

Monroe, S. A., M. Dyer, S. Stumpf, C. Bliss, and C. Parker. 2016. Water quality monitoring protocol for streams and springs in the Southern Colorado Plateau Network. Natural Resource Report NPS/SCPN/NRR—2016/1298. National Park Service, Fort Collins, Colorado.

**Standard Operating Procedure #10: Quality Assurance
Project Plan for Water Quality Monitoring, Version SCPN_
WATER_QUALITY_SOP10_20160901**

Contents

1 Introduction	3
2. Data quality objectives for measurement data	4
2.1 Water chemistry data	4
2.2 Water quality core parameters	5
2.3 Bacteria	7
3 Long-term data comparability	8
3.1 Personnel	8
3.2 Instrument replacement	8
3.3 Changes in technique or technology	8
3.4 Change of laboratory	9
3.5 Routine bias	9
4 Supplies and equipment selection criteria	9
5 Revision history	11

Standard Operating Procedure #10: Quality Assurance Project Plan for Water Quality Monitoring

Version SCPN_Water_Quality_SOP10_20160901				
Revision History Log				
Previous version number	Revision date	Section and paragraph	Change(s) made	Approved by
Only changes in this specific SOP will be logged here. Record the previous version number and date of revision; identify sections and paragraphs where changes were made, who approved the revision, and the new version number. More detailed descriptions of the changes made and the reason for the changes is provided at the end of this SOP.				

This Standard Operating Procedure (SOP) describes the Quality Assurance Project Plan for water quality monitoring in SCPN parks.

1 Introduction

A Quality Assurance Project Plan (QAPP) is the written document that outlines the procedures a monitoring project will use to ensure that the data it collects and analyzes meet project requirements (USEPA 2002a)

Quality assurance (QA) refers to the overall management process which includes organization, planning, data collection, documentation, evaluation, and reporting. QA provides the information needed to ascertain the quality of data and whether it meets the requirements of the program. QA ensures that data will meet defined standards of quality with a stated level of confidence.

Quality control (QC) refers to procedures used to assess data quality and ensure that

defined standards are met. Since data quality errors can occur in the field, the laboratory, or in the office, QC must be a part of all activities of the water quality monitoring program.

The SCPN Streams and Springs Water Quality Monitoring Program is intended to be a long-term (30+ years) project. In this QAPP, full consideration is given to potential changes in equipment, laboratory methods, and staff. The plan will be managed and implemented by SCPN personnel. Decisions regarding development of the SCPN Streams and Springs Water Quality Monitoring Program and long term monitoring plan have been made with input from park staff and SCPN water resource professionals. The program will be responsive to changing park and network needs. Broader-based, long-term decisions will be approved by the NPS-WRD, with additional guidance from the SCPN technical committee. These include

approval of WRD-funded work plans and approval of the overall water quality monitoring plan.

This SOP describes how to assess whether water chemistry data, core water quality parameter data, and bacteriological data collected using methods described in this protocol are credible and meet project objectives.

2. Data quality objectives for measurement data

This section explains how to evaluate whether data collected from this project is credible, and of known and documented quality. Data credibility necessary for the intended uses will be achieved when the data is:

- consistent both over time and between staff members
- collected and analyzed using standardized and acceptable techniques
- comparable to data collected in other assessments using the same methods
- used appropriately to make decisions based on sound statistics

Measurement sensitivity, precision, and bias are to be assessed in the following ways.

Sensitivity is the smallest amount of change that can be detected by a measurement. This is a concern for water quality parameters that are present in very low levels, or for those parameters that are likely to change in very small amounts over time. For laboratory analyses, sensitivity is expressed as detection limits—the lowest concentration that a method can reliably quantify. For field instruments, sensitivity is specified by the instrument manufacturer. NPS also recommends documenting Alternative Measurement Sensitivity (AMS) in the field for each instrument (Irwin 2008) in order to quantify instrument noise, as well as natural heterogeneity in a well-mixed stream.

For SCPN water quality monitoring, measurement **precision** will be considered as the agreement among repeated measurements of

the same property under identical or substantially similar conditions. SCPN will use two methods for calculating measurement precision: relative percent difference (RPD) and alternative measurement sensitivity (AMS). RPDs will be calculated for laboratory replicate samples and for turbidity samples, for which sample sizes are limited. AMS will be used for multiparameter sonde measurements, for which sample sizes are larger.

Bias is the systematic or persistent distortion of a measurement process that causes errors in one direction. SCPN will assess bias for laboratory analytes by collecting blank, spike, or reference-material samples (Wilde et al. 1999). For field parameters, bias will be assessed by comparing instrument measurements against known standards.

2.1 Water chemistry data

The National Water Quality Laboratory (NWQL) performs all laboratory analyses for SCPN samples. Data quality is assured by using standard quality control strategies for measurement sensitivity, precision, and bias; known and possible interferences; sample preservation and storage requirements; stability of reagents and standards; instrument performance; safety information; and method comparability to superseded methods, if any (USGS 2005).

2.1.1 Measurement sensitivity

For chemical analyses at the NWQL, measurement sensitivity is expressed as the long-term method detection limit (LT-MDL). It is determined from multiple years of data and is updated annually, or when methods change. It is the statistically-derived concentration of the analyte, where the risk of reporting a false positive in a blank sample is no more than 1%. The LT-MDLs are well below the state water quality criteria for SCPN's suite of analytes; therefore, these estimates meet the minimum requirements stated in Irwin (2008).

Minimum reporting levels (MRL) are “the smallest measured concentration of a

constituent that may be reliably reported by using a given analytical method” (Timme 1995) and are used to assure data quality. REporting levels are normally set to concentrations equal to twice the LT-MDL. Values above the LT-MDL, but below the MRL are reported by the NWQL, but flagged with an “E” (for “estimated”). The validity of these data are qualified due to increasing variability of the data near the LT-MDL at the low end of the instrument’s calibration range. In the SCPN water quality database, the numeric results are not reported; rather, the results are “present < quantitation limit”.

2.1.2 Measurement precision

Precision is the variability of the different observations or measurements (of the same thing) when compared to each other, or how repeatable observations or measurements are. The precision of laboratory analysis is monitored and documented by the contract laboratory. In addition, SCPN staff will monitor laboratory precision in measuring analytes by submitting field replicates at least once per fifteen samples. The purpose of collecting replicate samples is to identify and/or quantify the variability in all or part of the sampling and analysis system. Field replicates will consist of two different stream samples collected in succession. To evaluate measurement precision, SCPN will calculate the relative percent difference (RPD) between two replicate samples. The RPD should be within 30%.

RPD is calculated as

$$RPD = \left[\frac{S_1 - S_2}{(S_1 + S_2)/2} \right] \times 100$$

where

S_1 = the larger test result value,

S_2 = the smaller test result value

2.1.3 Measurement bias

The NWQL utilizes extensive QA/QC procedures to minimize measurement bias. These

procedures include analyzing blanks and standards.

Many method detection limits (MDL) at the NWQL are in parts-per-trillion concentrations, particularly for trace metals. As a result, it is very easy to contaminate water samples. To determine the bias caused by contamination due to equipment handling, sample collection, and transport, a field blank will be collected at least once every 15 sampling events. This type of blank is to be collected at the sample site following the sample collection steps described in SOP #6, but using inorganic blank water (ordered from NWQL) rather than stream water. In addition, a blank will be collected in the laboratory every time there is a change in equipment or method. Blanks will be analyzed for the same suite of parameters as for the native site water. If blank contamination is an issue, it may be necessary to collect different types of blanks to isolate the sample collection step responsible for the contamination. Detailed information on the many types of blanks can be found in U.S. Geological Survey (2006).

In order for sample data to be valid, ideally all results from a blank sample should be <MDL. If this is not the case, then the stream water sample results for that analyte must be questioned and, possibly not reported if the result is near the concentration found in the blank. When low-level contamination of trace elements has occurred, EPA guidance (EPA 1989) says that sample results may be considered positive if the concentration of the analyte exceeds five times the maximum amount detected in any blank. In addition, if the analyte is present in the sample(s) at less than five times the amount in the blank, then it must be treated as a non-detect, and that concentration will be the detection level for that analyte in that sample.

2.2 Water quality core parameters

Water quality core parameters include water temperature, pH, specific conductance, dissolved oxygen, turbidity and discharge. Each of these parameters are measured with a specific instrument designed for the

purpose. Selection criteria for water quality equipment should include considerations of measurement range and accuracy, data storage capacity, expected battery life, and instrument durability.

2.2.1 Measurement sensitivity

The measurement sensitivities for the Hydrolab MS-5 and HOB0 Pendant data loggers, as specified by the manufacturer, are provided in Table 10-1.

Because the field measured values of core parameters are always expected to be in the quantitative range of and well above the detection limits of the sensor, detection limits will not be reported. Irwin (2008) describes a method called Alternative Measurement Sensitivity Plus (AMS+), where measurement variability is quantified by analysis of 7 repeated measurements of an environmental sample. It includes instrument noise and natural heterogeneity. The results will be reported in NPSTORET for each site visit.

AMS+ is calculated as:

$$AMS+ = 3.708 * sd_7$$

where 3.708 = the two-sided t-value for 99% confidence for sample size 7,

sd_7 = standard deviation of 7 in-situ measurement values.

We use the AMS+ calculation, which uses the two sided t value because we are interested in both the side (distinguishing a value from zero, such as with a MDL) and the two-sided issue of sensitivity: how large of a difference between two individual values can we justify as actually being a true difference from one

another and not due to measurement process noise (Irwin 2008).

Like the AMS+, AMS is calculated in the lab during premobilization with measurements made in a calibration standard solution. The intent is to evaluate instrument noise alone without confounding influence of stream variability. AMS will be calculated during each premobilization and recorded on the premobilization datasheet.

2.2.2 Measurement precision

We can evaluate the precision of core water quality measurements obtained from a multiparameter sonde by conducting a field stabilization check. This is done by placing the probe in the stream and comparing the ranges of the fluctuations to the field stabilization criteria in Table 10-2.

If probe measurements do not consistently meet these criteria, then the sonde must be sent in for maintenance or repair and measurements must be made with a different sonde, if possible. If alternate measurements cannot be made with another sonde, the data anomalies must be noted on the datasheet and the data should not be used.

For turbidity measurements, precision will be evaluated by measuring three sub-samples from a single stream sample. The sample must be properly mixed before pouring each aliquot. Failure to do this will result in biased readings. For turbidity over 5 NTU, results should be within 10% of one another. If not, the reported result should be flagged.

2.2.3 Measurement bias

Bias when collecting multiparameter sonde

Table 10-1. Manufacturer-specified sensitivities for Hach Hydrolab MS-5 and Onset HOB0 Pendant

Core parameter	Accuracy	Resolution
Temperature (Hydrolab)	± 0.10°C	0.01°C
Temperature (Onset HOB0 pendant)	± 0.53°C	0.14° at 25°C
pH	± 0.2	0.01
Specific conductance	± 1% of reading; ± 1 µS/cm	0.1 µS/cm
Dissolved oxygen	± 0.01 mg/L for 0-8 mg/L; ± 0.02 mg/L for >8 mg/L	0.01

Table 10-2. Acceptable post-calibration ranges for core parameter measurements with multi-parameter sondes (Gibs et al 2007).

Core Parameter	Stabilization Criteria
Temperature	± 0.2 °C
pH	± 0.2 S.U. ¹
Specific conductance	± 5 μ S/cm if ≤ 100 μ S/cm or $\pm 3\%$ if ≥ 100 μ S/cm
Dissolved oxygen	± 0.3 mg/L, $\pm 3\%$ saturation

¹If drifting persists or if measuring low conductance waters (≤ 75 μ S/cm), allow ± 0.3 pH units

measurements can be avoided by frequent calibration, as described in SOPs #4 and #5 of this protocol. For pH, specific conductance, and dissolved oxygen, we will evaluate potential bias by measuring these parameters in the appropriate solution immediately after calibration. These results are to be compared to the stabilization criteria in Table 10-2. If the difference between the calibration solution standard value and measured value is outside the stabilization criteria, then the sonde should be re-calibrated or another sonde should be used.

Performance of the pH sensor will be tracked over time by recording the millivolts generated by the sensor in each standard calibration solution. These numbers will be used to calculate the slope percentage of the pH probe, as described in SOP #4. If the slope percentage is 94% or less, the pH sensor should be replaced (Ritz and Collins 2008).

Bias in temperature readings will be assessed by comparing field thermometers to a NIST-certified thermometer across a range of temperature conditions once per year, as described in SOP #4.

Bias in turbidity readings will be assessed in the laboratory by calibrating the turbidimeter as described in SOP #4 of this protocol, and obtaining readings from each calibration standard. In the field, bias will be assessed by measuring the Gelex secondary standards prior to measuring stream samples. Readings should be within $\pm 5\%$ of the values that were assigned to the Gelex standards immediately after calibration of the turbidimeter; these ranges will be printed on each data-

sheet. Readings outside of this range indicate that the turbidimeter calibration is no longer acceptable, and the instrument should be re-calibrated prior to further measurements.

Discharge is the field measurement most likely to exhibit significant bias due to a change in crew members. To minimize this bias, all technicians will be trained using the USGS On-Demand Web-stationed Training for discharge estimates (<http://pubs.usgs.gov/fs/2007/3099/>). Each crew will compare results with an experienced hydrologist or aquatic ecologist at least once after the training. In the event that measurements rated “good” are collected by both the experienced and inexperienced crew, the results should differ by 5% or less.

2.3 Bacteria

2.3.1 Measurement sensitivity

The IDEXX Quantitray-2000 can detect bacteria numbers from <1 to >2420 MPN per 100 mL sample. The IDEXX method for quantifying bacteria in a stream sample produces an estimate, or most probable number (MPN), of bacteria cells present in 100 mL of water. The test utilizes a 97-well tray where bacteria are incubated and the positive wells are counted. It is not an exact count of the number of bacteria per 100 mL because it is possible for more than one bacterial organism to be present in a well. Thus, two identical samples may yield different numbers of positive wells. Positive wells are entered into a calculator (provided by IDEXX), and results are given as an MPN plus a 95% confidence interval.

2.3.2 Measurement precision

As with the laboratory samples, SCPN will collect a replicate at least once for every 15 stream samples collected. The replicate will consist of another sample collected in the same manner as the original sample to capture variability in the stream water and in the sample handling process. Because bacteria data are variable, SCPN will follow EPA guidance (USEPA 2014) and consider a RPD of <50% to be acceptable. The RPD will not be calculated for results below 20 MPN. In three years of bacteria sample collection by SCPN, numbers have been highly variable below this threshold and RPDs well above 50% have been common.

2.3.3 Measurement bias

Bias of bacteria samples in the form of sample contamination will be assessed by collecting and processing field blanks at least once per 15 stream samples. These samples will be filled at the site with blank water purchased from IDEXX and handled in the same manner as a stream sample.

Bacteria trays are to be incubated within a narrow temperature range— $35 \pm 0.5^\circ\text{C}$. Incubation typically occurs during multi-day trips, so that the incubator may be subjected to many changes in temperature during the 24-hour incubation period. Incubator temperature will be logged during all bacteria sample incubations. Incubation duration and the time between sample collection and start of incubation will be noted on the data-sheets. Data obtained from samples whose incubation temperatures (temperature range or time period) were violated will be flagged appropriately in NPSTORET and wherever the data appear.

3 Long-term data comparability

SPCN ensures data comparability, including comparability to other agencies by following the guidelines laid out in these SOPs, which follow standard methods for data collection. However, over the life of the monitoring program, changes in sensors, instruments, methods, and staff will occur. It is important to quantify and track the effect these

changes may have on the data. This ensures that internal data comparability is achieved through the life of the monitoring program, and that any changes detected during trend analysis are due to environmental conditions rather than bias from the observer, equipment, or method. This section outlines the major types of changes that can occur during the life of a monitoring program, and how these changes can be tracked and their effects mitigated.

3.1 Personnel

Staff turnover could cause bias in the dataset due to individual differences in executing the procedures described in these SOPs. This may be avoided with thorough training of new personnel by experienced staff, which includes the yearly crew workshop described in SOP #2. Although most water quality measurements are made with electronic instruments, the exact location of a sample collection site is subjective, as is instrument selection.

3.2 Instrument replacement

Water quality equipment can break or wear out over time, requiring replacement of the instrument. Changing instruments and sensors introduces the potential for bias. When an instrument is replaced by a similar instrument (same technology) or if a sensor is changed, a minimum of 30 discrete measurements should be made on three objects or samples of known value with each instrument. Each measurement should be unique, not a series of readings recorded by a datalogger. Sonde measurements should be made on samples in a controlled environment to avoid natural variation.

3.3 Changes in technique or technology

Changes in methods may be necessary as new techniques or technologies are developed that prove to be superior to older techniques. Rather than using older methods for the sake of continuity, this monitoring program should evolve and adopt new meth-

ods. However, there must be a clear understanding of how results from older methods can be used with new methods. This requires careful tracking of changes in methods or indicators so that future data users will not misinterpret a change in trend.

When a change in method occurs, at least 30 paired measurements should be made. More measurements are necessary when indicators change because very few indicators respond in exactly the same way to environmental stimuli. The more complex the relationship between the measured value and what it is intended to represent, the more measurements will be needed to ensure that data from the two indicators or methods can be compared through time.

3.4 Change of laboratory

When a change in methods is made by the laboratory, the laboratory will be responsible for estimating any bias introduced by the change. When a change in laboratory is made, a minimum of seven samples will be split and replicates should be sent to both labs for analysis. Data from these analyses will be used to calculate a percent difference (PD) and fraction of change:

$$PD = \left(\frac{new - old}{old} \right) \times 100$$

$$\text{Fraction of change} = \frac{old}{new}$$

where

old = data from the older or original method

new = data from the new or replacement method.

Ideally, the average PD (two-sided 95% t-distribution confidence interval) will be within 20% of the mean of the old or the new method measurements. If this is not the case, the number of overlapping measures should be increased until this criterion is met (Irwin 2008).

3.5 Routine bias

In addition to the types of bias determination utilized when a change of methods, instruments, or observers occurs, SCPN will also use instrument calibration data to calculate long-term “running” averages for instrument bias. These calibrations will be done routinely on all instruments that can be calibrated against known standards. The individual bias data calculations and long-term statistics will be automatically generated using NPSTORET or another database. The goal will be to keep long-term bias for an individual parameter smaller than the USGS acceptance/rejection criterion in SOP #5, Table 8. The bias should be compared against manufacturer ratings and concern should be raised if bias is significantly higher than the rating. Ideally, the bias should be smaller than the manufacturer’s rated precision, but this is often not the case, for various (often unknown) reasons. It could be a result of internal damage or a small crack in a sensor, but it could often result from improper cleaning or maintenance. Full inspection, maintenance and cleaning should be performed before sending any instruments out for repairs (see SOP #4).

4 Supplies and equipment selection criteria

To ensure that the quality of water quality measurements and analytical lab results are consistent over time and are comparable to other projects, SCPN will use supplies and equipment that meets established guidelines, particularly those established by USGS. Guidance for selecting supplies and equipment for water sample collection, such as filters, blank water, and tubing are described in Lane et al. (2003). Guidance for selecting current meters are described in Turnipseed and Sauer (2010). All of the meters used by SCPN are described in this document except for the Hach FH950, which is a newer meter. For making measurements at USGS gaging stations, only USGS-approved meters listed in Turnipseed and Sauer (2010) will be used in order to comply with that agency’s requirements.

Specifications for sensor-calibration solutions and multiparameter sondes, including accuracies and ranges of pH, temperature, specific conductance, and dissolved oxygen sensors are described in Gibbs et al. (2012). Also described are general care and storage of multiparameter sondes. Descriptions of various types of turbidity meters and calibration standards are in Anderson (2005), and choosing a specific meter largely depends on the anticipated turbidity range. SCPN uses a Hach 2100P, which is commonly used in water monitoring projects, mentioned in Anderson (2005), and is capable of measuring turbidities typical of SCPN streams.

5 Revision history

Provide the following information for each revision made to SOP 10.

New version:

SCPN_Water_Quality_SOP10_YYYYMMDD

Date:

Author:

Sections affected:

Changes & reasons: