

4.0 ESTABLISHING REFERENCE CONDITION

4.1 Purpose

The establishment of reference conditions prior to implementing CERP project activities is one of the primary goals of the assessment and it is against these reference conditions that changes in the PMs that occur during and following CERP projects will be evaluated. The reference condition, or baseline, will include a minimum of four cumulative years of MAP data supplemented with relevant non-MAP data covering longer time series. In addition, the reference condition must be cumulative in order to include the full range of variability including capturing the extremes. The issue is confounded because of the heterogeneity in the temporal and spatial expression of the different PMs. For example, because hydrology and water quality changes may occur rapidly, changes in these parameters relative to a reference condition can be ambiguous. Changes in ecological responses are generally slower, thereby complicating the establishment of reference conditions as well as detecting change. It is critical that data sets used to establish the reference conditions meet the MAP guidance criteria.

The objectives of this component of the guidance process follow.

- 1) Describe the non-MAP data sources used in the assessment
- 2) Ensure that monitoring data meet the guidance criteria or provide rationale that justify their inclusion
- 3) Ensure that the data have been entered into the CERP's centralized environmental system, and updated appropriately.

4.2 Historic and Current Databases

The availability and quality of the monitoring data used to characterize the CERP reference conditions (baseline variability) for each of the performance measures may vary and may include MAP and non-MAP data (e.g. project level monitoring data). PIs and Module Groups should consider including all appropriate non-MAP data in their assessment efforts and consequently, identification and selection of the monitoring data used for establishing reference conditions and measuring changes in PMs must meet minimum criteria.

4.3 Criteria for the Use of Non-MAP Databases

To ensure that a wide range of data, in addition to those collected by the MAP, is available to estimate reference conditions for the PMs, we propose the following criteria be used to assure the data are comparable to, and consistent with, those being collected by the MAP. Observational data (existing water quality monitoring in the Greater Everglades), should include one or more of the following bullets listed below:

- The data were collected for sufficient duration to provide estimates of both intra- and inter-annual variability.

- Data used to detect change (either alone or in combination with other data) must be sufficient to assess the distributional characteristics of the PM.
- The sampling design used to collect the data should include the following elements: representation, randomization, replication, stratification, and blocking as appropriate.
- The data must be sufficient for power analysis.
- Data should be consistent and compatible with MAP data management.
- The data should be validated according to Quality Assurance Systems Requirements (QASR) procedures and meet CERP Zone criteria and standards whenever possible.
- Model-derived databases should be subjected to calibration and verification before being used to establish reference conditions for PMs.

4.4 Object-Oriented Databases

Sub-regional, regional, and system-wide models for hydrology, water quality, and ecology will be developed by CERP projects over the next few years or decades. These models will require and subsequently generate large amounts of spatial and temporal data for calibration, verification, and application. All the data (pre-and post-processed model simulation data) should be stored in a separate object-oriented database or in a centralized database system. Data quality objectives must be identified early in the process and be consistent with QASR (www.evergladesplan.org/pm/program_documents/qasr.cfm). In addition, these databases must comply with data validation and standardization requirements as described in the CERP Zone to assure accuracy and validity of the spatial and temporal data required by hydrologic, chemical, and ecological models.

Databases that store spatial and time series hydrological parameters (e.g., rainfall, stage, ground-water levels, evapotranspiration, aquifer characteristics), water quality parameters (e.g., nutrients, salinity, contaminants of concern, etc.) and ecological parameters (e.g., species densities and distributions, survival rates, habitat use, and movement, etc.) must meet the criteria to validate their use at appropriate spatial or temporal scales. Criteria include, but are not limited to: data standardization, data documentation, and data quality assurance.

4.4 Data Integrity

The integrity of databases and model output is integral to an accurate assessment of the changes in PMs and the progress of the CERP. The responsibility for assuring data quality will be shared between the individual PIs and the Module Group project managers. The Module Groups will be responsible for identifying data to be used in the assessment. As part of this responsibility, PIs will be responsible for ensuring that all data used in module level assessments satisfy the data integrity guidelines described for the CERP Zone. Ensuring that quality assurance and quality control (QA/QC) and data

validation processes provide accurate, defensible environmental data is critical to the overall success of CERP and each of its projects. The design, development and documentation of a QA program does not of itself ensure the quality of the data generated. Data quality is of utmost importance to the assessment of any data and is facilitated by adherence to the QASR, which provides guidelines for data quality and entities contributing data to CERP are required to conform to these guidelines. The Quality Assurance Oversight Team (QAOT) is responsible for ensuring compliance with the QASR and overall guidance on monitoring procedures, QA/QC and data validation.

4.5 Data Management

The RECOVER team of CERP has reached a point in the development of the program at which integrated assessments need to be initiated. The integrated assessments will depend on accessing and integrating a wide array of biological, chemical, and physical data collected by numerous organizations and scientists. To establish a baseline on which CERP projects can be evaluated, resources and efforts were initially focused on selecting sampling locations and collecting data. As a result, the data collected to this point are held by many PIs, in a variety of databases and with different standards for data quality, documentation, and coding. The purpose of this guidance is to identify the challenges that the Module Groups will face and to begin organizing the data to allow the integrated assessments to be accomplished.

4.5.1 Data Sources

The broad range of data being collected as part of CERP includes biological, chemical, physical, and geological data. Many of the projects within CERP are highly specialized and require interdisciplinary teams to integrate information from many sources. Data collection has focused on specific questions (not on integrated analyses) and data have been collected using different temporal and spatial scales. As the data are integrated the scales will sometimes need to be coordinated and summarized so that physical data on very small time scales but large spatial scales can be integrated with biological data collected on very broad time scales in localized spatial scales. Once these calculations are done they will need to be either stored or documented adequately to ensure that other researchers performing similar analysis know what analyses were done before and how the data normalization and integrations were performed.

The Module Groups face significant challenges in meeting the objectives of the integrated assessment. The first step is to clearly identify the questions/hypotheses that are to be used as the basis of the assessment. The identification of these questions will focus the analyses and will also help the data management staff understand the requirements for data, storage, documentation, analysis, and integration.

Second, a substantial amount of the physical data that will be required is already available in on-line, accessible relational databases. Many of these databases use industry-wide data and coding standards, ensuring that they can be easily integrated into and combined with data collected by PIs.

Third, there is an existing relational information system (ERDP) that houses data similar to the biological monitoring data being collected for CERP. This data management system has been developed, tested, and refined over a period of years and can serve as a prototype for the CERP central data management system that will house the CERP biological data.

Finally, several programs within SFWMD are designed to improve and integrate information systems. These programs receive broad funding and are supported by a staff knowledgeable in the collection and management of Everglades-related data.

4.6.2 Data Access and Distribution

Much of the data required for the MAP assessments is managed in a number of unrelated systems, including individual personal computers and cannot be easily accessed. In many cases the data will only be available by direct request from a PI. Many data sets may not be in relational databases and lack adequate documentation to allow other users to understand the applicable uses and the limitations of the collected data. To assure consistent reporting, the format and requirements for metadata will be coordinated with the RECOVER data stewards.

Different naming conventions (e.g., biological taxonomy, location names) used by the PIs will likely lead to inconsistencies that will make integrating the data difficult. Overcoming these difficulties will require direct input from the PIs and cross-references of different naming conventions before any meaningful integration can be accomplished.

The initial designs for collecting these data are focused on specific projects with little consideration for secondary uses or integration of the data for higher level questions. This will result in a variety of data collection methods, sampling and analytical methodologies, data documentation and spatial and temporal scale sampling. These discrepancies will need to be considered before data can be integrated. In some instances the data will need to be adjusted or summarized before data can be aggregated for integrated assessments.

The RECOVER MAP Statement of Work (SOW) Template presently specifies that PIs must submit all data for contract or interagency agreement closeout. Any data derived from the project will be provided to the AT at predetermined intervals and all data and results must be made available to the AT at the completion of the project.

4.6.3 Guidance Recommendations

To meet the data analysis obligations, the Module and Data Management groups should begin preparing data and resources for integrated assessments as soon as possible. An effective approach would be to develop a set of priority assessment hypotheses that will be used to assess the progress and success of CERP. These hypotheses should then be used to guide the organization of the data and analyses. For each assessment hypothesis,

a group of experts should be brought together who are knowledgeable about the domains of the hypothesis and who can identify relevant data and resources for performing the respective assessments. The core elements that should be addressed include, but may not be limited to, the following:

- Identify major groups of similar data
- Identify specific fields required to adequately describe and accommodate data
- Identify sources of data, for example, who collected them, why they were collected, and how they were organized into a database or table
- Identify locations of the databases (e.g., individual PI computers, government databases, web addresses, notebooks) and how they are currently shared (e.g., on-line, subscription database, CD, sneaker net, hard copies)
- List standards that apply to the data
- Indicate if there are naming conventions for locations, species, fields etc.
- Describe spatial and temporal scales of sampling
- Describe how various data sources will combined for the integrated analyses
- Provide metadata (e.g., written documentation of the original question being addressed, sampling plan, analytical methods, extent of the data, limitations, etc.)
- Document quality assurance performed on the data.

This information can then be used to develop a data analysis plan for the specific hypothesis. The data analysis plan should specify what preprocessing of the data, if any, will be required to combine diverse sources of information in order to conduct an assessment.