each NMHS, the local archive might be as simple as a collection of files in a directory, or as elaborate as a computerized database management system.

5.5 Data Quality Control

5.5.1 Data Quality Flags

Each NMHS is responsible for quality controlling the GAW precipitation chemistry data prior to submitting the data to the WDCPC. The purpose of the quality control is to identify in the database those data which are valid or of questionable quality. This is to be done by reviewing the field and laboratory data for problems and inaccuracies, and by attaching a "data validity flag" to each datum, e.g., reporting SO_4^{2-} concentration as 2.1 V or 5.4 X where V= Valid and X=Mishandled or Contaminated. Data and their attached data validity flags are to be reported to the WDCPC following the protocols described in the following section.

Table 5.1: Data Quality Flags.

Analytical Data Flags	
Flag Code	**Definition**
V	Valid measurement above the method detection limit
L	Measurement is below the method detection limit
D	Measurement is at or below the method detection limit but has been set equal to the method detection limit
X	Sample mishandled or contaminated
M	Missing measurement
Gauge Value and Sample Mass Flags	
V	Valid measurement
X	Malfunction of the standard precipitation gauge/precipitation chemistry sampler
M	Missing Measurement

5.5.2 Data Quality Control Checks

In addition to the quality control checks that are performed in the field and laboratory as the precipitation chemistry measurements are being taken and analyzed (see Chapter 3, Sections 3.2.1.4, 3.3.2.3, 3.5, and Chapter 4, Section 4.3), there are three specific checks that can be applied to individual sample values. They are: ion balance, measured versus calculated conductivity agreement, and sea salt corrections.

The principle of electroneutrality in precipitation requires that total anion equivalents equal total cation equivalents. According to this principle, ion balance and conductivity balance in a precipitation sample should be checked by the methods described in Chapter 4, Section 4.3.13. Stations located near oceans or whose air masses are strongly affected by marine aerosols require corrections to their sulfate concentrations to account for the effects of sea salt sulfate. When the sea salt contribution to the sulfate concentration is very high, the non-sea salt sulfate concentration will be prone to high uncertainty, which may require the sample to be invalidated.

Sea salt sulfate corrections are usually based on the concentrations of Na^+, Mg^{2+} or (in rare cases) Cl^- found in seawater (Keene, et. al. 1986 and Hawley, et al. 1988). The QA/SAC Americas has a sea salt sulfate correction algorithm that calculates the expected contribution of sea salt sulfate to the total sulfate concentration measured in the precipitation sample. In the WDCPC database three pieces of data are kept:

(a) The total sulfate concentration as reported by the user (variable SO4)
(b) The expected non-sea salt concentration (as determined by the algorithm – variable XSO4)
(c) The correction method used.
(d) The sea salt sulfate correction algorithm is discussed in detail in Appendix H.

One issue with the sea salt correction algorithm is that it can generate negative values of non-sea salt sulfate when the total sulfate concentration is relatively low. Small negative values are likely due to round-off error or measurement uncertainty. If negative values are persistent and frequent, then they may be an indication of other problems such as drift in the calibration of the analytical instrument for one or more of the analytes. The presence of frequent negative values in a dataset should spark an investigation into the cause of the problem and might lead to a reanalysis of the samples.

Currently, QA/SAC Americas applies the sea salt sulfate correction algorithm only to sites located within 100 km of the ocean.

5.5.3 Significant Digits[1]

Databases and the scientific literature are full of examples in which authors did not pay adequate attention to the presentation of their results by displaying either an insufficient or an excessive number of digits. Partly, it is the problem of psychology - an author is trying to display his/her result in the best possible light. The problem becomes very serious when any assessment of results obtained by different scientific groups, at different times, and by different methods is carried out. The key point of the problem can be formulated like this: recording of measured values and related calculations reflect the true accuracy of the results with respect to significant digits.

The digits in a recorded value possess two qualities: significance and certainty. Certain digits carry real information. Significant digits comprise all the certain digits plus one additional digit

---

[1] **This section has been adapted from Yu. A. Zolotov** (ed.) *Basis of Analytical Chemistry. Part I: General Problems. Methods of Separation*. Moscow, <<Sysshaya Shkola>> Publishing House, p. 383, **1996.**

which is uncertain. Hence, the uncertainty of a result is equal to +/- 1 for the last significant digit (aka the first uncertain digit).

Rounding off any result must be done up to the first uncertain digit. Also, it must be remembered that the rounding off operation should not be performed until the final arithmetic operation is complete.

It is important that all GAW precipitation chemistry data be submitted to the QA/SAC Americas with the correct number of significant digits. To assist NMHSs with this, the following examples:

For Values Derived Using Addition and/or Subtraction. The significance level associated with an addition or subtraction operation is determined by the significance level of the algebraic summand with the fewest number of decimal digits. In the following example

$$50 + 2 + 0.55 \approx 53$$

the significance level of the result is determined by the summand “2” and the result should be recorded as “53”.

For Values Derived Using Multiplication and/or Division. As a simplified approach, the significance level of numbers derived from a multiplication or division operation is equal to the significance level of the co-factor with the fewest number of significant digits. This is illustrated in the following example

$$1.5 \cdot 2.35 \approx 3.5$$

in which the significance level of the result is determined by the co-factor “1.5” and the result should be recorded as “3.5”. This is the simplified approach. As a strict approach, the number of significant digits should be based on the relative uncertainties of the co-factors used in the calculations. Here, the relative uncertainty of a co-factor is calculated as the ratio of the absolute uncertainty of the number to the number itself. After determining the relative uncertainty of each co-factor, the relative uncertainty of the product or quotient should be set equal to the sum of the relative uncertainties of the co-factors. For example:

$$98 / 87.25 = 1.1232092...$$

In this case the relative uncertainty of the first co-factor is “$1/98 = 1 \times 10^{-2}$” and for the second co-factor: “$0.01 / 87.25 = 1 \times 10^{-4}$”. The relative uncertainty of the quotient is: “$0.01 + 0.0001 = 1 \times 10^{-2}$”. Hence the second digit to the right of the decimal point is the most uncertain and therefore represents the last significant digit. Therefore, the reported value should be rounded off to 1.12.

For Values with Multiple Zeros. A special situation occurs whenever a number ends with the one or more zero digits. In these cases, the reported result should be expressed in exponential form, such as 2.300×10^3. Here the leftmost zero is the last certain digit, while the last zero is the last significant digit.

For Exponentiated Values. The number of significant digits for values derived from exponentiation is based on the relative uncertainty of the product of the relative uncertainties of the base and the exponent. For the following example:

$$3.5^{2.5} = 9.8821...$$

the relative uncertainty of the base is 0.1/3.5=0.02851É. The relative uncertainty of the result is $0.02851 \cdot 2.5 = 0.0714 \approx 0.1$. Hence, the final result should be reported as 9.9.

For Logarithms. When taking the logarithm of a measurement, you should give as many digits in the mantissa (the part to the right of the decimal point) as there are significant figures in the measurement. For example:

$$\mathrm{Ln}(45.123) = 3.80939209$$

In this case the significant portion of the result is 3.80939 .

5.5.4 Treatment of values below the detection limit (BDL) and zero values

A major problem with the historical precipitation chemistry database is the treatment of concentrations and measurements below the detection limit (BDL). There are numerous occurrences of zero values throughout the older data. Since descriptive metadata is lacking and probably will never be found, these zero values are ambiguous. There are several possible reasons that a data originator may have coded a zero value:

(a) The value represents a missing datum.
(b) The sample was so contaminated or the deviation from established protocols so severe, that the value had to be invalidated.
(c) The concentration is below the detection limit.
(d) The datum has been lost.

Because zeros are ambiguous, they must be considered meaningless. This problem alone demonstrates the need for complete and descriptive metadata. All future data that is reported to the QA/SAC Americas MUST NOT report zero values. See the *Data Reporting Manual of The WMO World Data Centre for Precipitation Chemistry* for instructions for how to describe missing or invalid data.

Beyond the zero reporting problem is the more generalized difficulty with values below the detection limit. Without properly documented metadata, is it difficult, if not impossible, for the QA/SAC personnel to determine the detection limit from an examination of the data. Therefore, when reporting a datum that is BDL, it must be flagged as such. When reporting a BDL to the QA/SAC it is expected that in all cases the actual value obtained by the analysis will be reported.

This leaves open the question of how a detection limit is determined. Unfortunately, there is little consensus on this issue. Many definitions are based on the standard deviations obtained through the repeated analysis of blanks and other test samples. Inherent in these approaches is a tradeoff between flagging an excessive number of samples when the detection limit is set to a higher value and lowering the quality of the analysis when the detection limit is set to a lower value. In heavily polluted areas, the tradeoff is a moot point since the usual concentrations will be well above even the most restrictive detection limit criteria. However, at oceanic global stations concentrations of many species may lie at or below the detection limit. This implies a tendency to overestimate the concentration and deposition in relatively clean environments.

Given the lack of general agreement concerning the definition of the detection limit, data originators are given authority to decide on a definition that is appropriate for their needs. However, for most precipitation chemistry analyses, this definition is a reasonable guideline:
The detection limit is double the standard deviation determined from the measurements of blank samples.

Many analytical instruments today are manufactured with specific calibration routines, that usually use a set of calibration solutions. For these instruments, another method for determining the detection limit is recommended. After plotting the calibration curve, extrapolate the curve to the abscissa. The intersection of the calibration curve with the abscissa is the detection limit.

5.6 Data Submission to the QA/SAC Americas

After collection and quality control, GAW precipitation chemistry data must be submitted to the WMO GAW World Data Centre for Precipitation Chemistry, via the QA/SAC Americas:

WMO Quality Assurance - Science Activity Center for the Americas
Attention: WMO/GAW Data Manager
Illinois State Water Survey
Health and Environmental Applications Laboratory
2204 Griffith Dr.
Champaign, IL 61820-7463
USA

Telephone: +1 (217) 300-7420
E-mail: manager@qasac-americas.org
Website: http://www.qasac-americas.org

Each NMHS is responsible for submitting the GAW data to the QA/SAC Americas.

To assist data originators with preparing the data for data submission, Microsoft Excel templates have been created by the QA/SAC Americas for both metadata and chemistry data. Templates may be obtained by contacting QA/SAC Americas at addresses above.

The QA/SAC Americas has software that will read the Excel files and load the data into the WDCPC database.

Electronic submission of data files to the QA/SAC Americas is preferred. The following media and submission methods are recommended:

- e-mail (as an attachment)
- physical media sent by mail (physical media will not be returned.)

Data should be submitted to the QA/SAC Americas in a timely manner. The data should be submitted in annual increments by December 31 of the year following its collection.

5.7 WMO GAW Inter-laboratory Comparison Studies

Precipitation chemistry laboratories are encouraged to participate in as many inter-laboratory comparison studies as possible. This is an excellent way to not only gauge laboratory performance, but also to detect problems in analysis procedures. The WMO has been conducting intercomparisons of precipitation chemistry laboratories for well over 20 years (Visit the QA/SAC Americas website: http://www.qasac-americas.org for more information and access to the study results). Currently, the QA/SAC Americas in conjunction with the Illinois State Water Survey manages the GAW inter-laboratory comparison program on behalf of the WMO. Sets of three simulated rainwater samples are sent out to the participating laboratories twice a year.

All laboratories that analyze samples taken at GAW precipitation chemistry stations must participate in the sponsored WMO GAW Inter-laboratory comparison Studies and submit their analysis results by the deadline established by the QA/SAC Americas. GAW stations whose laboratories do not participate in the WMO GAW Inter-laboratory Comparison Studies will have their data sequestered from the WMO database. The basis for this policy is that the QA/SAC Americas cannot assess laboratory performance or data quality without using the study results.

The foregoing policy serves two broad objectives:

- To identify poorly performing precipitation chemistry laboratories and to assist the QA/SAC in correcting quality control problems at those laboratories in a timely manner
- To provide a means of describing the quality of the data produced by each laboratory to the international scientific community.

In this way, the data archived at the WDCPC will be of known and documented quality which, in turn, will serve to improve environmental assessments and models.

To enhance and maintain data of known quality within the WDCPC precipitation chemistry database, the QA/SAC Americas plans to incorporate the results of the laboratory intercomparison studies into the database of the World Data Centre. After each laboratory is assessed for performance, data from those GAW stations whose laboratories performed poorly will be flagged as qualified or invalid.

5.8 Data Archiving

The World Data Center for Precipitation Chemistry (WDCPC) archives precipitation chemistry data in electronic form. The types of data include wet inorganic ions, wet trace metals, and wet organic acids. In the future, the WDCPC may start accepting data from dry deposition monitoring efforts, but there are no immediate plans to do so. Generally, the data archived have been validated through control checks by the data originators and through assurance by the QA/SAC Americas staff. However, in extreme circumstances, the WDCPC will also accept raw and un-validated data (often referred to as "Level 0 data") to ensure its long-term retention. Level 0 data will be stored in a location separate from the precipitation chemistry database, and it will be made available only by special request.

Hardcopy documentation will not be retained in the WDCPC. Where appropriate, hardcopy data will be converted to electronic form and the electronic copy will be retained.

While the WDCPC is primarily concerned with precipitation chemistry data derived from WMO/GAW sources, it will archive bona-fide precipitation chemistry data of known quality from non-GAW sources as well. Non-GAW data will be archived only when it is given free of charge by the data originator, and when the data originator agrees to the terms of the WMO data distribution policy. Where possible, the WDCPC seeks to co-operate with other environmental archives.

5.9 Data Distribution

The WDCPC acknowledges and abides by the draft principles set by the WMO Congress in Resolution 40 (Cg-XII). A brief excerpt of these principles are:

(a) WMO Members have common ownership of the data held at the WMO World Data Centres (WDCs).
(b) The WDCs will provide data on a free and unrestricted basis.
(c) WDCs shall not accept in their holdings data for which there are restrictions for free and open access.
(d) Data archives of the WDCs must include readily accessible and comprehensive information describing the data sets, including quality assessments.
(e) WDCs should to the greatest extent possible, use media as well as processing and communication systems which are compatible with internationally accepted standards and protocols.

The WDC Managers in September 1998 endorsed the Resolution 40 principles at their meeting held at the Norsk institutt for luftforskning (NILU)[2]**.** They added the following recommendation concerning data originators who place restrictions onto the data:

> If data are from a GAW site or laboratory, then the WDC manager may insist that the data be posted without restriction. If the data are from a non-GAW site, then the WDC Manager may either refuse to accept the data or store the data in a protected or

---

[2] **Norwegian Institute for Air Research**

inaccessible directory.

5.10 References

Hare, E. W. and Carty, E. J. (1999) WMO/GAW Programme: Recommendations for a Harmonised, Metadata Header for Data File Submissions to the WMO World Data Centres. Meteorological Service of Canada.

Hawley, M.E., Galloway, J. N. and Keene, W. C. (1988) Standard error calculations for non-sea salt constituents in marine precipitation. *Water, Air, and Soil Pollution* **42**, 87-102.

Keene, W. C., Pszenny, A. A. P., Galloway, J. N. and Hawley, M. E. (1986) Sea-salt corrections and interpretation of constituent ratios in marine precipitation. *J. Geophys. Res.* **91**(D6), 6647-6658.

Olsen, A. R., Voldner, E. C., Bigelow, D. S., Chan, W. H., Clark, T. L., Lusis, M. A., Misra, P. K., and Vet, R. J. (1990) Unified wet deposition data summaries for North America: data summary procedures and results for 1980-1986. *Atmospheric Environment* **24A**(3), 661-672.