

Springs Monitoring Protocol Implementation Plan for Park Units in the Southern Colorado Plateau Network

Standard Operating Procedure (SOP) # 12

Data summary, analysis and reporting

Version SCPN_Springs_SOP12_20180930

Revision History Log:

Prev. Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #

This SOP is adopted from the Northern Colorado Plateau Network's SOP #10 Data summary, analysis, and reporting, Version 1.00 (June 2016) in Weissinger, R., M. Moran, S. Monroe, and H. Thomas. 2017. Springs and seeps monitoring protocol for park units in the Northern Colorado Plateau Network, Version 1.00. Natural Resource Report NPS/NCPN/NRR—2017/1439. National Park Service, Fort Collins, Colorado.

This SOP provides an overview of data summary, analysis and reporting procedures. Raw data are compared to previously collected data to catch misidentified species and values that represent unlikely amounts of change between years. Once data quality issues are resolved, analyses can be conducted. Two general types of analyses are performed: 1) summary reporting of the most recent monitoring observations (summary reports and annual briefs), and 2) comprehensive assessments of the long term data set (power to detect trend of the overall sampling design, evaluation of trends, and inclusion of other parameters that may explain variation or trends over time).

1.0 Data Summaries

Data should be summarized for the following variables:

Vital Sign	Reporting Variables (reporting unit)	Test for Long-Term Trends
Water Quantity	Flow (nearest 0.01 L/min) Stage (nearest 0.01 m)	Seasonal Kendall tau
Water Quality	pH (nearest 0.1) specific conductance (nearest 10 μ S/cm) water temperature (nearest 1° C)	Seasonal Kendall tau
Vegetation Species Composition	Endemic plant counts (hanging gardens only) Native species % cover and richness	General Additive Model

	Exotic species % cover and richness Wetland species % cover and richness Wetland: Upland cover ratio	
--	--	--

Water Quantity and Water Quality

Summarize flow, stage, pH, water temperature, and specific conductance annually (March-November) and seasonally (March-May, June-August, September-November) for each sentinel spring. Present the number of samples, median, range, and standard deviation of measurements. Document missing samples and indicate the reason. Graphical representations such as values over time and boxplots by month are also helpful ways of presenting data for each site.

Vegetation Data

Endemic plant counts

Summarize by species for each hanging garden and each hanging garden's geomorphic surfaces. For data summaries, count code data may be presented. For data analysis and graphing, numeric substitutions are necessary.

Count Code	Definition	Substitution
N	Species not seen and presumed absent.	0
1-500	Count data	none
A	501-1000 individuals	750
B	1001-10,000 individuals	5000
C	> 10,000 individuals	10000

Species richness

Species richness is the number of unique species encountered. For each spring, calculate richness of all species, nonnative species, species by life form (e.g., tree, shrub, etc.), and species by wetland status group (obligate, facultative wet, facultative upland, upland). Species richness data should be included in site summaries but still need to be evaluated for their power to detect trends.

Vegetation cover

Percent cover is calculated for each species, native and exotic, and wetland status group for each geomorphic surface by summing the mid-points of the cover class recorded for each species. A wetland to upland cover ratio is calculated by summing obligate and facultative wet cover and dividing by the sum of upland and facultative upland cover. Vegetation cover data should be included in site summaries but still need to be evaluated for their power to detect trends.

Cover Class	Definition	Mid-Point
1	<1% canopy cover	0.05
2	1 - <10% canopy cover	5
3	10 - < 25% canopy cover	12.5
4	25 - < 50% canopy cover	37.5
5	50 - < 95% canopy cover	72.5

6	95% or greater canopy cover	97.5
---	-----------------------------	------

2.0 Data Quality Assessment

Before data analysis and reporting, data should be additionally evaluated for reasonable levels of change based on historical conditions at each spring. For instance, a measure may appear to be valid based on the database domain, but represent an unreasonable (and unlikely) change in conditions since the last field visit.

Summarized attributes should be compared between years, and all suspect values (large increases or decreases) should be flagged and resolved before the data are analyzed. In addition, data should be evaluated for the omission of historically dominant species or the appearance of dominant species not detected in previous years. New occurrences of a species should be compared with the park-level species list and species-range limits.

Suspect measures or species should be evaluated by field crew members and by field verification if possible. Any corrections made to the database should be recorded in the QAQC log. Suspect data that are not resolved should also be flagged in the QAQC log so that it can be excluded from analyses, if appropriate. Suspect data that cannot be resolved in the office should be brought to the attention of the field crew at the next monitoring visit so that they may eventually be resolved.

3.0 Common issues with time series data

Representativeness

The target population for hydrologic data is base spring-flow at each site. If data are sampled during a flood event, they may be used for data summaries but should be excluded from the dataset prior to trend analyses.

Outliers

Data that are more than 2 standard deviations from the mean are considered outliers. Check each outlier for data transcription or human errors and for comments on the site visit that may explain unusual conditions. Unless the data are found to be erroneous, retain the data in the dataset and use analysis methods that are robust to outliers (e.g., ranking, non-parametric, etc.).

Missing data

Missing samples should be avoided as much as possible but are inevitable in a monitoring program. Missed samples become less important over time, and data should not be estimated. For data above or below detection limits (e.g., if flow is unmeasurable, but present), do not use substitutions. Use analysis methods that are robust to uncertain data, e.g., Kendall tau, etc., see Helsel and Hirsch (2002).

Serial correlation

Repeated measures over time can often be serially correlated, introducing bias into analyses. Model residuals should be tested for autocorrelation. If autocorrelation is present, it can be specifically accounted for in mixed models by including a temporal correlation structure within the model (see, for example, Zuur et al. 2009). A seasonal Kendall test that accounts for serial dependence is also available, although it is only robust with large sample sizes (Hirsch and Slack 1984).

3.0 Summary Reports and Annual Briefs

Summary Reports

The primary objectives of the summary, or “status”, report are to 1) document monitoring activities for the period of record, 2) summarize data and describe the current condition of resources, and 3) provide information to park managers in a timely manner that increases data utility and communication. In addition to tabular and/or graphical summaries of monitoring data, the report should include important assessment points of management interest.

Assessment Points

- Presence of target exotic plants or animals
- Novel disturbance – e.g., trespass cattle, increased social trailing, etc.
- Shift in pH or conductivity indicating possible aquifer contamination
- Repeated presence of *E. coli* indicating possible fecal contamination

4.0 Comprehensive Assessments

Comprehensive analyses and reports are completed periodically on the long term data set. Comprehensive reports should be structured like typical scientific manuscripts (introduction, methods, results, discussion, and literature cited sections) and may include any of the analyses described in this section.

Sampling-Design Power to Detect Trends

The current sampling design was developed using legacy and pilot data that may not accurately reflect the monitoring data over time. The sampling design should be checked periodically to ensure that the assumptions are being met and to explore improvements that may better meet the original monitoring and sampling objectives. Sampling designs are evaluated by calculating variance and power to detect trends of varying magnitudes. For the springs monitoring program, the desired power is a better than 80% likelihood of not obtaining a false negative, i.e., not detecting a trend when in fact, a trend exists.

Variance estimates are derived from all years of relevant monitoring data, not just observations from the current field season. Changes in methods that would affect variance (e.g., more precise methods, new instruments, etc.) may necessitate a new analysis period. To achieve the desired power, alternative survey designs can be evaluated to determine appropriate enhancements, such as an increase in sample size or more frequent visits to established springs, to achieve the original sampling objective. Power-to-detect-trend is evaluated with alternative numbers of panels, sample sizes, and re-visit schemes. A trend may be reported as a yearly rate or a rate over a set number of years. This conversion is outlined for increasing (eq. 1, Table 2) and decreasing (eq. 2, Table 2) rates of trend. Results are then used to determine which survey designs best meet the original sampling objectives.

Table 2. Formula for converting a yearly rate of trend to a rate of trend over a set number of years.

Conversion for rate of trend	
$I = e^{rt}$	(1)
$D = 1 - e^{rt}$	(2)
where;	
I = % increase over t years	
D = % decrease over t years	
r = the yearly rate of increase or decrease	
t = the total time period of increase or decrease	

Site-based water quantity and water quality data

The current monitoring design for water quantity data is estimated to detect a 5% annual change in 6 years of sampling (Weissinger and Moran 2013). Irvine et al. (2012) developed a power analysis for water quality data based on repeated iterations of the seasonal Kendall tau test using pilot data representative of future sampling conditions. Data are deseasonalized by creating residuals from the monthly median, then the residuals are randomized within months and across years. Multiple permutations of the tests are compared to develop an estimate of the power to detect trend with varying sampling scenarios and time periods. This approach is appropriate for future power analyses of monthly water quality and water quantity data from this program.

Population-based vegetation data

A preliminary analysis of power to detect trends in endemic species used simulated data for *Aquilegia micrantha* taken from a double observer pilot study. Variability documented within the pilot study was used to simulate a dataset with the proposed sampling design's revisit design. For a sample size of 10 springs per year, it was estimated that a 10% change in *Aquilegia micrantha* could be detected over 20 years using a Poisson regression model (Tipton 2013).

The current sampling design must be tested for each species and vegetation metric after adequate data have been collected; we recommend this analysis occur after 6 years of data collection. If methods for power analysis using General Additive Models (GAMs; see trend assessment section below) have been developed, that would be the appropriate assessment for these types of data. In the absence of this type of analysis, the Poisson regression model used by Tipton (2013) can be used.

Trend Assessments

Long-term trend analyses are performed after at least 6 years of monitoring data are collected. Data are analyzed for the period of record, and for any period where 1) a change is expected to have occurred due to a known impact at or up-gradient of the spring or 2) a change is apparent after a graphical inspection of the data record. To more conservatively inform managers of potential changes in springs ecosystems, data will be evaluated at a statistical significance of $p \leq 0.10$.

When significant trends are detected, reasons for these trends should be evaluated by literature reviews, comparison with similar data sets, and additional assessments using ancillary information. When unidirectional trends are not detected, correlations may help us understand system drivers. Local climate records and remotely-sensed vegetation data can provide valuable context for springs data.

Frequentist Methods

Site-based analyses

For site-specific analyses, such as water quantity and water quality data from sentinel sites, Helsel and Hirsch (2002) cover various data analysis scenarios based on the distribution of the data (parametric, nonparametric, or mixed). Monotonic trend tests look for gradual, uni-directional changes over time, while step trends evaluate before and after a change at a specific point in time. The monotonic trend hypothesis is more commonly used for general monitoring unless there is a reason to test for a step trend.

Monthly water quantity and water quality data in this monitoring program have been designed to be tested for long-term trends using the seasonal Kendall tau, a non-parametric test for monotonic trends in data that have a strong seasonal pattern, such as spring-flow. The test accounts for seasonal variations by comparing ranks of data from the same recurring time intervals; for example, in a monthly sampling scenario, March values are compared only to other March values, July values to July values, and so forth. Deseasonalizing is done by subtracting seasonal medians from each of the values to be regressed. A Mann-Kendall test for trend is done for each season, with the Seasonal Kendall test statistic being the sum of the several Mann-Kendall test statistics.

The data can also be fitted with a non-parametric equivalent of a regression line called a Sen slope, which is calculated as the median annual slope of all possible pairs for a season. This slope corresponds to the estimated median annual change in the parameter, and the percent annual change can be calculated as $(\text{median annual Sen slope} / \text{median annual value}) \times 100$.

For step trends, in which a sudden jump in the data is expected or observed to have occurred, the non-parametric Seasonal Rank Sum test is likely to be an appropriate test for springs water quantity and water quality data. The Rank Sum test (also known as the Mann-Whitney “U” test or Wilcoxon Rank Sum test) tests the assumption that two populations (before and after a change) have the same distribution. The Seasonal Rank Sum test computes the rank-sum statistic separately for each season, sums the test statistics, and then evaluates the overall summed test statistic.

Population-based analyses

Hanging garden vegetation data from this monitoring program are designed to be tested for long-term trends in populations using mixed-effects models (Larsen et al. 1995). The Waterpocket Fold Navajo Sandstone aquifer at Glen Canyon National Recreation Area is the initial population in the sampling design.

General Additive Models (GAMs) are a non-parametric extension of Generalized Linear Models that have no a priori assumptions about the shape of the response (i.e., linear, quadratic, etc.; Wood 2006). Given the still preliminary nature of the vegetation data being collected by this program, GAM analysis is an appropriate first step for determining whether simpler linear models can be used for each population. Because smoothed functions do not have simple interpretable parameters to report, when the comparison of GAMs indicates that linear responses are adequate, estimate parameters and the confidence intervals of those estimates via conventional linear models.

Given differences among springs in spring size, aspect, and habitat availability, spring identity is treated as a categorical (factor) fixed effect. We have no a priori reason to expect linear trends or linear responses to predictors. Therefore, each model should be fitted with and without smoothing via penalized smoothing using the gam function. For example, to test for temporal trends in vegetation data for each spring, accounting for an interaction between time and the individual spring, fit both linear (Equation 1) and smoothed (Equation 2) models of the forms:

$$\text{Equation 1: } \textit{Vegetation} \sim \textit{Spring} + \textit{Date} + \textit{Spring:Date}$$

$$\text{Equation 2: } \textit{Vegetation} \sim \textit{Spring} + s(\textit{Date}, \textit{by}=\textit{Spring})$$

The degree of smoothing in the second model is selected using generalized crossvalidation. Models can be compared using the Akaike Information Criterion (AIC). Generally the smallest AIC is considered the best fit model, but models with $\Delta\text{AIC} < 4$ are roughly equivalent (Burnham and Anderson 2002).

The output from this model is a measure of the intercept and slope, the significance of the slope and intercept being different from zero, and measures of the variance and covariance of the regression parameters. A significant slope indicates a significant trend at the alpha level used to assess significance or below.

Bayesian Methods

Bayesian methods for trend assessment identify the probability of different slope-parameter values (Gelman et al. 2000, Wade 2000) given the observed data. This differs from frequentist methods which calculate the probability of observing data given a specific value for a parameter, such as the null hypothesis. Assessing the probability of different slope values is potentially important for various reasons. First, sampling designs are largely based on a few key features. The minimum detectable trend at a reasonable power level for many of the other responses will likely be much higher. Estimating the probability of slope values may be the best way to provide insight into trends for these attributes. Second, even for the key features around which sampling designs were based, Bayesian methods may reveal the potential of declining trends before rejecting the null hypothesis that the regression slope equals 0 with conventional statistics.

Bayesian methods use a likelihood distribution generated from the sample data that provides the probability of observing the data for every possible parameter value. Combining this with a prior distribution produces a posterior distribution which is used to make inference about parameters. The prior distribution reflects previous knowledge about the distribution of parameters which can be developed as the monitoring program progresses (Efron 2013). Once an informed prior is in place, the probability of various increasing and decreasing slopes should be evaluated based on the posterior distribution and the consequences of not alerting management of a potential trend. The result of this assessment is the derivation of the probability of different levels of decline/increase. Even when conventional statistics do not reveal a significant trend, relatively high probabilities (e.g., ≥ 0.70) of trends with an informed Bayesian model may warrant special attention.

Including Explanatory Variables

Model-based methods

Model-selection procedures consider a range of potential factors influencing directional change in resource conditions. Even if a significant trend occurs as a function of time, the relationship between other explanatory variables and spring measures should be evaluated to determine potential ‘causal’ factors of trends. These other explanatory variables can include biophysical features of the springs (e.g., aspect, potential solar radiation, etc.) as well as other measures (e.g., climatic conditions).

Model selection methods (Burnham and Anderson 2002) evaluate the explanatory power of different models of trends in spring measures. Multiple models are considered, each with different combinations of independent variables and interaction terms. All models should be based on hypothesized, ecologically meaningful relationships. Adjusted AIC (Akaike’s Information Criterion) or Bayesian Information Criterion (BIC) generated with a mixed-effects model are used to determine the best predictive model as well as the importance of independent variables; the latter penalizes additional parameters more than AIC does.

The primary model in this model-selection procedure is the response of interest vs. time (see frequentist methods above). Alternative models can be based on measures of climate, land use, and land condition, as well as other measures that may be considered

important. When evaluating other variables, they are either included in an interaction with time or time is replaced with the variable.

The result of the model-selection procedure is the selection of the model (i.e., set of independent variables) which best predicts the trend of the response variable. This, in turn, aids in explaining reasons for the observed trend. For instance, if alternative models with one or more driver attributes (e.g., drought days) accounted for more variance than just using time, trends may be attributed to these attributes in a correlational relationship, but not necessarily a causal relationship. If alternatives to time account for less variance, other avenues of investigation should be considered to explain reasons for observed trends.

Correlations

There will be instances where a monitored attribute fails to exhibit a trend (i.e., a direction change), but instead may have periods of decline followed by recovery. In other words, natural variation may be more evident than a precipitous decline or increase, and it may mirror the variation in a key driver attribute. Discerning this variation and reasons for it can involve many complex procedures. As a first step, however, correlational methods can provide useful insight and are relatively simple.

For attributes that fluctuate between declines and increases, correlations with climate, land cover/use, land condition, and other potential drivers should be considered. Strong correlations between a response and driver may be sufficient to understand reasons for the observed variation over time (e.g., strong correlation between flow and winter precipitation).

Another type of correlation assessment is path analysis (Lindquist 2000). This method accounts for the inter-correlated structure among factors by decomposing correlations to direct and indirect effects. For instance, changes in flow in an alluvial spring may be influenced by precipitation alone or by vegetation cover alone or by the influence of precipitation on vegetation cover. A path analysis would decompose the effects of climate and of vegetation cover on flow, while considering the combined influence of climate and vegetation. The result of a path analysis is the indication of the strength of direct and indirect factors on a response, thus providing insight into reasons for the observed changes in a response. Path analysis is based on correlation values and the simultaneous solution of structural equations. Simple examples of the construction and use of path analysis are provided in Lindquist (2000) and Van Bruggen and Arneson (1986); details of path analysis are provided in Pugese et al. (2003).

5.0 References

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: A practical Information Theoretic approach. Second edition. Springer, New York, NY, USA.

Efron, B. 2013. Bayes' theorem in the 21st century. Science 340: 1177-1178.

- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2000. Bayesian Data Analysis. Chapman & Hall/CRC, New York.
- Helsel, D.R. and R.M. Hirsch. 2002. Statistical methods in water resources. Ch. A3 in Techniques of Water-Resources Investigations of the United States Geological Survey, Book 4, Hydrologic Analysis and Interpretation.
<http://water.usgs.gov/pubs/twri/twri4a3/>
- Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 20: 727-732.
- Irvine, K. M., K. Manlove, and C. Hollimon. 2012. Power analysis and trend detection for water quality monitoring data: An application for the Greater Yellowstone Inventory and Monitoring Network. Natural Resource Report NPS/GRYN/NRR—2012/556. National Park Service, Fort Collins, Colorado.
- Larsen, D. P., N. S. Urquhart, and D. L. Kugler. 1995. Regional scale trend monitoring of indicators of trophic condition of lakes. Water Resources Bulletin, 31:117-140.
- Lindquist, J. E. 2000. A method of estimating direct and indirect effects of Armillaria root disease and other small-scale forest disturbances on canopy gap size. Forest Science, 46:356-362.
- Pugesek, B. H., A. Tomer, and A. von Eye. 2003. Structural equation modeling. Cambridge University Press, Cambridge, UK.
- Tipton, J. 2013. Rare plant power analysis for the National Park Service at Arches, Canyonlands, and Natural Bridges. Internal report to NPS, Moab, UT.
- Van Bruggen, A. H. C. and P. A. Arneson. 1986. Path coefficient analysis of effects of *Rhizoctonia solani* on growth and development of dry beans. Phytopathology, 76:874-878.
- Wade, P. R. 2000. Bayesian methods in conservation biology. Conservation Biology, 14:1308-1316.
- Weissinger, R., and M. Moran. 2013. Proposed sampling design for springs, seeps, and hanging gardens monitoring in the Northern Colorado Plateau Network. Natural Resource Report NPS/NCPN/NRR—2013/618. National Park Service, Fort Collins, Colorado.
- Wood, S. N. 2006. Generalized additive models: an introduction with R, volume 66. CRC Press, Boca Raton, Florida, USA.
- Zuur, A.F., E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, NY.