APPENDIX A. DATA QUALITY OBJECTIVES AND ASSESSMENT

This appendix describes the methods by which the Data Quality Objectives (DQOs) specified in Table A.1 were established and provides methods for determining whether laboratories are meeting DQOs, a process commonly referred to as *Data Quality Assessment*. A goal of GAW precipitation chemistry laboratories should be able to meet or exceed the DQOs in this table; though some laboratories may require improvements in their analytical and quality control methods. Each of the following sections refers specifically to the DQOs specified in Table A.1.

A.1   Detection Limits

The *Detection Limits* presented in Table A.1 were established from a review of the 2001 detection limits provided by all laboratories participating in the European Monitoring and Evaluation Programme (EMEP). Listed are the median detection limits of all participating EMEP laboratories.

The preferred method of determining *Detection Limits* in the GAW Precipitation Chemistry Programme requires making 30 (or preferably more) repetitive chemical analyses of a low- concentration solution. The solution must be stable with a concentration at or near the detection limit of the analytical method. Ideally, this means that the concentration should be between a blank level (i.e., pure deionized water) and the expected detection limit. In practice, the concentration can be as high as five times the anticipated detection limit but not higher. For example, if the anticipated detection limit of a sulfate analysis is 0.04 mg L^{-1}, then the solution from which the actual detection limit is established should have a stable sulfate concentration of less than 0.20 mg L^{-1}. Ideally, the concentration would be between 0.01 and 0.04 mg L^{-1}. The 30 or more repetitive analyses can be done in either a single analytical batch or in a number of batches analyzed over a period of time, the latter being preferred. The *Detection Limit* for each analyte is set equal to 3 times the standard deviation of the 30 or more repetitive analyses. When compared against the DQOs listed in Table A.1, those *Detection Limits* better than or equal to the DQOs are considered to have met the DQOs.

A.2   Overall Precision

Overall Precision is the precision of the complete precipitation chemistry measurement system, which includes both field and laboratory components. The DQOs for *Overall Precision* listed in Table A.1 were set at 1.5 times the precision values measured by one specific GAW network, that is, the Canadian Air and Precipitation Monitoring Network (CAPMoN). The precision values for this network are published in Table I of Sirois and Vet (1999). Here, and in Sirois and Vet (1999), *Overall Precision* is expressed as the Modified Median Absolute Difference (M.MAD), which is described in detail below. The factor of 1.5 was chosen arbitrarily to account for the wide variation in measurement methods and capabilities across the GAW member countries. It is worth noting that for Mg^{2+}, Na$^+$ and K$^+$, the *Overall Precision* as quantified by the M.MAD in Sirois and Vet (1999) was less than the analytical detection limit for these ions, in which case the DQO values were set equal to the detection limit times the factor of 1.5. The DQO for pH >5.0 was chosen arbitrarily since no precision estimates were available.

The estimation of *Overall Precision* by GAW member countries should be done by operating two identical precipitation chemistry collectors and standard gauges simultaneously at the same site for a period of one year or more. The paired concentration and depth data from the individual samples should be used to calculate the M.MAD. The M.MAD is a non-parametric estimator of the spread of the frequency distribution that is relatively insensitive to the presence of outliers and a consistent estimator of the standard deviation when the underlying frequency distribution is normal. The mathematical expression of the M.MAD is as follows:

$$M.MAD = \frac{1}{0.6745} \, Median(\, |\, x_i - Median(x_i)\, |\,) \qquad \text{Eq. A-1}$$

where x_i = variable of interest.

GAW Guidelines for Precipitation Chemistry Measurements – Appendix A (Updated - 30 June 2018; Original Version – November 2004)

A description and sample calculation of the M.MAD are given below. Readers are referred to Sirois and Vet (1999) for complete details.

A.2.1 Calculation of the Modified Median Absolute Difference (M.MAD)

1) For each set of paired concentration data, *i*, from Sampler 1 and Sampler 2, calculate the between-sampler error, e_i, as

$$e_i = \frac{1}{\sqrt{2}} (C_{1i} - C_{2i})$$ Eq. A-2

where C_1 and C_2 represent the concentrations from Samplers 1 and 2 for the ith sample and $1/\sqrt{2}$ accounts for the fact that the errors in the two measurements are assumed to be drawn from the same distribution.

2) Calculate the Overall Precision (which is defined as the spread of the e_i values) equal to the M.MAD as follows:
    a) From all values of e_i collected from the paired measurements, calculate the median value of e_i or Median(e_i).
    b) For each e_i, subtract the Median(e_i) and take its absolute value, $|e_i -$ Median(e_i)$|$.
    c) Determine the median $|e_i -$ Median(e_i)$|$, i.e., Median $|e_i -$ Median(e_i)$|$.
    d) Multiply the Median $|(e_i -$ Median $e_i)|$ by (1/0.6745), a factor that sets the M.MAD as a consistent estimator of the standard deviation, when the underlying distribution of $|e_i -$ Median(e_i)$|$ values is normal.

3) Compare the calculated M.MAD value with the DQO in Table A.1.

A.2.2 Example Calculation of the M.MAD

The following is a sample calculation of *Overall Precision* based on sulfate concentration data collected from duplicate CAPMoN samplers.

Sampler 1	Sampler 2	e_i	Median(e_i)	$|e_i$ -Median(e_i)|	Median $|e_i$-Median(e_i)|
5.234	5.453	-0.155	0.003	0.158	0.018
2.343	2.328	0.011		0.008	
2.359	2.335	0.017		0.014	
4.778	4.167	0.432		0.429	
0.736	0.733	0.002		0.001	
0.737	0.767	-0.021		0.024	
3.772	3.793	-0.015		0.018	
1.345	1.329	0.011		0.008	
11.787	11.426	0.255		0.252	
2.987	2.995	-0.006		0.009	
3.080	3.050	0.021		0.018	
1.095	1.098	-0.002		0.005	
1.636	1.631	0.004		0.001	
1.086	1.082	0.003		0.000	
3.207	3.314	-0.076		0.079	
1.756	1.788	-0.023		0.026	
1.772	1.778	-0.004		0.007	
3.118	3.102	0.011		0.008	
1.842	1.765	0.054		0.051	
2.719	2.677	0.030		0.027	
3.231	3.185	0.033		0.030	
1.239	1.289	-0.035		0.038	
4.392	4.354	0.027		0.024	
4.108	4.145	-0.026		0.029	
3.766	3.798	-0.023		0.026	
3.668	3.699	-0.022		0.025	
1.056	0.917	0.098		0.095	
2.560	3.580	-0.721		0.724	
0.863	0.870	-0.005		0.008	
1.202	1.207	-0.004		0.007	
2.812	2.809	0.002		0.001	
0.448	0.441	0.005		0.002	
15.412	14.285	0.797		0.794	
2.224	2.207	0.012		0.015	
2.621	2.631	-0.007		0.010	
0.920	0.928	-0.006		0.009	
1.642	1.601	0.023		0.020	
3.338	3.317	0.015		0.012	
5.910	5.839	0.050		0.047	

M.MAD = (1/0.6745) x 0.018 = 0.027 mg L^{-1}
DQO for *Overall Precision* from Table A.1 = 0.06 mg L^{-1}
Since the M.MAD < DQO, the *Overall Precision* for sulfate meets the DQO.

A.3   Laboratory Precision

Laboratory Precision is the precision of the analytical measurements made by GAW precipitation chemistry laboratories. The DQOs for *Laboratory Precision* listed in Table A.1 were set at 1.5 times the analytical precision of the CAPMoN laboratory (GAW laboratory of Environment Canada) as published in Table II of Sirois and Vet (1999). The metric used for the analytical precision of the CAPMoN Laboratory was the M.MAD calculated from a large number of between-run replicate analyses. The factor of 1.5 was chosen to expand the tolerance of the DQOs beyond that of a single laboratory in order to account for the wide variety of laboratory methods and measurement capabilities within the GAW Programme.

GAW Guidelines for Precipitation Chemistry Measurements – Appendix A (Updated - 30 June 2018; Original Version – November 2004)

Laboratories in the GAW Precipitation Chemistry Programme should estimate their *Laboratory Precision* by making between-run replicate analyses of 30 or more (preferably more than 30) precipitation chemistry samples covering a broad range of ion concentrations. As with *Overall Precision*, the *Laboratory Precision* is calculated as the M.MAD of these replicate data pairs, as described in Section A.2.

A.4 Inter-Network Bias

Inter-network Bias is a measure of relative accuracy or comparability between different networks that make the same measurements at the same time and place, and hence under the same environmental conditions. *Inter-Network Bias* is only a partial measure of *Overall Accuracy* but it is particularly useful in that it is readily quantifiable and easily understood.

Inter-Network Bias is quantified by making simultaneous side-by-side measurements of two or more networks at the same site. Each network must use its own instrumentation, standard operating procedures, analytical laboratory and data management methods to produce its data. The multiple networks' data sets are then inter-compared to determine the *Inter-Network Bias.* Collocated sampling must take place for at least one year in order to capture the full range of concentrations and environmental conditions that affect the measurements.

The DQOs for *Inter-Network Bias* shown in Table A.1 were established based on a number of assumptions about collocated sampling, namely:

- Collocated collectors from different networks sample the same population of precipitation events.
- Each network's measurement system produces a unique distribution of sampling data that reflects that network's field and laboratory measurement methods.
- Perfectly comparable networks theoretically can produce identical data distributions but, in practice, are unlikely to do so. Even collocated samplers from the same network typically produce different data distributions.
- Highly comparable networks will typically produce data distributions that differ by at least as much as collocated collectors from a single network. In other words, two different networks will generally differ by more than the *Overall Precision* of an individual network.
- Networks that are significantly different from each other (to a given probability level) will typically have data distributions that differ from each other by as much as or more than the precision of the individual networks.

Based on these assumptions, the DQOs for *Inter-Network Bias* in Table A.1 were set at 1.5 times the inter-network biases measured in a network inter-comparison study that involved two highly precise GAW networks, namely, the United States National Atmospheric Deposition Programme/National Trends Network and the Canadian Air and Precipitation Monitoring Network. The *Inter-Network Biases* measured in this study were published as the M.MAD in Table I of Sirois et al. (2000). The multiplicative factor of 1.5 was chosen arbitrarily as a way of expanding the tolerance in the GAW Programme to account for the wide variation of measurement methods and capabilities across GAW member countries. For Mg^{++} and K^+, the *Inter- Network Bias,* as quantified by the M.MAD in Sirois et al. (2000), was less than the analytical detection limit so the DQO values in Table A.1 were set equal to the detection limit times 1.5. The DQO for pH >5.0 was chosen as 2 times the DQO for pH <5.0, since quantitative bias estimates were not available for that pH level.

The *Inter-Network Bias* DQOs for sample depth and standard gauge depth measurements were arbitrarily set at ±5% for rain, ±15% for snow, and ±10% for mixed rain/snow relative to the total annual precipitation depth for each type of precipitation. The DQOs were expressed in relative terms (i.e., percentages) because of the difficulty in establishing absolute values (in millimetres) for the 50

or more national standard gauges (Sevruk and Klemm, 1989) and the more than 14 types of precipitation chemistry samplers in use today -- each having its own wind speed, exposure, wetting and evaporation errors (see Sevruk 1989 and Goodison et al., 1998 for a discussion of these errors).

GAW member countries can calculate *Inter-Network Bias* from collocated sampler data by following the method described in the next paragraph (it is assumed in this method that only two networks are collocated at the same site). Readers are referred to Sirois et al. (2000) for a more detailed description of the statistical model and methodology.

Calculation of *Inter-Network Bias* is best done by collocating a precipitation chemistry sampler and standard gauge from more than one network at the same site. After a minimum of one year of collocated sampler operation, tabulate the paired ion concentrations from the two samplers for all precipitation sampling periods. In cases where the sampling periods of the two networks are the same (e.g., daily-versus-daily, weekly-versus-weekly), this tabulation is straightforward. For cases where the sampling periods are different (e.g., daily-versus-weekly sampling periods), the concentration values of the shorter sampling periods must be converted to precipitation-weighted mean concentrations of the longer sampling periods. For example, if a daily sampler is collocated with a weekly sampler, the daily concentrations must be converted into weekly precipitation-weighted mean concentrations for the same weekly periods as the weekly sampler. To do this, the standard gauge depth must be used as the weighting factor, <u>not the sample depth</u>. Weekly sample depths and standard gauge depths must also be calculated and tabulated.

1) For each pair of samples for each sampling period, calculate the between-network difference in concentration, sample depth and standard gauge depth as $\Delta C_i = C_{1i} - C_{2i}$ where C_1 and C_2 represent the concentrations and depths measured by Networks 1 and 2 for sampling periods, $i = 1$ to n.

2) Test the frequency distribution of the between-network differences, ΔC_i, for normality (or non-normality). A test such as the Kolmogorov-Smirnov test can be used (see, for example, Gibbons, 1985).

3) (a) If the frequency distribution of the between-network differences is normal (which is unlikely), use the *student-t test* to test whether the mean value of the between-network differences, ΔC_i, is significantly different from zero.
(b) If the distribution is non-normal, use the *sign test* to test whether the median value of the between-network differences is significantly different from zero. The result will determine whether a statistically significant *Inter-Network Bias* exists. Note that special multiple testing techniques must be used if more than two networks are being compared in order to establish significant differences between networks while at the same time preserving simultaneous confidence intervals across multiple tests. One such test is the Friedman Rank Sum Test described in Hollander and Wolfe (1973) and used in Vet et al. (1988).

4) When the frequency distribution of the between-network differences is normal, calculate the mean value of the between-network differences. This value represents the *Inter-Network Bias.* When the frequency distribution of the between-network differences is not normal, calculate the median value of the between-network differences. This value represents the *Inter-Network Bias*. In either the normal or non-normal case, compare the *Inter-Network Bias* with the DQO in Table A.1. If the value is less than or equal to the DQO, the two networks are considered to have met the DQO for *Inter-Network Bias.*

5) For the sample depth and standard gauge depth measurements, calculate the annual totals of the rain, snow and mixed rain/snow events for each instrument. For each precipitation type, subtract the total depth measured by instrument 1 from the total depth measured by instrument 2 and divide by the average total depth of the two instruments. Compare this number to the DQOs specified in the table.

GAW Guidelines for Precipitation Chemistry Measurements – Appendix A (Updated - 30 June 2018; Original Version – November 2004)

A.5   Inter-Laboratory Bias

Inter-laboratory Bias is a measure of laboratory comparability. The DQOs for *Inter-Laboratory Bias* in Table A.1 were determined using data from WMO/GAW Inter-laboratory Comparison Studies 36 through 55. These 20 studies were conducted between 2007 and 2016 and the number of participants in each study ranged from 69 to 81 laboratories. Laboratories were challenged with 3 samples, each of different ionic concentrations. These 20 Inter-laboratory Comparison Studies of 3 samples each generated 60 sets of analytical measurements. The frequency distribution of measurements in each set was determined from which an *Acceptable Range* of measurements was calculated, as defined in equation A-3.

$$\textit{Acceptable Range (in percent)} \ = \ \pm \ [0.5 \bullet \frac{IQR}{Median} \bullet \ 100] \qquad \text鈥{Eq. A-3}$$

IQR designates the interquartile range of the frequency distribution, i.e., the middle 50% of the reported values. The *Acceptable Range* of a set of measurements, then, is half of the IQR expressed in relative terms by dividing by the median of the set.

The *Acceptable Ranges* of the 60 sets of measurements from the Inter-laboratory Comparison Studies were ranked from lowest to highest and the DQO for *Inter-Laboratory Bias* for each analyte was set at or close to the 90th percentile. This value was chosen to avoid the influence of outliers and was considered to represent a reasonable and attainable DQO for all GAW laboratories.

Individual GAW laboratories can use the DQO for *Inter-Laboratory Bias* in Table A.1 to evaluate their WMO/GAW Inter-laboratory Comparison results by following these steps:

1) For each measurement, calculate your laboratory bias as:

$$Bias = 100 \ \bullet \ \frac{(C_{lab} - MedianC)}{MedianC} \qquad \text{Eq. A-4}$$

   C_{lab} = laboratory's reported measurement
   $MedianC$ = median measurement of all laboratories

2) Compare your laboratory bias to the DQO for *Inter-Laboratory Bias* in Table A.1. The DQO is met if the bias is equal to or less than the tabulated value. If not, the laboratory did not meet the DQO for that measurement in that sample.
3) Follow the same procedure for every measurement in every sample.

Example: Laboratory reported a sulfate concentration of 2.31 mg L^{-1}. The median sulfate concentration of the 70 participants in the study was 2.54 mg L^{-1}.
Bias = 100 • (2.31 – 2.54) / 2.54 = -9.1%.
DQO for Inter-*Laboratory Bias* for sulfate (Table A.1) = $\pm$ 5%.
Conclusion: *Bias* > DQO, therefore laboratory did not meet the DQO for sulfate for that sample.

A.6   Calibration Levels

Multi-point calibration curves are mandatory for all ion determinations in GAW laboratories. As a general rule, at least 5 calibration solutions should be used to formulate each calibration curve. Even more solutions should be used when the analytical range is large, when the calibration curves are not linear, or when many of the concentrations in an analytical run are at or near the analytical detection limit.

The concentrations of the calibration solutions should vary evenly between the 2nd and 98th percentile values of the precipitation chemistry samples routinely analyzed by the laboratory. The exception to the foregoing is pH, for which two low-conductivity calibration solutions are recommended for calibration (pH = 4.0 and 7.0). To compensate for the paucity of calibration standards, several pH check solutions (i.e., stable low-conductivity solutions with a certified pH) should be included in each batch of samples to ensure that the pH meter is in control.

A.7   Data Ccompleteness

Data Completeness is an important data quality indicator when measurement data are summarized statistically over monthly, seasonal, quarterly or annual periods. This is because summary statistics such as the mean, median, and standard deviation) can be highly misleading when large amounts of data are missing or deemed invalid during the summary period. It is therefore incumbent upon all reporting agencies to ensure that they have a sufficient amount of data before reporting their summary statistics. A detailed analysis of the effects of missing data on precipitation chemistry statistics can be found in Sirois (1990).

Two DQOs for *Data Completeness* were selected for the GAW Precipitation Chemistry Programme:

The *Percent Precipitation Coverage Length* (%PCL). %PCL is the percentage of a summary period (e.g., month, season, year) for which there is a valid precipitation measurement. It is important to note that zero is a valid measurement when no precipitation occurred. In practice, %PCL is the percentage of the summary period that a standard precipitation gauge was operating properly and reporting precipitation depths. When precipitation occurred but there is no measurement of the depth, the depth measurement is missing and there was no knowledge of precipitation occurrence. Example: if a standard rain gauge reported data properly for 300 of 365 days in a year (including both precipitation and non-precipitation days), then the %PCL = 300/365 = 82.2%.

The *Percent Total Precipitation* (%TP). %TP for a given summary period (e.g., month, season, year) is the percentage of the total precipitation depth measured by a standard gauge *that was associated with valid precipitation chemistry measurement*. It is important to note that %TP must be calculated for each analyte, since for any given precipitation sample, one or more of the analyte measurements, but not all, may be invalid. Example: if a standard precipitation gauge measures a total of 1000 mm in a year but the total standard gauge depth associated with the valid precipitation chemistry samples was 782 mm, then the %TP = 78.2%.

The DQOs for *Data Completeness* in Table A.1 were taken from Olsen et al. (1990):

GAW DQO for Annual Data Completeness

%PCL:	%TP
Annual ≥90%	Annual ≥70%
Every quarter ≥60%	Every quarter ≥60%

GAW DQO for Seasonal Data Completeness

%PCL ≥90%
%TP ≥70%

Each GAW site should calculate its data completeness values for seasonal, annual and quarterly periods as follows:

$$\% PCL = 100 \bullet \frac{M}{N}$$

Eq. A-5

where
   M = number of days in the year, season, or quarter during which the standard gauge worked
   properly and reported data. A standard gauge that detects no precipitation on a dry day is
   operating properly and is included in M.
   N = total number of days in the year, season or quarter.

$$\% TP = 100 \bullet \frac{M}{N}$$

Eq. A-6

where
   M = total precipitation depth during the summary period associated with samples having valid,
      non-missing measurements.
   N = total precipitation depth associated with all samples collected during the summary period.

The calculated values of *%PCL* and *%TP* should be compared against the DQOs listed in Table A.1. Note
that networks must satisfy both the annual and quarterly *Data Completeness* DQOs when reporting their
summary statistics for annual periods. This prevents biases from occurring in the annual statistics caused by
large amounts of missing data within one or more seasons of the year.

The *Data Completeness* DQO for the standard gauge depth measurements has been set at
%PCL = 95% for a given annual period and %PCL = 90% for each of the calendar quarters within the
annual period. This means that a standard gauge depth measurement must be made for every sampling
period, with little to no tolerance for lost or missing data. The stricter requirement imposed for the standard
gauge depth measurements is because missing and lost data induce a direct negative bias into the
chemical deposition and loading estimates discussed elsewhere in this manual.

References

Gibbons, J. D. (1985). *Nonparametric Statistical Inference*, Second Edition. Marcel Dekker, Inc,
    Statistics Textbooks and Monographs: Volume 65, New York, NY.
Goodison, B. E., Louie, P. Y. T. and Yang, D. (1998) WMO Solid Precipitation Measurement
    Intercomparison Final Report, World Meteorological Organization Report WMO/TD No. 872,
    Geneva, Switzerland.
Hollander, M. and Wolfe D.A. (1973). *Nonparametric Statistical Methods*. John Wiley & Sons, Inc.,
    New York, NY, 503 p.
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. *Atmos. Environ.* **24A**(3), 661-672.
Sevruk, B. (1989) Reliability of Precipitation Measurement. International Workshop on Precipitation
    Measurements, St. Moritz, Switzerland, 3-7 December 1989, Report WMO/TD No. 328, World
    Meteorological Organization, Geneva, Switzerland.
Sevruk, B. and Klemm, S. (1989) Catalogue of National Standard Precipitation Gauges, World
    Meteorological Organization Report WMO/TD No. 313, Geneva, Switzerland.
Sirois, A. and Vet, R. (1999) The precision of precipitation chemistry measurements in the Canadian
    Air and Precipitation Monitoring Network (CAPMoN). *Environ. Monitor. and Assess.* **57**, 301-329.
Sirois, A., Vet, R. J. and Lamb, D. (2000) A comparison of the precipitation chemistry measurements
    obtained by the CAPMoN and NADP/NTN networks. *Environ. Monitor. and Assess.* **62**, 273-303.
Sirois, A. (1990) The effects of missing data on the calculation of precipitation-weighted mean
    concentrations in wet deposition. *Atmos. Environ.* **24A**: 9, 2277-2288.
Vet, R. J., Sirois A., Jeffries, D. S., Semkin, R. G., Foster, N. W., Hazlett, P. and Chan, C. H. (1988)
    Comparison of bulk, wet-only, and wet-plus-dry deposition measurements at the Turkey Lakes