

Cumberland Piedmont Network

National Park Service
Inventory and Monitoring Division



Standard Operating Procedure 6

Data Management and Quality Assurance / Quality Control—Version 2.0

Please cite this as:

Cumberland Piedmont Network. 2024. Ozone and Foliar Injury Monitoring. Data management and quality assurance/quality control—Version 2.0. Standard Operating Procedure. NPS/CUPN/SOP—6. Cumberland Piedmont Network, Mammoth Cave, Kentucky.

Summary

This standard operating procedure (SOP) describes tasks to be performed for data management including data entry, quality assurance/quality control (QA/QC), and general stewardship. The crew leader oversees field data collection and reviews completed field forms, but it is the responsibility of the project lead to ensure that all data management activities have been completed (data have been entered from field forms, GIS data have been processed from GPS units, voucher specimen data have been processed, etc.). Sections of this SOP were adapted from SOP 9 *Data Management and Quality Assurance/Quality Control—Version 1.0* of the Gulf Coast Network Monitoring Vegetation Protocol Implementation Plan (Carson et al. 2018).

Revision Log

Revision Date	Author	Changes Made	Reason for Change	New Version #
June 2020	Teresa Leibfreid	Added info about database development, data management overview and timeline, QA/QC procedures, and recent I&M guidance	The original protocol was published prior to completion of the database and prior to I&M QA/QC guidance	2.0

Introduction

There are three primary data components to manage in foliar injury monitoring:

1. **Foliar injury data:** The ozone foliar injury database is a customized Microsoft Access application. The current table structure is based on the NPS Natural Resource Database Template (NRDT; NPS 2007). The core structure includes a Locations table that holds locational information about each site, which is linked to an Events table containing information about each sampling event. The project-specific tables and forms needed for foliar injury were developed in 2013 through a cooperative agreement with the University of North Carolina and are based upon field forms from the US Forest Service Phase 3 Field Guide – Section 9 (USFS 2008). Hardcopy field forms are scanned after all necessary reviews and stored on the Cumberland Piedmont Network (CUPN) server. Site sketches and digital photos are also stored on the server with links established in the database.
2. **GIS data:** The spatial component (site locations) of foliar injury monitoring is managed in ArcGIS following NPS-GIS metadata standards. The centroid points of site locations are entered into ArcGIS using information from hardcopy field forms and data from GPS units. Since monitoring sites are now established, GIS layers will only be updated if new sites are added, or old sites dropped.
3. **Specimen and Archival Data:** Leaf specimens, photographic vouchers, and archival material will be managed following NPS museum collection guidelines and will be catalogued in the Museum Collection Management System (MCMS). For the ozone monitoring project, specimens consist of leaves from plants suspected to have foliar injury that are collected by the field crew. Specimens are processed annually, after each field season (see SOP 5 *Post-Sampling—Version 2.0* [CUPN 2024a]). The archives include all other materials related to the project such as field forms, site photographs, and maps. Archives are processed after the completion of the park rotation schedule, every six years. For more information, see the CUPN In-House document: Natural History Collections (which is available on the CUPN internal server or by request to the network).

Data Management Overview and Timeline

Project information management is an ongoing or cyclic process. The stages of this cycle are briefly summarized in the following list. Additional details on data stewardship are provided in subsequent sections of this SOP.

1. **Data Acquisition:** Data are entered onto paper field data forms during field monitoring which occurs in July and August. This step also includes collection of specimens with suspected injury and within-season duplication of data acquisition at select sites for QA/QC purposes. For detailed information on data acquisition procedures see SOP 4 *Sampling—Version 2.0* (CUPN 2024b).

2. **Field Data Forms:** Reviewing and initialing the field form for legibility, completeness, and accuracy prior to departing the field site is an important part of the QA/QC process. As soon as possible after returning to the office and prior to data entry, the project lead also checks each field form for realistic entries, correct spelling, and current taxonomic names. These checks result in validated paper field forms. The field forms are scanned after results from the foliar injury expert are entered and stored in the appropriate folder on the network server. The scans of site sketches are also saved as separate files and posted to the proper folder for linkage to the database. For naming conventions and folder locations, see section below on Naming Conventions and Version Control.
3. **Data Entry and Photo Processing:** Data are entered into the ozone database from the reviewed field forms. See *CUPN In-House Document: Ozone Foliar Injury Data Entry* (available on the CUPN internal server or by request to the network) for instructions on entering data into the database. The timing of data entry depends upon whether suspect specimens were collected. Specimens with suspected injury are sent to a foliar injury expert for confirmation. After receipt of the results (which may take a few months), data are entered into the database. If no specimens are collected with suspected injury, data entry will proceed within two months after fieldwork. Site photos are also downloaded, renamed, and posted to the ozone folder for linkage to the database. For naming conventions and folder locations, see section below on Naming Conventions and Version Control.
4. **Database Verification:** Electronic data are compared to paper field forms during data entry as an initial step in verifying the electronic dataset. The data entry person reviews the record for completeness and accuracy before moving onto the next record. The project lead conducts the verification step on 100% of ozone records, followed by a second person checking a 10% subset. Errors will be corrected when found during data entry or verification. Notation of corrections will be made on the field forms using a unique color, with the date and name of person making the correction. A record of errors will also be maintained and discussed as part of the protocol's annual operational review.
5. **Database Validation:** Although data may be correctly transcribed from the original field forms, their values may not always be accurate or logical. The logic and accuracy of a verified database is tested through database validation. For example, the amount and severity of plant injury is entered using integer values (0 to 5) to indicate scale. If a "no injury" or "0" code is entered for "amount," then the severity code must also be "no injury" or "0." As often as possible, components of the data validation process have been built into the database (e.g., value range restrictions, and species pick-lists). Although many validation problems can be addressed by a well-designed database, the project lead must also review and validate the data after verification is complete. The project lead is the authority on the methodologies and species involved.
6. **Data Quality Review:** After data are verified and validated, the next step involves analyzing data from repeat sampling events to document observer variability and compare the values obtained to the protocol's data quality standards. The data quality standards for foliar injury

data are partly based upon the measurement quality objectives used by the US Forest Service (2008). There are four components. The first two involve comparing values for amount and severity on the first ten bioindicator plants using a random starting point as described in SOP 4 (CUPN 2024b). The third component involves comparing the species evaluated during one sampling event per season. These first three components are evaluated on the designated QA/QC site to calibrate the teams, especially the standard team. The fourth component involves evaluation of all suspected injury by an outside expert on foliar injury and happens on every park site where injury is found. See Table 1 for a summary.

Table 1. Data Quality Standards and Implementation (see SOP 4 [CUPN 2024b] for more details).

Foliar Injury Metric	Value	Data Quality Standards	Implementation	Justification
Amount	0–5	± 0 classes, at least 50% of the time (USFS 2008).	Measured by both teams on the first ten plants using a random starting point at the designated QA/QC site.	Teams' evaluations should be the same, except in marginal cases, which could happen around half of the time.
Severity	0–5	± 0 classes, at least 50% of the time (USFS 2008).	Measured by both teams on the first ten plants using a random starting point at the designated QA/QC site.	Teams' evaluations should be the same, except in marginal cases, which could happen around half of the time.
Species Detection	Varies	If one team finds ten or more plants of a species, then the other team also finds that species 100% of the time. For species detected by both teams, the absolute difference between plant counts can be no greater than twenty-five 100% of the time.	Measured by both teams for the annual sampling event at the designated QA/QC site. Selected plants can vary between the two teams.	If large numbers (more than 10 which is one third of the total plants that might be examined) of plants of a species are present at a site, then both teams should detect them. If one team detects many plants of a species, then the other team should detect more than just a few.
Injury	Yes/No	The Foliar Injury Expert has appropriate experience in the diagnosis of ozone induced injury. The Foliar Injury Expert examines all voucher specimens.	All submitted samples with suspected injury are evaluated by an expert in the diagnosis of foliar injury. The Foliar Injury Expert confirms, at the site/species level, ozone injury is present.	Ozone injury is correctly identified by samplers in the field (and is not confused with mimicking symptoms). The Foliar Injury Expert examination helps to protect against false positive errors.

Evaluation of whether field QA/QC targets are met is accomplished with Excel spreadsheets that were developed by the project lead. An example of a filled-in versions of the spreadsheet is shown in Figure 1. The spreadsheets evaluate whether the QA/QC goals for the amount index, the severity index, and the number of species detected were met. The user simply

enters the plant species, plant numbers and the amount/severity indices for both the primary and QA/QC observer in the indicated cells. Whether the QA/QC goals are met are then automatically generated in the spreadsheets cells.

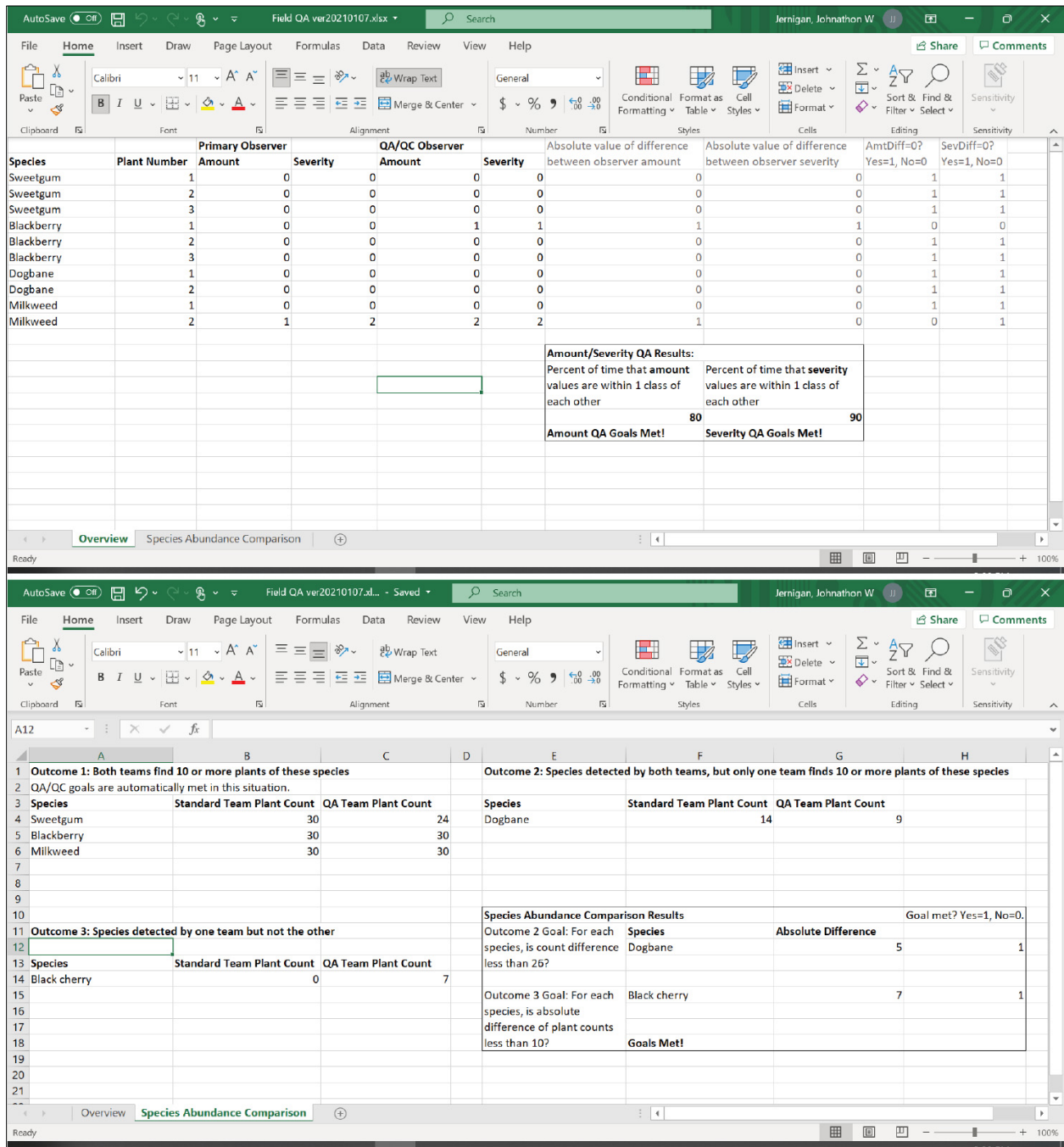


Figure 1. Example of completed field QA/QC evaluation spreadsheets.

Following guidance from the Inventory and Monitoring Division (IMD; NPS 2016), data will be published with the results of the data quality review. Data that have undergone all QA/QC procedures as defined in this protocol (and associated SOPs) are considered “accepted” and

are then available for reporting/distribution (see SOP 7 *Data Analysis and Reporting—Version 2.0* [CUPN 2024c]).

7. **Data Documentation and Distribution:** Final data that have been verified, validated, and accepted, are documented with metadata and made available to the public. On an annual basis, the data products are posted to IRMA (Integrated Resources Management Applications), which is the NPS clearinghouse for natural resource data and metadata, at <https://irma.nps.gov>. The network data manager has created an [Ozone Damage Monitoring reference](https://irma.nps.gov/DataStore/Reference/Profile/2302022) (<https://irma.nps.gov/DataStore/Reference/Profile/2302022>) to track database uploads to IRMA over time. Refer to the IRMA website for upload and product linking instructions.

Data Stewardship

Good data stewardship practices play a critical role in long term data security and help prevent irretrievable loss of information.

Data Updates

Data sets are rarely static. They often change through additions, corrections, and improvements made following the archival of a data set. There are three main caveats to this process:

- Only make changes that improve or update the data while maintaining data integrity.
- Be prepared to recover from mistakes made during editing.
- Document changes.

Any editing of archived data is accomplished jointly by the project lead and data manager. Every change must be documented and accompanied by an explanation that includes pre- and post-edit data descriptions.

Naming Convention and Version Control

Version control allows project staff to manage changes made to data sets or documents and provides an auditable trail of individual versions. To safeguard against loss or damage to data, changes or edits are not started without having the ability to back up and start over. In its simplest form, version control is saving a copy of a file before beginning work on it and using an established naming convention to do so.

This project utilizes a front-end/back-end Microsoft Access application to manage data. The front-end contains the forms, queries, reports, etc. for interacting with the data which are stored in the back-end application or file. The working version of these files are stored in the following folder: Ozone\Data\Working_Data\Database

The recommended naming convention for both files is to begin with the network acronym, followed by an underbar (“_”) followed by the protocol name (e.g., CUPN_Ozone). For the front-end

application (i.e., the user interface) this is followed by another underbar and the version number. The back-end file (i.e., the one containing data) includes the date of the latest changes made to the data using a YYYYMMDD format. Thus, the recommended naming convention for this protocol would look like the following:

Frontend: *CUPN_Ozone_v1.7* and Backend: *CUPN_Ozone_20190920*

Thus, based on the names of these files the user can discern that these are indeed the Cumberland Piedmont Network's Ozone protocol applications. The front-end is version 1.7 and the data were last updated on September 20, 2019.

As noted above, any data editing/entry session begins by making a backup of the current back-end data file. This is a simple task that is accomplished by copying and pasting the file and adding “_backup” to the end of the file name. For example, if entering data to the above file on November 25, 2019, the following data files would be in your folder:

CUPN_Ozone_20191125 and *CUPN_Ozone_20190920_backup*

Maintain an edit log to keep track of what changes were made at each step (i.e., backup file) during data processing. This makes it easier to know which backup file to select, if for some reason you have to “start over” at some point in the process. For suggestions of what information should go into this edit log, consult with the data manager.

Data entry or editing is never done in a file labeled backup (or in a file with “Archive” in its name). File naming and proper version control practices are extremely important to the long-term integrity of the data. It can also save you and the data manager a significant amount of time and frustration. So, if any questions arise regarding this process, ask the data manager.

In addition to database naming conventions, field forms, site photos and sketches follow a standard format. The naming conventions and folder locations are summarized in Table 2.

Table 2. Naming Conventions and Folder Locations using ABLI Site 0001.

Folder	Folder Location	File Naming Convention
Database Front-end	Ozone\Data\Working_Data\Database	<i>CUPN_Ozone_v1.7.mdb</i>
Working Database Back-end	Ozone\Data\Working_Data\Database	<i>CUPN_Ozone_20190920.mdb</i>
Master Database Back-end	Ozone\Data\Validated_Data\Database	<i>CUPN_Ozone_ARCHIVE_20191231.mdb</i>
Field Forms	Ozone\Data\Working_Data\Scanned_Field_Forms\Foliar_Injury\ABLI	<i>ABLI0001_FOIN_20190814.pdf</i>
Site Sketches	Ozone\Images\ABLI\Foliar_Injury\ABLI0001\site_sketch	<i>ABLI0001_20190814_SK.jpg</i>
Site Photos	Ozone\Images\ABLI\Foliar_Injury\ABLI0001\site_photo	<i>ABLI0001_20190814_01.jpg</i>

It is also recommended that reports and presentations be appended with the date at the end of the file name. This will aid other users by making it clear which file is the last report or the most recent presentation given.

Directory Structure

The various databases, images, reports, and GIS coverages used and generated by the monitoring program create a large number of files and folders to manage. Poor file organization will lead to confusion and potentially to data corruption. In all cases it is recommended that folder naming conventions adhere to the following:

- No spaces or special characters in the folder name.
- Use the underbar (“_”) character to separate words in folder names.
- Limit folder names to 20 characters or fewer.
- Format dates as YYYYMMDD (for better sorting).

The Cumberland Piedmont Network has established a template directory structure for each vital sign monitoring project to serve as a basic starting point and provide consistency across projects. Figure 2 depicts the file organization structure for a typical monitoring project. A similar directory structure contains archived files. Both the working and archive directories are accessible via the Local Area Network (LAN) at Mammoth Cave National Park (MACA) with access permissions maintained by the data manager and MACA information technology (IT) staff.

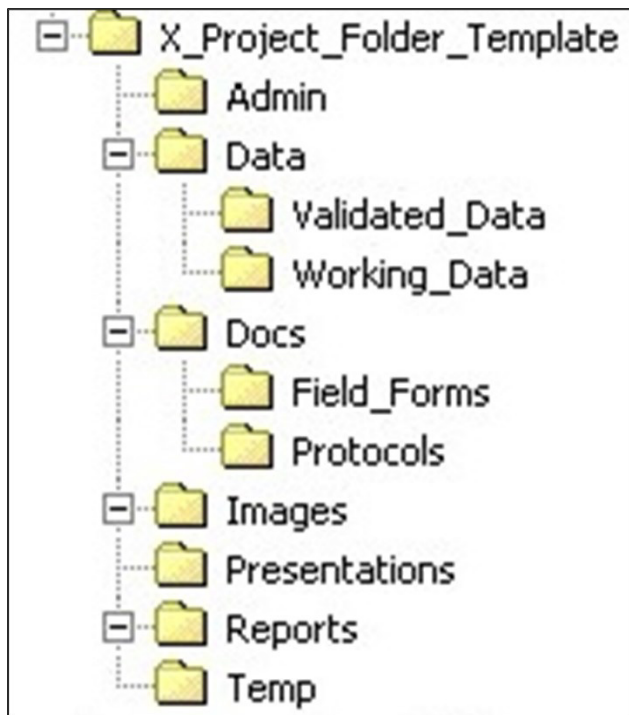


Figure 2. Screenshot of project folder structure for the Cumberland Piedmont Network.

Without due diligence file folders will become bloated over time with draft documents, multiple copies of the same files. To keep things clean, the project lead will review the project folders on an annual basis. The project lead will consult with other staff (as needed) before removing files created by others or with unknown origin. A good time for this review is at the end of each season once annual reporting requirements have been completed and disseminated.

Data Backups

Backups of all project data stored on MACA's server in CUPN's shared folders are regularly conducted. A full backup is conducted on the last Friday of each month with incremental backups each night. With this approach, data are backed up every night to a Storage Area Network (SAN). The backup is set to retain 100 days' worth of backup restore points and conducted by MACA IT staff.

Metadata

In 2018 IMD's Long-Term Data Management (LTDM) Governance Board formed a Metadata Working Group. The goal was to determine a metadata standard for the Inventory and Monitoring Division. A Remodeled Data Store will be designed to incorporate these recommendations. While the LTDM's work is not 100% complete the workgroup has produced a report (Sanders 2019) and a brief. The workgroup recommended that metadata be stored in an agnostic format in Data Store, allowing it to be output in a variety of standards including Ecological Metadata Language (EML) and Content Standard for Digital Geospatial Metadata (CSDGM). In the interim, the Cumberland Piedmont Network has created a [Data Store metadata record for the ozone database](https://irma.nps.gov/DataStore/Reference/Profile/2302022) (https://irma.nps.gov/DataStore/Reference/Profile/2302022). This record is updated annually as part

of IMD's Application Inventory Registry compliance (Kingston and Tucker 2017). All other ozone data products are cross referenced.

Protected Data

Each park in the Cumberland Piedmont Network has documented their requirements for the handling of potentially sensitive categories of data in a signed agreement between the network and the park. These agreements specify customized protected data handling guidelines. There are no protected data collected under this protocol. However, if ozone sensitive species lists are updated in the future, the status of each plant will be checked. The project lead would work with the data manager to apply the specified protections to the identified data categories. The current agreement (signed in 2019) and list of sensitive resources are available to internal NPS staff on the [Data Store](https://irma.nps.gov/DataStore/Reference/Profile/2260264) (<https://irma.nps.gov/DataStore/Reference/Profile/2260264>).

Literature Cited

- Cumberland Piedmont Network. 2024a. Ozone and Foliar Injury Monitoring. Post-sampling—Version 2.0. Standard Operating Procedure. NPS/CUPN/SOP—5. Cumberland Piedmont Network, Mammoth Cave, Kentucky. Available at: <https://irma.nps.gov/DataStore/Reference/Profile/2302029> (last accessed January 2024).
- Cumberland Piedmont Network. 2024b. Ozone and Foliar Injury Monitoring. Sampling—Version 2.0. Standard Operating Procedure. NPS/CUPN/SOP—4. Cumberland Piedmont Network, Mammoth Cave, Kentucky. Available at: <https://irma.nps.gov/DataStore/Reference/Profile/2302028> (last accessed January 2024).
- Cumberland Piedmont Network. 2024c. Ozone and Foliar Injury Monitoring. Data analysis and Reporting—Version 2.0. Standard Operating Procedure. NPS/CUPN/SOP—7. Cumberland Piedmont Network, Mammoth Cave, Kentucky. Available at: <https://irma.nps.gov/DataStore/Reference/Profile/2302031> (last accessed January 2024).
- Kingston S. D., and M. Tucker. 2017. To: National Park Service. Natural Resource Stewardship and Science Directorate. Inventory and Monitoring Division. Re: Data Management Brief - Application Inventory Registry (AIR) - AIR Compliance
- National Park Service. 2007. Natural Resource Database Template Version 3.2 documentation. Natural Resource Program Center, Office of Inventory, Monitoring, and Evaluation, Fort Collins, Colorado.
- National Park Service. 2016. Certification guidelines for Inventory and Monitoring data products. National Park Service, Inventory and Monitoring Division. Fort Collins, Colorado.
- Sanders G. 2019. Metadata recommendation and implementation plan Inventory and Monitoring Division of the National Park Service. Natural Resource Report. NPS/NRSS/IMD/NRR—2019/1956. National Park Service. Fort Collins, Colorado.
- United States Forest Service. 2008. Phase 3 Field Guide—Section 9. Ozone Bioindicator Plants (east). Southern Research Station. Knoxville, TN. 49pp.

List of In-House Documents Required by this SOP

1. In-House Document: Ozone Foliar Injury Data Entry. 2020. This document is maintained by the project lead and describes the procedure for entering data into the Ozone Foliar Injury database.
2. In-House Document: Natural History Collections. 2020. This document is maintained by the CUPN curatorial specialist and describes the steps used to process specimens and archives using the NPS Museum Collections Management System.