1.0 PROJECT MANAGEMENT – ORGANIZATION AND RESPONSIBILITIES

1.1 Title and Approval Page

Secondary Data Quality Assurance Project Plan

for

Application of Technical Approach for Establishing Nitrogen Thresholds and Allowable Loads for Three LIS Watershed Groupings: Embayments, Large Riverine Systems and Western LIS Point Source Discharges to Open Water

Contract Number EP-C-12-055
Task Order 0023

Prepared for:

U.S. Environmental Protection Agency
Region 1 – New England
5 Post Office Square
Boston, MA 02109

Prepared by:

Tetra Tech, Inc.
10306 Eaton Place, Suite 340
Fairfax, VA 22030

December 15, 2016

QAPP 476, Revision 0

Effective Date with Signatures: January 11, 2017

This quality assurance project plan (QAPP) has been prepared according to guidance provided in the following documents to ensure that environmental and related data collected, compiled, and/or generated for this project are complete, accurate, and of the type, quantity, and quality required for their intended use:

- *EPA Requirements for Quality Assurance Project Plans* (EPA QA/R-5, EPA/240/B-01/003, U.S. Environmental Protection Agency, Office of Environmental Information, Washington DC, March 2001 [Reissued May 2006]). http://www.epa.gov/quality/qs-docs/r5-final.pdf
- *EPA Guidance for Quality Assurance Project Plans* (EPA QA/G-5, EPA/240/R-02/009, U.S. Environmental Project Agency, Office of Environmental Information, Washington DC, December 2002a). https://www.epa.gov/sites/production/files/2015-06/documents/g5-final.pdf
- *New England QAPP Guidance for Projects Using Secondary Data, Revision 2* (U.S. Environmental Protection Agency, New England, Quality Assurance Unit, Office of Environmental Measurement and Evaluation, Boston, MA, October 2009a). https://www.epa.gov/sites/production/files/2015-06/documents/EPANESecondaryDataGuidance.pdf
- *Guidance for Geospatial Data Quality Assurance Project Plans* (EPA QA/G-5G U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC, March, 2003). http://www.epa.gov/sites/production/files/documents/guidance_geospatial_data_qapp.pdf

Tetra Tech, Inc., will conduct work in conformance with procedures detailed in this QAPP.

Approvals:

12/15/16
Dakis Mosso
Date
Co-Task Order Leader
Tetra Tech, Inc.

12/15/16
Michael Paul, Ph.D.
Date
Co-Task Order Leader
Tetra Tech, Inc.

12/15/16
John Hochheimer, Ph.D.
Date
Technical Monitor
Tetra Tech, Inc.

12/15/16
Susan Lanberg
Date
Quality Assurance Officer
Tetra Tech, Inc.

12/15/16
Jon Butcher, Ph.D., PH
Date
Water Quality Modeling Reviewer
Tetra Tech, Inc.

12/15/16
Peter Cada, GISP
Date
Geospatial Data Steward
Tetra Tech, Inc.

12/15/2016
Saumya Sarkar
Date
Geospatial Data Steward QC Officer
Tetra Tech, Inc.

12/15/16
Jon Harcum, Ph.D.
Date
Database and Statistical Lead
Tetra Tech, Inc.

12/15/2016
Lei Zheng, Ph.D.
Date
Statistical Analyst
Tetra Tech, Inc.

12/15/16
Alex DeWire
Date
Database Developer
Tetra Tech, Inc.

12/15/2016
[illegible] Bissen
Date
Script QC Officer
Tetra Tech, Inc.

1-11-17
Ray Cody
Date
Task Order Contracting Officer's Representative
U.S. Environmental Protection Agency Region 1

1/10/2017
Nora Conlon, Ph.D.
Date
Quality Assurance Manager
U.S. Environmental Protection Agency Region 1

1/4/17
Leah O'Neill
Date
Project Team Leader
U.S. Environmental Protection Agency Region 1

1/4/17
Mark Tedesco
Date
Project Technical Lead
U.S. Environmental Protection Agency LIS Office

1/3/17
David Pincumbe
Date
Project Technical Advisor
U.S. Environmental Protection Agency Region 1

1/9/17
Bob Nyman
Date
Project Technical Advisor
U.S. Environmental Protection Agency Region 2

1.2 Table of Contents

Tables

Figures

Appendices

Acronyms and Abbreviations

AVGWLF	ArcView GIS Generalized Watershed Loading Function
CCMP	Comprehensive Conservation and Management Plan
chl-a	chlorophyll-a
Co-TOL	Co-Task Order Leader
DO	dissolved oxygen
EFDC	Environmental Fluid Dynamics Code
EPA	(U.S.) Environmental Protection Agency
ESRI™	Environmental Systems Research Institute, Inc.
GIS	geographic information system
GISP	Geographic Information Systems Professional
HSPF	Hydrological Simulation Program–Fortran
ICIS	Integrated Compliance Information System
LINAP	Long Island Nitrogen Action Plan
LIS	Long Island Sound
LISS	Long Island Sound Study
LRM	Literature Review Memorandum
LSPC	Loading Simulation Program in C++
MS4	municipal separate storm sewer system
MWWTP	municipal waste water treatment plant
NE SPARROW	New England SPAtially Referenced Regressions on Watershed Attributes
NGDP	National Geospatial Data Policy
NHDPlusV2	National Hydrography Dataset Version 2
NPDES	National Pollutant Discharge Elimination System
NYHOPS	New York Harbor Observing and Predicting System
OEI	(EPA) Office of Environmental Information
ORD	(EPA) Office of Research and Development
OST	(EPA) Office of Science and Technology
PARCC	precision, accuracy, representativeness, comparability, and completeness
PWS	Performance Work Statement
QA	quality assurance
QAPP	quality assurance project plan
QAU	Quality Assurance Unit
QC	quality control
SOP	standard operating procedure
SWAT	Soil and Water Assessment Tool
SWEM	System-wide Eutrophication Model
TMDL	total maximum daily load
TN	total nitrogen
TOCOR	Task Order Contracting Officer’s Representative
USGS	U.S. Geological Survey
WASP	Water Quality Analysis Simulation Program
WPCF	water pollution control facility

1.3 QAPP Distribution List

This quality assurance project plan (QAPP) will be distributed to the U.S. Environmental Protection Agency (EPA) staff listed on the title page through the EPA Task Order Contracting Officer's Representative (TOCOR). The Tetra Tech, Inc., Co-Task Order Leaders (Co-TOLs) listed on the title page will distribute the approved document to all Tetra Tech staff involved in this project.

Name	Phone number E-mail	Mailing address or regional office
Tetra Tech, Inc.		
James Bisese Script QC Officer	202-236-1077 james.bisese@tetratech.com	Tetra Tech, Inc. 3775 Iris Avenue, Suite 4 Boulder, CO 80301
Jon Butcher, Ph.D. Water Quality Model Reviewer	919-485-2060 jon.butcher@tetratech.com	Tetra Tech, Inc. 1 Park Drive, Suite 200 P.O. Box 14409 Research Triangle Park, NC 27709
Peter Cada, GISP Geospatial Data Steward	919-485-2071 peter.cada@tetratech.com	Tetra Tech, Inc. 1 Park Drive, Suite 200 P.O. Box 14409 Research Triangle Park, NC 27709
Alex DeWire Database Developer	703-385-1428 alex.dewire@tetratech.com	Tetra Tech, Inc. 10306 Eaton Place, Suite 340 Fairfax, VA 22030
Jon Harcum, Ph.D. Database and Statistical Lead	864-650-5815 jon.harcum@tetratech.com	Tetra Tech, Inc. 313 Kings Way Clemson, SC 29631
John Hochheimer, Ph.D. Technical Monitor	703-385-1696 john.hochheimer@tetratech.com	Tetra Tech, Inc. 10306 Eaton Place, Suite 340 Fairfax, VA 22030
Susan Lanberg Quality Assurance Officer	703-385-1906 susan.lanberg@tetratech.com	Tetra Tech, Inc. 10306 Eaton Place, Suite 340 Fairfax, VA 22030
Dacia Mosso Co-Task Order Leader	703-385-1437 dacia.mosso@tetratech.com	Tetra Tech, Inc. 10306 Eaton Place, Suite 340 Fairfax, VA 22030
Michael Paul, Ph.D. Co-Task Order Leader	919-485-2073 michael.paul@tetratech.com	Tetra Tech, Inc. 1 Park Drive, Suite 200 P.O. Box 14409 Research Triangle Park, NC 27709
Sam Sarkar Geospatial Data Steward QC Officer	919-485-2069 saumya.sarkar@tetratech.com	Tetra Tech, Inc. 1 Park Drive, Suite 200 P.O. Box 14409 Research Triangle Park, NC 27709
Lei Zheng, Ph.D. Statistical Analyst	410-356-8993 lei.zheng@tetratech.com	Tetra Tech, Inc. 10711 Red Run Blvd., Suite 105 Owings Mills, MD 21117

Name	Phone number E-mail	Mailing address or regional office
U.S. Environmental Protection Agency		
Ray Cody Task Order Contracting Officer's Representative	617-918-1366 cody.ray@epa.gov	U.S. Environmental Protection Agency Region 1 – New England 5 Post Office Square Mail Code: OEP06-1 Boston, MA 02109
Bob Nyman Project Technical Advisor	212-637-3809 nyman.robert@epa.gov	U.S. Environmental Protection Agency Region 2 290 Broadway Mail Code: 24TH FL New York, NY 10007
Leah O'Neill Project Team Leader	617-918-1633 oneill.leah@epa.gov	U.S. Environmental Protection Agency Region 1 – New England 5 Post Office Square Mail Code: OEP06-1 Boston, MA 02109
Greta Perry Contracting Officer	513-487-2331 perry.greta@epa.gov	U.S. Environmental Protection Agency 26 West Martin Luther King Drive Mail Code: NWD1 Cincinnati, OH 45268
David Pincumbe Project Technical Advisor	617-918-1695 pincumbe.david@epa.gov	U.S. Environmental Protection Agency Region 1 - New England 5 Post Office Square Mail Code: OEP06-1 Boston, MA 02109
Nora Conlon, Ph.D. Quality Assurance Manager	617-918-8335 Conlon.Nora@epa.gov	U.S. Environmental Protection Agency Region 1 – New England Regional Laboratory 11 Technology Drive Mail Code: EQA North Chelmsford, MA 01863
Mark Tedesco Project Technical Lead	203-977-1542 tedesco.mark@epa.gov	USEPA Region 2 - Long Island Sound Office Stamford Government Center 888 Washington Boulevard Stamford, CT 06904
Steven Winnett Alternate Task Order Contracting Officer's Representative	617-918-1687 winnett.steven@epa.gov	U.S. Environmental Protection Agency Region 1 – New England 5 Post Office Square Mail Code: OEP06-2 Boston, MA 02109

(This page is intentionally blank.)

1.4 Project Organization

The purpose of this quality assurance project plan (QAPP) is to describe the quality system that Tetra Tech, Inc., will implement to support U.S. Environmental Protection Agency (EPA) Region 1 in the application and refinement of a technical approach (methodology) for establishing nitrogen thresholds and allowable loads consistent with achieving desired water quality conditions and uses. The work is organized by three watershed groupings (embayments, large riverine systems, and western Long Island Sound [LIS] point source discharges to open water) in LIS. This technical approach would then be used to calculate allocations for watersheds, with an initial focus on priority embayments, tributaries, and western LIS. Funding for this project is provided by EPA Region 1, under contract number EP-C-12-055, Task Order 0023.

Tetra Tech does not currently expect that mechanistic numeric modeling, sampling, or primary data collection will be performed for Task Order 0023. This QAPP follows a format that has been recommended by EPA Region 1 for projects using secondary data (USEPA 2009a). Because this Task Order relies heavily on secondary data analysis to assist EPA in the numeric interpretation of narrative nitrogen criteria to inform permit writers, only the secondary data collection and analyses are described in this QAPP and not any sample manipulations or methods that were used to generate the primary data.

If the EPA Task Order Contracting Officer's Representative (TOCOR) requests that mechanistic numeric modeling be performed for this project in the future, Tetra Tech will update this QAPP (or prepare a standard operating procedure [SOP] that covers these activities as an attachment to this QAPP) to ensure that it covers modeling activities. Tetra Tech will ensure that the updated QAPP or SOP complies with EPA's (2016) *Design and Implementation of New Tools for Quality Assurance in Modeling*, EPA's (2009b) *Guidance on the Development, Evaluation, and Application of Environmental Models*, and EPA's *Guidance for Quality Assurance Project Plans for Modeling* (USEPA 2002b). Tetra Tech will submit the updated QAPP or SOP to EPA for review and approval prior to performing modeling activities.

The project organization chart, presented in Figure 1, includes relationships and lines of communication among all participants and data users. The responsibilities of these persons are described below.

The EPA TOCOR, Ray Cody, will provide overall project and program oversight for Task Order 0023. He will review and approve the project work plan, QAPP, and other materials developed to support the project and will coordinate with contractors, reviewers, and others to ensure technical quality in all deliverables and adherence to the contract. EPA's Project Team consists of the Project Team Leader, Leah O'Neill, the Project Technical Lead, Mark Tedesco, and Project Technical Advisors, David Pincumbe and Bob Nyman. EPA's Project Team will assist the EPA TOCOR in managing and coordinating project activities.

The EPA Quality Assurance (QA) Manager is Nora Conlon. Her responsibilities include reviewing and approving the QAPP and participating in any Agency reviews of work performed, as appropriate.

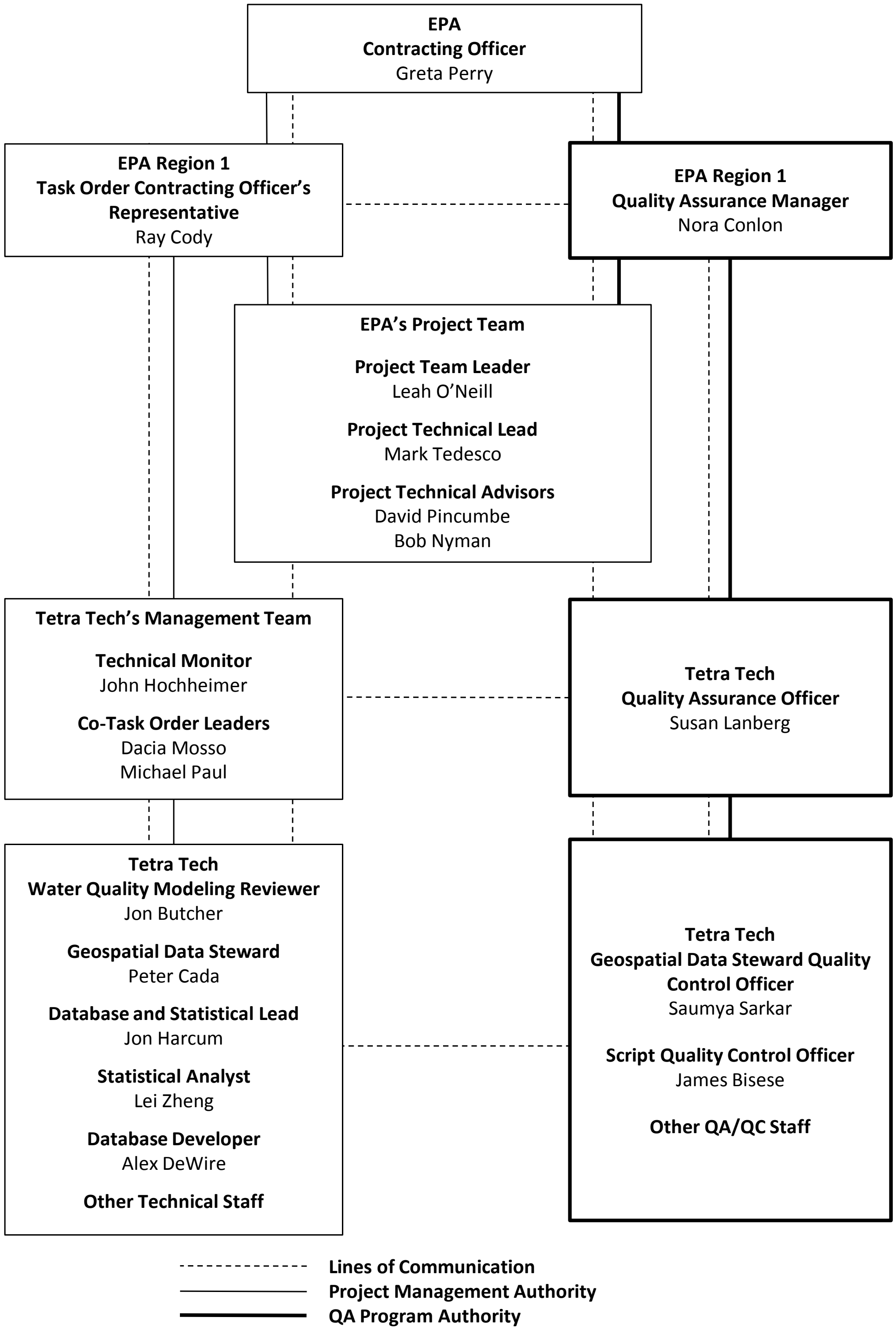

Figure 1. Project Organization

Tetra Tech's Management Team consists of the Tetra Tech Technical Monitor, John Hochheimer, and the Tetra Tech Co-Task Order Leaders (Co-TOLs), Dacia Mosso and Michael Paul. The Tetra Tech Technical Monitor will provide technical oversight for the overall project and the Tetra Tech Co-TOLs will supervise the overall project. Specific responsibilities of the Tetra Tech Co-TOLs include the following: performing or overseeing staff responsible for selecting, compiling, and evaluating existing data; coordinating project assignments; establishing priorities and schedules; ensuring completion of high-quality projects within established budgets and schedules; providing guidance, technical advice, and evaluation of the performance of those assigned to the project; implementing corrective actions; preparing or overseeing preparation and review of project deliverables, responses to EPA, action memos, and any other materials developed to support the project; and providing support to EPA in interacting with the Project Team, technical reviewers, and others to ensure that technical quality requirements are met in accordance with EPA's objectives.

Jon Butcher, the Water Quality Modeling Reviewer, will evaluate water quality models (e.g., System-wide Eutrophication Model [SWEM], New England SPAtially Referenced Regressions on Watershed Attributes [NE SPARROW] model, ArcView GIS Generalized Watershed Loading Function [AVGWLF] model) to determine whether their outputs are suitable for use in in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS.

Peter Cada, the Geospatial Data Steward, will supervise the geospatial information operations performed for this project and ensure they comply with the EPA *National Geospatial Data Policy* (NGDP; USEPA 2005) and the EPA *National Geospatial Data Policy Procedure for Geospatial Metadata Management* (USEPA 2007) described in Section 1.7.1 of this QAPP. His responsibilities include complying with applicable procedures and standards to meet project objectives and produce documented results or products of known quality. He will ensure that geospatial data compiled for this project conform to data exchange protocols, and applicable data standards as defined and maintained by EPA's Office of Environmental Information (OEI). He will also document geospatial metadata for all spatial data elements compiled into data sets for this project, in accordance with the provisions of FGDC-STD-001-1998, *Content Standard for Digital Geospatial Metadata*.

Jon Harcum (the Database and Statistical Lead) with support from Lei Zheng (the Statistical Analyst) and Alex DeWire (the Database Developer), will oversee database development and statistical analyses to explore relationships among and develop stressor-response relationships for response measures and nitrogen species within hydro-subregions of three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS to develop nitrogen thresholds.

Other Tetra Tech technical staff will assist in selecting, compiling, and evaluating secondary data, and aid in completing data analyses. Technical staff will implement the QA/quality control (QC) program, complete assigned work on schedule and with strict adherence to the established procedures, and complete required documentation.

The Tetra Tech QA Officer for this Task Order is Susan Lanberg, whose primary responsibilities include the following: providing oversight support to the Tetra Tech Co-TOLs in preparing the QAPP, reviewing and approving the QAPP, and, with the assistance of assigned QC Officers, monitoring QC activities to determine conformance with QA/QC requirements.

A QC Officer is a technical staff member who is familiar with the project tasks but does not participate in the task or subtask that he or she checks. QC Officers ensure independence of the quality function during execution of specific project tasks, oversee work to ensure that QC is maintained throughout data compilation and analyses; notify the Tetra Tech QA Officer of any difficulties experienced in meeting the QAPP requirements; and coordinate with the Tetra Tech QA Officer on the ongoing requirements for corrective action investigations. QC evaluations include checking electronic data sets, reviewing calculations, double-checking work as it is completed, and documenting these reviews to ensure that the standards set forth in the QAPP and technical proposal are met.

The Tetra Tech Geospatial Data Steward QC Officer, Saumya Sarkar, will provide oversight to ensure that geospatial information operations performed for this project comply with the EPA NGDP (USEPA 2005) and the EPA *National Geospatial Data Policy Procedure for Geospatial Metadata Management* (USEPA 2007) described in Section 1.7.1 of this QAPP. The Tetra Tech Script QC Officer, James Bisese, will independently review and test the statistical scripts prepared for this project to ensure that they are performing as intended, and yielding desired and accurate outputs.

Other QA/QC staff, including technical reviewers and technical editors selected as needed, will provide review oversight of the content of the work products and ensure that the work products comply with EPA's specifications.

All key technical project personnel will have expertise in environmental sciences, review this QAPP, and be knowledgeable of the quality system for this project. They will use this expertise and knowledge to obtain and analyze existing data to support EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS. A description of the experience of key Tetra Tech staff relevant to this project is provided below.

Tetra Tech Co-TOL, Ms. Dacia Mosso, is an associate director and environmental scientist at Tetra Tech with 17 years of experience in aquatic resources, water quality standards, effluent limitations guidelines, regulatory processes, nutrient criteria, and environmental policy. She has applied her strong technical writing and editing skills to numerous guidance documents, outreach products, and website content for EPA. Ms. Mosso also has provided technical and project management support on many EPA projects to develop and implement water quality criteria and standards. Ms. Mosso has worked on numerous response to comments efforts for EPA guidance/regulations, managing logistics and staff, developing coding outlines and QA procedures, analyzing comments and content, coding comment letters, and performing QA. Currently, Ms. Mosso is the deputy program manager on a major EPA contract in the Office of Science and Technology (OST).

Tetra Tech Co-TOL, Dr. Michael Paul, is an aquatic ecosystem ecologist/biogeochemist with more than 19 years of experience in the research and management of aquatic ecosystems. His work, which has included teaching, research, and public policy, has focused on the ecology of freshwater ecosystems, including more than 12 years of experience in water quality standards development across the nation. He has provided technical support in assessment and criteria development for more than 29 states, tribes, and federal government agencies, has developed instructional materials for and led instructional workshops on assessment, analysis, and criteria development across the nation, and has co-authored EPA guidance on the statistical analysis of bioassessment data, the design, sampling and analysis of bioassessment for large rivers, and the application of stressor-response analysis for nutrient criteria development. Dr. Paul has also been involved in several EPA Office of Research and Development (ORD) ecological risk assessment projects involving causal analysis, multiple stressor analysis, and the effects of climate change on state water quality assessment and criteria programs. Dr. Paul currently co-manages the national nutrient criteria support center for EPA OST and is project manager of a 5 year, Ecological Risk Support contract in ORD. Dr. Paul has authored more than 15 peer reviewed scientific papers, proceedings, and book chapters and more than 31 technical reports.

Tetra Tech Water Quality Modeling Reviewer, Dr. Jonathan Butcher, is a registered Professional Hydrologist and environmental engineer with more than 29 years of experience in watershed planning, risk assessment, and development, application, and communication of hydrologic, hydraulic, and water quality models. Dr. Butcher has led technical efforts to support EPA, state, and local governments in a variety of total maximum daily load (TMDL), wasteload allocation, watershed modeling, and water body restoration and protection studies. He is a nationally recognized expert in the application of Hydrological Simulation Program–Fortran (HSPF), Soil and Water Assessment Tool (SWAT), and other watershed models and has worked with model developers to test, debug, modify, and improve modeling code. Dr. Butcher has developed numerous lake, reservoir, and estuarine response models using Water Quality Analysis Simulation Program (WASP), Environmental Fluid Dynamics Code (EFDC), CEQUAL-W2, and a variety of other tools. He has led comprehensive modeling and assessment projects for reservoir source water protection using linked watershed and receiving water models in a variety of locations. He has developed multimedia models for mercury and persistent organic pollutant TMDLs and is currently leading the development of 20 large-scale watershed models across the United States to evaluate hydrologic impacts of climate change. Dr. Butcher's research interests also include development of TMDLs to address narrative criteria for sediment and nutrients. He is experienced in use of numerous lake, river, and estuarine hydrodynamic, hydrologic, and water quality models, and has conducted flow, sediment, dissolved oxygen (DO), nutrient, algae, and toxics modeling on a variety of water systems

Tetra Tech Geospatial Data Steward, Mr. Peter Cada, is an environmental scientist with more than 10 years of experience in the areas of environmental and water resources analysis, watershed planning and restoration, urban storm water and agricultural runoff management, and watershed and water quality modeling. Mr. Cada received a Certificate in Geospatial Analysis from the Nicholas School of Environment, providing advanced skills in the use of geographic information system (GIS) software packages to support a variety of water resource projects. He is also experienced with database construction and management and statistical analyses. Since joining Tetra Tech, he has served as GIS lead for numerous projects managing a variety of large and complex datasets, geospatial analyses, and GIS-based figure production. Because of this, Mr. Cada has become the designated lead GIS applications person for the Tetra Tech Water Resources Group. He has performed technical and GIS analyses for numerous watershed models and management plans; technical, GIS and field analyses of potential stream and wetland restoration project sites; field validation for water quality modeling; stream habitat and geomorphic assessment; and pollutant load evaluation for land use planning. He also has experience training individuals and groups in the use of Environmental Systems Research Institute, Inc. (ESRI) ArcGIS software (versions 9.x and 10.x). He is a certified Geographic Information Systems Professional (GISP) by the GIS Certification Institute.

Tetra Tech Database and Statistical Lead, Dr. Jon Harcum, is a principal engineer/hydrologist at Tetra Tech specializing in ground and surface water quality hydrology, statistics, data analysis, and water quality monitoring network design. He has managed or served as principal investigator on more than 50 environmental projects using multiple subcontractors. Dr. Harcum has direct experience for EPA using software, including ArcInfo, MATLAB, Minitab, R, SAS, SeaDAS, and Virtual Beach in addition to basic Microsoft data tools. He has developed EPA guidance documents and training material for nutrient management plan development; calculating TMDLs in streams and rivers; tracking, evaluating, and reporting implementation of forestry, urban, and agriculture nonpoint source control measures; determining the effectiveness of nonpoint source controls; and the TMDL process. He supported development of EPA’s risk-based assessment of sediment quality on a national level, which included assessing toxicity, tissue residue, and sediment chemistry data to evaluate human and ecological risks. Dr. Harcum’s research has included evaluating and comparing numerous trend detection and ambient condition estimation procedures commonly used to analyze water quality data that have data limitations such as multiple detection limits, irregular sampling, and missing values.

Tetra Tech Geospatial Data Steward QC Officer, Mr. Saumya Sarkar, is a civil engineer experienced in watershed model development and applications, GIS analysis, and software programming for water resources applications. He has supported EPA, state, and local governments in a variety of watershed modeling, wasteload allocation, TMDL, and climate change studies. He has expertise in the application of SWAT, HSPF, Loading Simulation Program in C++ (LSPC), and other watershed models and supported the development of 20 large-scale watershed models across the United States to evaluate hydrologic and water quality impacts of climate change. He is leading SWAT model development and climate change analysis for Lake Champlain basin to support EPA in the formulation of a TMDL for phosphorus. He has developed several other SWAT models to support TMDL development, watershed management plans, and climate change analysis throughout the United States. Mr. Sarkar has also supported development of several HSPF models for Minnesota for their statewide Nutrient Reduction Strategy. He has also supported development of several HSPF and LSPC models in California and North Carolina for development of TMDLs and watershed management plans. Mr. Sarkar is skilled in GIS and has aided in spatial analyses, geo-processing and data analysis for several projects. He also developed software applications for water resources including tools for atmospheric weather data development for watershed and receiving water models, statistical analysis of water quality data, and analysis of watershed modeling results using visual basic for applications and Python.

Tetra Tech Statistical Analyst, Dr. Lei Zheng, is a senior aquatic ecologist and statistician within Tetra Tech's Center for Ecological Sciences. He is a research scientist with more than 20 years of broad experience in aquatic environmental sciences, specialized in data analysis and environmental statistics. During his professional career, he has been involved in a range of multidisciplinary research on spatial analysis, species distribution modeling, and responses of species distribution to climate change. He has applied his strong background in experimental design and statistical analysis to integrate complicated information into simple models and indices for environmental management. He has been active in the field of causal analysis, working with both EPA and state agencies to identify environmental stressors and human disturbance that cause the degradation of aquatic ecosystems. Dr. Zheng also has experience handling Big Data (geospatial datasets), Relational Databases, Quantitative Data Analysis, Experimental/Survey Design, Metric/Index Development and Calibration, Data Visualization, Cost/Benefit Evaluation, and Predictive Model Development. He is an expert in R Programming (an instructor for numerous workshops), works with Python programming, and offers an array of other statistical/analytic expertise.

1.5 Purpose of Study, Background Information, and Problem Definition

Hypoxia, defined as DO levels of less than 3 milligrams per liter (mg/L), is a common occurrence in LIS bottom waters during the summer, affecting up to half of its area in some years. In LIS, nitrogen is the primary limiting nutrient for algal growth. Impairments linked to excess discharges of nitrogen include harmful algal blooms, low DO, poor water clarity, loss of submerged aquatic vegetation and tidal wetlands, and coastal acidification.

The Long Island Sound Study (LISS) has focused on understanding the drivers to hypoxia and developing tools to support nitrogen management. The LISS developed and, in 1998, adopted a plan entitled Phase III Actions for Hypoxia Management that identified the sources and loads of nitrogen to LIS and recommended nitrogen reduction targets.

In 2000, Connecticut and New York incorporated these targets into a Total Maximum Daily Load to Achieve Water Quality Standards for Dissolved Oxygen in Long Island Sound (LIS TMDL). The LIS TMDL allocated a 58.5 percent nitrogen reduction to in-basin sources of enriched nitrogen (with a 10 percent reduction allocated to nonpoint sources and the remainder assigned to point sources). In addition, the LIS TMDL identified actions and schedules to reduce nitrogen from tributary sources (25 percent reduction to point sources, 10 percent reduction to nonpoint sources) and atmospheric sources (an 18 percent reduction), and to implement non-treatment alternatives (e.g., bioextraction, aeration) necessary to fully attain DO water quality standards.

Based on monitoring and modeling efforts to this day, current and planned actions by the states are expected to fall short of fully implementing the 2000 TMDL and will be insufficient to address other adverse impacts to water quality in LIS, and its embayments and near shore coastal waters.

EPA has developed a Nitrogen Reduction Strategy (Strategy; http://longislandsoundstudy.net/issues-actions/water-quality/nitrogen-strategy/) to aggressively continue progress on nitrogen reductions, in parallel with the states' continued implementation of the 2000 TMDL, and achieve water quality standards throughout LIS and its embayments and near shore coastal waters. The strategy recognizes that more work must be done to reduce nitrogen levels, further improve DO conditions, and address other nutrient-related impacts in LIS. EPA's Strategy expands the focus of the 2000 TMDL to include other nutrient-related adverse impacts to water quality, such as loss of eelgrass, that affect many of LIS's embayments and near shore coastal waters. Expanding the focus in this way will help restore and protect these important habitats from a range of nutrient-caused impairments, ensuring that water quality standards are achieved in near shore waters as well as supporting the attainment of water quality standards in the open water portion of the Sound.

The Strategy is organized by the three customized watershed groupings: embayments, large riverine systems, and priority western LIS point source discharges to open water. Common to each grouping is the need to:

- Develop nitrogen thresholds that will translate the narrative water quality standard into a numeric target
- Identify where nitrogen watershed loading results in threshold exceedances
- Assess options for the load reductions from point and nonpoint sources that would be needed to remain below thresholds

Nitrogen loads will need to be customized for each watershed grouping and a specific allocation proposed for priority embayments/subwatersheds. Customizing the application of nitrogen thresholds for each grouping recognizes their distinct watershed and receiving water characteristics.

1.6 Overview of Project Tasks

The purpose of this Task Order is to support EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS. The specific tasks are defined below. Technical direction will be provided to Tetra Tech for clarification purposes through written communication provided by the EPA TOCOR. Additional background and more details regarding the Task Order 0023 performance work statement (PWS), including the approach for analyzing data, are provided under the individual task descriptions in Section 1.6.1. Additional details on the type and amount of data gathered, and how data sources will be evaluated for use, are provided in Section 1.6.2.

1.6.1 Task Order 0023 Tasks

Task 0. Work Plan and Budget Development

Tetra Tech prepared a detailed Work Plan and budget response to the PWS.

This Task Order includes work to be conducted in two terms:

- A Base Period (September 30, 2016 through September 29, 2017) and an Option Period 1 (September 30, 2017 through March 27, 2018 [contract vehicle expiration]).

- Work to be conducted in the Base Period includes: Task 1, Task 2 and Task 3 for Primary Tier Watershed Groupings.
- Work to be conducted in Option Period 1 includes: Task 1B (as appropriate) and Task 3B for completion of work not conducted in the Base Period for Primary Tier Watershed Groupings, and the Secondary Tier Watershed Groupings.

Task 1. Project Management and Administration

This task includes subtasks related to administration, management and coordination of the project. A description of EPA's Project Team and Tetra Tech's Management Team is provided in section 1.4 of this QAPP.

The EPA Project Team will coordinate interactions and meetings with multiple stakeholders, including the following:

- Connecticut Department of Energy and Environmental Protection
- New York State Department of Environmental Conservation
- Massachusetts Department of Environmental Protection
- New Hampshire Department of Environmental Services
- Vermont Department of Environmental Conservation
- New England Interstate Water Pollution Control Commission
- Connecticut River Watershed Council
- LIS Science and Technical Advisory Committee
- LIS Citizens Advisory Committee
- Other local watershed groups, and additional partners where appropriate as determined by EPA Project Team Leader

The EPA Project Team Leader will convene a Technical Stakeholder Group to consist of qualified stakeholders that will assist the EPA Project Team. A primary responsibility of the EPA Project Team Leader will be coordinating with all stakeholders. Tetra Tech will provide assistance to the EPA Project Team Leader as generally described herein.

Tetra Tech will provide monthly progress reports and invoices, per the terms of the contract.

Subtask 1A. Kickoff Meeting

Tetra Tech initiated a project kickoff meeting with the EPA Project Team at EPA's LIS Office. The meeting was held on October 12, 2016. Tetra Tech prepared a draft agenda for the meeting. At this meeting, EPA made available additional technical references and other supplemental data and information to Tetra Tech.

A week following this meeting, Tetra Tech summarized its understanding of the project kickoff meeting (e.g., action items; scheduling adjustments) and transmitted these by email to the EPA Project Team Leader for distribution to the EPA Project Team.

Subtask 1B. Conference Calls, Meetings and Project Team Support

Following the kickoff meeting, Tetra Tech will provide for monthly conference calls (as needed) to keep the EPA Project Team updated as to the status of the project. These calls will use either EPA's or Tetra Tech's teleconferencing facilities and will use remote desktop sharing, if appropriate.

Tetra Tech will provide presentation materials and present routine summaries of its progress to the Technical Stakeholder Group and external partners.

Tetra Tech will briefly summarize its understanding of each conference call (e.g., action items; scheduling adjustments) and/or meeting and transmit these by email to the EPA Project Team Leader for distribution to the EPA Project Team and Technical Stakeholder Group.

It is possible that the calls and/or meetings could generate a need to respond to, or otherwise address, comments from the EPA Project Team and/or the Technical Stakeholder Group. Tetra Tech will work with the EPA Project Team Leader (with Technical Direction if necessary) to assist the EPA Project Team Leader and the EPA Project Team to develop formal responses to comments.

Task 2. Development of Quality Assurance Project Plan (QAPP)

Tetra Tech developed a draft QAPP for TO 0023. Tetra Tech will submit the QAPP to the EPA TOCOR and EPA Project Team Leader, and eventually to EPA's Regional Quality Assurance Unit (QAU) for approval. Tetra Tech began consideration and development of the QAPP upon initiation of the project. The QAPP had been provided to EPA in draft within three months of the Project Kickoff Meeting. All comments developed from the review have been incorporated by Tetra Tech into a final QAPP for submittal by the EPA Project Team Leader to the QAU. Because the project contemplates application of a technical approach for the numeric interpretation of narrative nitrogen criteria which may implicate an iterative process of refinement, the QAPP may require periodic adjustment to account for modifications arising from iteration.

Task 3. Application and Refinement of Technical Approach

The steps, or "algorithms" for this technical approach include the following:

- Collecting nitrogen information (e.g., loading, nitrogen concentration)from existing analytical results (from a number of sources), as described in Parts A through I below
- Finding the relationship of this nitrogen information (e.g., loads, concentrations) to LIS water quality model results and chlorophyll-a (chl-a) data, as described in Parts A through I below
- Using these results to make recommendations for nitrogen thresholds for identified watershed groupings
- Estimating reduction levels by prioritized watersheds necessary to meet LIS-specific, regional, and developing Long Island Nitrogen Action Plan (LINAP) thresholds
- Proposing nitrogen load allocations among categories of nitrogen sources

Specific tasks proposed to accomplish this technical approach include:

Part A
Summarize nitrogen loads (watershed loading and embayment area normalized loading) and sources from each coastal embayment from Vaudrey et al. 2016 and other technical assessments (e.g., the Nature Conservancy, and those developed through the LINAP and the 2016 Suffolk County Sub-watersheds Wastewater Plan). Identify and record the location of uncertainty estimates for loads where available, to be used as part of a qualitative treatment of load variability as applicable. [*All LIS embayments*]

The source data and location where the data may be obtained include, but may not be limited to the following:

- Vaudrey research
- The Nature Conservancy
- LINAP (including contractors supporting LINAP)
- U.S. Geological Survey (USGS)

Part B
Summarize flow, total nitrogen (TN) load, and TN concentration for all regulated point source discharges including wastewater treatment plant discharges, major industrial point source discharges, and municipal separate storm sewer system (MS4) stormwater discharges. For some sources, particularly MS4 stormwater discharges, where measured values are not available, apply estimates. [*Entire LIS Watershed*]

The source data and location where the data may be obtained are as follows:

- Data compiled from National Pollutant Discharge Elimination System (NPDES) Integrated Compliance Information System (ICIS) database and states
- Data submitted by state to EPA for progress under TMDL
- Large, direct discharging wastewater treatment facilities (a.k.a. municipal waste water treatment plants (MWWTP); water pollution control facilities (WPCF) discharging to the open waters of LIS
- MS4 sources
- Industrial discharge information

Part C
Summarize tributary nitrogen loads using published monitoring and modeling results. Identify uncertainty estimates where available as part of loading estimates. [*Large Riverine only*]

The source data and location where the data may be obtained are as follows:

- USGS data
- SWEM model outputs for nearshore waters and embayments
- NE SPARROW Model

- AVGWLF Model
- Dr. Evans AVGWLF work

Part D
Summarize existing water quality data (nutrients and primary response variables) from the following sources, including but not limited to: [*All watershed groupings*]

- Connecticut Department of Energy and Environmental Protection
- EPA National Coastal Assessment
- Long Island Sound Integrated Coastal Observing System
- Suffolk County
- New York City Department of Environmental Protection
- Interstate Environmental Commission
- Dr. Gobler, Stony Brook University
- SWEM model outputs for nearshore waters and embayments
- Datasets from local watershed groups

Part E
Relate tributary loads to areas of influence using both SWEM outputs and other work (e.g. Dr. Whitney, University of Connecticut), to track how waters from the major rivers are distributed throughout LIS. This will be accomplished using the hydrodynamic components of the SWEM or New York Harbor Observing and Predicting System (NYHOPS) models and applying a particle tracking routine to the hydrodynamic model. The particle tracking model output will yield the relative average contribution of water from different local tributaries and the mixed water of central LIS to each grid cell in specific areas of interest such as embayments, in the hydrodynamic model domain [*Large Riverine and Western LIS*].

- SWEM
- University of Connecticut
- NYHOPS http://hudson.dl.stevens-tech.edu/maritimeforecast/

Part F
Compare both nitrogen loading rates and nitrogen concentrations to LIS chl-a concentrations using both water quality data and model outputs. [*All watershed groupings*]

- Assess relationship between nitrogen and chl-a
- Identify nitrogen conditions that would result in chl-a levels supportive of light levels suitable
    - To restore eelgrass habitat to 2000 acres to meet LIS Comprehensive Conservation and Management Plan (CCMP) target goal
    - To support eelgrass where it historically occurred

- If system is macroalgae dominated, relate nitrogen to macroalgae and identify levels of nitrogen suitable to eliminate use impairment for macroalgae

Part G

Recommend nitrogen thresholds for all watershed groupings. For specific embayments consider designated uses, flushing rate, and embayment area in setting thresholds. In recommending thresholds for LIS, consider the applications of: [*All watershed groupings*]

- Existing thresholds across the region and LINAP thresholds under development
- LIS-specific thresholds based on relationship between chl-a and nitrogen (based on F above)

Part H

Estimate reduction levels by prioritized watersheds necessary to meet LIS-specific, regional, and developing LINAP thresholds. [*All watershed groupings*]

- Where threshold is nitrogen load based, compare current loads to threshold loads and identify reductions
- Where threshold is nitrogen concentration-based, assess SWEM or other suitable model outputs to identify load reductions to achieve nitrogen loading threshold
- Summarize existing model outputs or reports on estimating residence time for coastal embayments (e.g. Dr. Whitney, University of Connecticut)
  - Residence time estimates not available for all embayments, where available it should be used to adjust allowable nitrogen loading thresholds
  - Where it is not available thresholds should be applied without consideration for residence time

Part I

Propose allocations among categories of nitrogen sources for Primary Tier Embayments (in the Base Period) and Secondary Tier Embayments (in Option Period 1) (as described in Subtask 3.B), distinguishing between regulated and nonregulated sources. [*Embayments only*]

Subtask 3A. Literature Review

Tetra Tech will perform a comprehensive literature review of the technical references provided herein, and other related references that EPA has provided (e.g., at the Project Kickoff Meeting) and/or which Tetra Tech in its professional judgment may recommend for the project.

The literature review will help Tetra Tech better understand the project objectives and clarify the science underlying the technical approach, identify data gaps and/or matters requiring future empirical data collection, and/or other matters or issues not at this time transparent but which an independent critical review could afford to the Project and EPA Project Team.

Tetra Tech will summarize its literature review in a detailed Literature Review Memorandum (LRM).

Subtask 3A Deliverables

The LRM will be provided in draft within three (3) months of the Subtask 1A Kickoff Meeting, coincident with the Task 2 QAPP deliverable.

Subtask 3B. Application of Technical Approach to Priority Waters within Watershed Groupings

Based in part on its literature review in Subtask 3A and following QAPP approval by the QAU, for this Subtask Tetra Tech will provision to work cooperatively with EPA and its stakeholders to apply the technical approach to derive nitrogen thresholds and propose allowable load allocations for a number of priority waters within each of the three watershed groupings, as follows:

Primary Tier Watershed Groupings (Base Period)

Embayments

There are roughly 110 coastal subwatersheds (i.e., embayments) to LIS. However, EPA has identified a small subset of priority embayments. These priority embayments are:

Connecticut

- Stonington Harbor / Pawtucket River
- Saugatuck Estuary
- Norwalk Harbor
- Mystic Harbor
- Niantic Bay
- Farm River
- Southport Harbor / Sasco Brook

New York

- Northport-Centerport Harbor Complex
- Port Jefferson Harbor
- Nissequogue River
- Stony Brook Harbor
- Mt. Sinai Harbor

Large Riverine Systems

There are three major tributaries representing the large riverine watersheds of the LIS. These are the Connecticut, Housatonic, and Thames Rivers. For this task, Tetra Tech will consider the Connecticut River.

Western Long Island Sound

This is the open water areas defined by the Long Island Sound Report Card as the Eastern Narrows and Western Narrows in the western LIS. For this task, Tetra Tech will consider the western LIS as defined above.

Secondary Tier Watershed Groupings (Option Period 1)

EPA has identified a small subset of watersheds as a secondary priority tier. These secondary tier watersheds, to be considered in the following sequential order, are:

Large Riverine

- Housatonic River
- Thames River

Embayments

- Mamaroneck River
- Hempstead Harbor
- Specific areas adjacent to the Northport/Centerport Harbor Complex (Huntington Bay, Lloyd Harbor, Duck Island Harbor)
- Oyster Bay – Cold Spring Harbor Complex
- Manhasset Bay
- Pequonnock River
- Byram River, CT
- Sasco Brook
- New Haven Harbor

EPA has chosen the above-referenced priority and secondary tier watersheds in part because it anticipates this Subtask will require a collaborative process of iterative refinement where application of the technical approach to the very first one or few of the priority watersheds will require more substantial effort, and that once the technical approach becomes refined, application of the technical approach to later watersheds can proceed in an increasingly efficient manner.

1.6.2 Type and Amount of Existing Data That Will Be Gathered and How Data Sources Will Be Evaluated for Use

As described in Section 1.6.1, Tetra Tech will compile sources provided by EPA's Project Team. Tetra Tech will also use the following steps to identify additional data sources:

- Determine the universe of sources that are acceptable (e.g., peer reviewed journal articles, federal/state reports, university research)
- Determine search tools that are acceptable (e.g., Google Scholar, EBSCO Host, professional society websites)
- Select and document key search terms
- Perform searches

Depending on the types of information found using initial search criteria, Tetra Tech might need to update them (e.g., key search terms might need to be added, additional sources types might need to be searched). The search criteria will be documented and updated in the LRM.

Tetra Tech expects to collect information on LIS water quality, including information on nitrogen loads and sources from each coastal embayment; flow, TN load, and TN concentrations for all regulated point source discharges; tributary nitrogen loads; existing water quality data (nutrients and primary response variables) for all watershed groupings; and information relating tributary loads to areas of influence to track how waters from the major rivers are distributed throughout LIS. Tetra Tech will inventory the data it collects and the data provided by EPA, and subsequently identify the types of available data and corresponding geographical and temporal representation. Tetra Tech, in consultation with EPA, will determine how the data sources will be selected or rejected for use based on the temporal and geographical limits, and quality of the data sources.

Tetra Tech will primarily rely on data that have been published in federal, state, or local government reports; peer-reviewed journal articles, as well as data cited in one of these existing data sources. Tetra Tech will also evaluate data in consultation with EPA from additional sources (e.g., unpublished university research data, data from volunteer monitoring groups) provided by or obtained from discussions with federal or state employees. Tetra Tech expects that project-specific QAPPs or similar documentation describing the performance criteria evaluated and met will be available for the federal, state, and local government sources, and from additional sources provided by or obtained from discussions with federal or state employees. If this documentation is not readily available, Tetra Tech, in consultation with the EPA TOCOR and Project Team will determine how much effort should be expended to find reports or metadata that might contain that information.

Tetra Tech will summarize its literature review, including the type and amount of data collected (e.g., geographical representation, temporal representation), in a detailed LRM. Tetra Tech will also describe its approach for evaluating data (e.g., data source types, QAPPs or similar QA documentation available for a particular source) for use in supporting this project in the LRM.

1.7 Quality Objectives and Criteria

Data of known and documented quality are essential to the success of projects that involve preparing and applying technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers. For Task Order 0023, Tetra Tech will collect information on nitrogen loads and sources from each coastal embayment; flow, TN load, and TN concentrations for all regulated point source discharges; tributary nitrogen loads; existing water quality data (nutrients and primary response variables) for all watershed groupings; and information relating tributary loads to areas of influence to track how waters from the major rivers are distributed throughout LIS. Project quality objectives and criteria for existing data are discussed below.

1.7.1 Project Quality Objectives

The basic requirements of this project are that Tetra Tech must ensure that the Task Order tasks are performed in conformance with the procedures detailed in this QAPP and any revisions and tasks results are reproducible and transparent. The conclusions provided will be logical, consistent, and defensible.

Tetra Tech will follow the practices of using existing data described in EPA's (2002a) *Guidance for Quality Assurance Project Plans, EPA QA/G-5*, EPA's (2009a) *New England QAPP Guidance for Projects Using Secondary Data, Revision 2*; the *EPA New England Environmental Data Review Program Guidance* (USEPA 2013a), and the *EPA New England Environmental*

Data Review Supplement (USEPA 2013b). Tetra Tech expects to provide support to EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS. Tetra Tech will inventory the data it collects and the data provided by EPA, and subsequently identify the types of available data and corresponding geographical and temporal representation. Tetra Tech, in consultation with EPA, will determine how the data sources will be selected or rejected for use based on the temporal and geographical limits, and quality of the data sources.

Tetra Tech will describe how data sources were evaluated for relevance and quality in a detailed LRM. As work requirements dictate, Tetra Tech will add task-specific standard operating procedures in an appendix to this QAPP when necessary. The Tetra Tech Co-TOLs will communicate the level of QA/QC review associated with each deliverable in the monthly progress reports.

1.7.1.1 Compilation of Geospatial Data

Tetra Tech will review geospatial data in data sets used for the project to ensure that they conform to data exchange protocols and applicable data standards as defined and maintained by EPA OEI. Sampling stations will be mapped to verify that the data fall within the correct waterbody.

Tetra Tech will also check whether metadata are available for geospatial data used in this project. If Tetra Tech determines that a particular data set or data layer obtained from a secondary source is provided at a scale that does not fit into the overall dataset or area of interest, Tetra Tech (in consultation with the EPA TOCOR) will describe the extent to which this particular data set can be used or applied in the metadata file.

1.7.1.2 Acceptance Criteria for Quantitative Data

Tetra Tech will prepare high quality deliverables. Tetra Tech will ensure that written deliverables are edited for grammar, spelling, and logic flow. Technical information will be complete and presented in a logical, readable manner.

To the extent possible and appropriate, most of the data sets and literature will be taken from peer-reviewed journals or peer-reviewed books and reports. Therefore, it is assumed that the works have already been screened for appropriateness and rigor of study design and methods and for appropriate application of statistical analyses and modeling methods. Data will also be taken from technical reports prepared by, or field work conducted by, EPA, another federal agency, a state, a research organization, or a contractor working for EPA. Such reports will be considered credible if appropriate QA/QC procedures were followed and documented.

Relevance to the study— Tetra Tech is primarily concerned in this project with summarizing data and literature to provide support to EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS. As stated, sources of these data will include peer-reviewed scientific papers and other sources, including technical reports. This information must be in good electronic condition, and require only minimal QA/QC to standardize among the available sources.

Representative of the areas and times of study—Tetra Tech will only include data collected under a documented quality program. Both recent and long-term data will be considered in consultation with the EPA TOCOR and Project Team. Tetra Tech will limit the temporal extent of data retrieval and assembly from 1/1/2006 to 12/31/2015 to allow for interannual variability in the characterization of current nitrogen loads and concentrations in an average year. Tetra Tech expects that this timeframe will encompass the range of different hydrological conditions in the watershed groupings.

Individual observations: anomalous or extreme outliers—Individual data values might be in error because of transcription errors or equipment malfunctions. If the error results in an anomalous or unrealistic value, it can be detected and excluded from analysis. Tetra Tech will examine the data for anomalous values and reject values reported well beyond the range of observed variability. Tetra Tech will document the number of exclusions, the source of the data excluded, and the suspected cause of error or rationale for exclusion.

1.7.2 Discussion of Data Attributes

Data sets will be reviewed relative to the general precision, accuracy, representativeness, comparability, and completeness (PARCC) quality characteristics of the data provided or available for download, in accordance with *EPA's Guidance for Quality Assurance Project Plans* (USEPA 2002a). These attributes are described below.

Precision—Precision of data reported in the literature is generally unknown. Unless there is compelling evidence in an individual publication that the precision of the data are unacceptable, it is assumed that the data reported are of precision sufficient for use in this project. Tetra Tech will review geospatial data for this project to ensure that they meet a minimum accuracy of Geospatial Accuracy Tier 5 (i.e., 101 – 200 meter precision) to maximize the potential for secondary users.

Accuracy—Accuracy of data reported in data compilations is generally unknown. Unless there is compelling evidence that the data were collected incorrectly and are inaccurate, it is assumed that the data reported are of sufficient accuracy for use in this project. Tetra Tech will review geospatial data for this project to ensure that they meet a minimum precision of Geospatial Accuracy Tier 5 (i.e., 101 – 200 meter accuracy) to maximize the potential for secondary users.

Representativeness—See the discussion of representativeness in Section 1.7.1 above.

Completeness—Whenever possible, data will be downloaded electronically from various electronic sources to reduce scanning of hard copy documents. Tetra Tech technical staff will develop dedicated hard copy and electronic files. In addition, the following steps for assigning staff and general project procedures will be used to ensure the completeness and correctness of data used in the deliverables:

- All original work performed by any member of the technical staff will be subject to QC checks by a different member of the technical staff who is capable of performing the QC checks.
- All QC reviews will be documented.
- Use of data evaluation factors and limits.
- Members of the Tetra Tech staff are capable of collecting technical data and managing data sources.
- The Tt Co-TOLs will maintain a continuing dialog with the EPA TOCOR and Project Team on technical issues, including discussions regarding the inclusiveness and comprehensiveness of the data sources collected for the project.
- The Tt Co-TOLs will provide answers to specific management questions.

Comparability—Comparability is an issue when different sampling or analytical methods or different GIS file formats are used among different studies that are being combined among parts of a single study. It is expected that information provided from various monitoring studies and various file formats will be synthesized; for interpretation of these data, Tetra Tech will ensure that different values are expressed in comparable units and that differences among values are clearly tied to how those values were developed.

Tetra Tech, in consultation with EPA, will collect information on LIS water quality, including information on nitrogen loads and sources from each coastal embayment; flow, TN load, and TN concentrations for all regulated point source discharges; tributary nitrogen loads; existing water quality data (nutrients and primary response variables) for all watershed groupings; and information relating tributary loads to areas of influence to track how waters from the major rivers are distributed throughout LIS. Tetra Tech will inventory the data it collects and the data provided by EPA, and subsequently identify the types of available data and corresponding geographical and temporal representation. Tetra Tech, in consultation with EPA, will determine how the data sources will be selected or rejected for use based on the temporal and geographical limits, and quality of the data sources. This information will be documented in the LRM.

No unpublished data will be cited or included in the final report, with the exception of such data that are recommended by EPA for inclusion in the report. In the rare case that a paper or study represents a technical landmark and is considered seminal within a topic/subtopic but has not been peer reviewed or some aspect of study design, methods, or support of results or conclusions is reviewed and found to be poor, that will be documented and discussed (as a caveat) in the draft deliverables when the associated results are presented.

All the above described procedures will be enforced using the approach described in this section.

1.7.3 Data Management/Handling

The methods and procedures described in this document are intended to reduce the magnitude of measurement error sources and the frequency of error occurrence. The relevant quality objectives for this project include existing data quality and data handling. The project quality objectives related to data handling include the following:

- Maintaining and documenting (in monthly progress reports) a continuing dialog with the EPA TOCOR on technical issues, as appropriate
- Providing answers to specific questions from the TOCOR
- Establishing quality targets while recognizing the limits of the data
- Documenting and presenting results in the form of draft and final deliverables
- Ensuring that the following information used to develop the final deliverables is documented in the project files:
    - Data sources used
    - Assumptions employed
    - Methods applied to address complications with merging data for analyses
    - Statistical procedures employed

Measurement performance accuracy criteria commonly used by Tetra Tech for various data handling parameters are presented in Table 1.

Table 1. Measurement Performance Criteria[a]

Measurement parameter	Accuracy	Description
Data entry	≤ 5% incorrectly entered data	20 percent of manual data entries will be checked. All errors will be corrected.
Extraction/interpretation of pertinent data from existing sources for use in deliverables	≤10% incorrectly extracted/interpreted data	10 percent of extraction/interpretations of pertinent data from existing sources for use in deliverables will be checked. All errors will be corrected.

[a] Analytical truth is unknown for precision.

Data Transfers

As values are transferred to a spreadsheet (data entry) or into a deliverable (word processing), the person performing this action will review the transfer for accuracy and write the complete citation for the source of the data, following the appropriate style guide format, with the short citation entered in the database, text, or footnote, as appropriate (e.g., author, year). All data will have its source identified during deliverable preparation.

A Tetra Tech QC Officer or qualified designee will independently check transferred data using a standard level review. This standard review consists of independently checking each different file type (i.e., a file with different structure or legacy), confirming the first, last, and a selected middle portion of the data were transferred correctly. Evaluating the first and last portions of data helps confirm that no records were accidentally dropped. Selection of the middle portion of the data to check will be done by targeting unique and unusual record types that might stress the transfer process. More files (up to 10 percent) will be reviewed if files are processed individually while fewer checks (no less than 2 data files of each type) will be used for automated to semi-automated procedures. All identified data transfer errors will be corrected and the Tetra Tech QC Officer or his designee will perform a follow-up review of the corrected components to ensure that the errors have been corrected.

Data Calculations

A Tetra Tech QC Officer or qualified designee will perform standard-level reviews of data calculations (including conversions) for Task Order 0023. A standard-level of review consists of up to 10 percent independent recalculations of computations and graphs, but no less than 2 examples of each type of computation and 2 examples of each graphic type. More calculations (up to 10 percent) will be reviewed if data sets or points are processed individually while fewer checks (no less than 2 examples of each type of computation and 2 examples of each graphic type) are appropriate for automated to semi-automated procedures. Selection of which calculation types and graphs to check will include targeting unique and unusual record types that might stress the calculation and graphing process. All identified data calculation errors will be corrected and the Tetra Tech QC Officer or his designee will perform a follow-up review of the corrected components to ensure that the errors have been corrected.

Uncertainty in the data due to errors introduced during data manipulation could result in providing incorrect results. Reducing data uncertainty is of highest priority, and it is important to reduce uncertainty by using QC protocols. Tetra Tech staff will ensure secondary data management, statistical calculations, and geospatial data management are performed in accordance with *Tetra Tech's Quality Assurance and Quality Control Considerations for Secondary Data Management* (Appendix A), *Tetra Tech's Quality Assurance and Quality Control Considerations for Statistical Analyses* (Appendix B), and *Tetra Tech's Geospatial and Data Management QA/QC Procedures* (Appendix C), respectively.

All the above described procedures will be enforced using the approach described in this section of the QAPP.

2.0 DATA SELECTION AND MANAGEMENT

2.1 Sources of Existing Data

Tetra Tech expects to provide support to EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS. To help provide this support, Tetra Tech expects to collect information from the following sources:

- Vaudrey research
- The Nature Conservancy
- LINAP
- USGS
- NPDES ICIS database
- Data submitted by states to EPA for progress under a TMDL
- MWWTP
- WPCF discharging to the open waters of LIS
- MS4 sources
- Industrial discharge information
- SWEM model outputs for nearshore waters and embayments

- NE SPARROW Model
- AVGWLF Model
- Dr. Evans AVGWLF work
- Water quality data from state and local agencies, including Connecticut Department of Energy and Environmental Protection, EPA National Coastal Assessment, Long Island Sound Integrated Coastal Observing System, Suffolk County, New York City Department of Environmental Protection, and Interstate Environmental Commission
- Dr. Gobler, Stony Brook University
- Millstone Nuclear Power Plant monitoring of Niantic Bay
- Water quality datasets from local watershed groups
- Work from Dr. Whitney, University of Connecticut

Tetra Tech will inventory the data it collects and the data provided by EPA, and document the following information in the LRM:

- Type of data and collection dates
- Originating organization
- Report title, author and date
- Database names

As described in Section 1.7, Tetra Tech, in consultation with EPA, will determine how the data sources will be selected or rejected for use based on the temporal and geographical limits, and quality of the data sources. This information will be documented in the LRM.

Whenever possible, data will be downloaded electronically from various electronic sources to reduce scanning of hard copy documents. Tetra Tech technical staff will develop dedicated hard copy and electronic files.

Minimizing errors by using QC protocols is of the highest priority for this project. The following steps for assigning staff and general project procedures will be used to ensure the completeness and correctness of data used in the deliverables:

- All original work performed by any member of the technical staff will be subject to QC checks by a different member of the technical staff who is capable of performing the QC checks.
- All QC reviews will be documented.
- Data evaluation factors and limits will be used and documented in the LRM.
- Tetra Tech technical staff will follow *Tetra Tech's Quality Assurance and Quality Control Considerations for Secondary Data Management* (Appendix A), *Tetra Tech's Quality Assurance and Quality Control Considerations for Statistical Analyses* (Appendix B), and *Tetra Tech's Geospatial and Data Management QA/QC Procedures* (Appendix C).
- Members of the Tetra Tech staff are capable of collecting technical data and managing data sources.

- The Tetra Tech Co-TOLs will maintain a continuing dialog with the EPA TOCOR and Project Team on technical issues, including discussions regarding the inclusiveness and comprehensiveness of the data sources collected for the project.
- The Tetra Tech TOCOR will provide answers to specific management questions.

Tetra Tech will manage data collected for each major task or subtask for this project in Microsoft Excel spreadsheets. If needed, Tetra Tech will use Microsoft Access for larger data compilations. Tetra Tech will transmit intermediate data products to EPA using email and temporary urls. Because the total amount of water quality sampling data, GIS layers, and modeling output data for this project may result in excess of 500 megabytes of files, Tetra Tech anticipates submitting key project data deliverables on external hard disks to EPA or through an EPA SharePoint or comparable secure file sharing platform as directed by EPA.

The Tetra Tech Co-TOLs will store all computer files associated with the project in a project subdirectory (subject to regular system backups) in the Fairfax, Virginia, and Research Triangle Park, North Carolina, offices, and will copy the files to disk for archive for 3 years subsequent to project completion (unless otherwise directed by EPA). Tetra Tech's computer systems and network servers are specified, inventoried, and maintained (either directly or by external service contract) by a dedicated information technology staff. All interim (working) and final electronic files used in development of project deliverables will be stored on Tetra Tech's secure network which is backed up daily.

2.2 Intended Use of Existing Data

Data from available studies will be used to provide support to EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS. As described in Section 1.6.1, Tetra Tech expects to use existing data sources for specific purposes, including those sources listed below.

A. Nitrogen loads and sources from each LIS coastal embayment
    - Vaudrey research
    - The Nature Conservancy
    - LINAP
    - USGS

B. Flow, TN load, and TN concentration for all regulated point source discharges, including wastewater treatment plant discharges, major industrial point sources discharges, and MS4 stormwater discharges, for the entire LIS watershed
    - Data compiled from NPDES ICIS database and states
    - Data submitted by state to EPA for progress under TMDL
    - Large, direct discharging wastewater treatment facilities discharging to the open waters of LIS
    - MS4 sources
    - Industrial discharge information

C. Tributary nitrogen loads for large riverine only
    - USGS data (Mullaney 2016; USGS monitoring estimates for the Connecticut River; Essex and Old Lyme monitoring stations in brackish waters of lower Connecticut River and LIS)
    - SWEM model outputs for nearshore waters and embayments
    - NE SPARROW model
    - AVGWLF model

D. Existing water quality data (nutrients and primary response variables) for all watershed groupings
    - Connecticut Department of Energy and Environmental Protection
    - EPA National Coastal Assessment
    - Long Island Sound Integrated Coastal Observing System
    - Suffolk County
    - New York City Department of Environmental Protection
    - Interstate Environmental Commission
    - Dr. Gobler, Stony Brook University
    - SWEM model outputs for nearshore waters and embayments
    - Millstone Nuclear Power Plant monitoring of Niantic Bay
    - Datasets from local watershed groups

E. Data used to relate tributary loads to areas of influence using both SWEM outputs and other work for large riverine and western LIS
    - SWEM
    - University of Connecticut
    - New York Harbor Observing and Predicting System (NYHOPS)

Tetra Tech will summarize its literature review in a detailed LRM, including how and when data that are found to have limitations (e.g., qualified lab data) will be used in the project.

2.3 Limitations on the Use of Existing Data

Tetra Tech will evaluate existing data sources using the selection criteria described in Section 1.7 of this QAPP. To ensure a high level of professional and technical judgment in data set and paper review and selection, experienced technical staff in each topic will conduct the data and literature search, selection, and data set and paper review. If, despite the use of peer-reviewed sources, substandard, out-of-date, or poorly supported data sources or papers are found, they will be rejected from inclusion in the final selection of papers for review on the basis of the lead reviewer’s judgment that they are of poor quality, supported by applying the selection criteria listed above. In the rare case that a paper or study represents a technical landmark and is considered seminal within a topic/subtopic but has not been peer reviewed or some aspect of study design, methods, or support of results or conclusions is reviewed and found to be poor, that will be documented and discussed (as a caveat) in the draft deliverables when the associated results are presented. Tetra Tech will consider the following series of general questions (Boslaugh 2007) when evaluating data collected for this project:

- What was the original purpose for which the data were originally collected?
- What kind of data are they, and when and how were the data collected?
- What data processing and/or recording procedures have been applied to the data?

Tetra Tech will also consider the following series of questions more specific to water quality data when evaluating a water quality sources (USEPA 2009a):

- Were the data generated under an approved QAPP or other documented sampling procedure?
- When multiple data sets will be combined, were the data sets generated using comparable sampling and analytical methods?
- Were the analytical methods sensitive enough (detection limits) to meet project needs?
- Is the sampling method indicated (e.g., grab, composite, calculated)?
- Was the sampling effort representative of the waterbodies of interest in a random way, or could bias have been introduced by targeted sampling?
- Are the data qualified? Are sampling and laboratory qualification codes or comments included? Are the qualification codes defined?
- Is sufficient metadata available about variables like sampling station location, date, time, depth, rainfall, or other confounding variables?

Tetra Tech will summarize its literature review in a detailed LRM, including the limitations on the use of existing data collected for this project.

3.0 ASSESSMENTS AND OVERSIGHT

The QA program under which this project will be performed could include performance and system audits, with independent checks of the data obtained from data-gathering activities and subsequent analyses. Final versions of any deliverable that is to be published or widely distributed by EPA must be reviewed by the Tetra Tech Co-TOLs or authorized designee and a qualified editor or authorized designee to ensure that all revisions have been properly made and that the deliverable is consistent with overall contract goals and requirements and does not contain information that could expose EPA to liability.

The Tetra Tech QC Officers will assist the Tetra Tech Co-TOLs in conducting internal reviews of the deliverables prepared for this project. They will identify and document all issues that could affect quality and make recommendations for improving the reports. When internal reviews have been completed, the Tetra Tech TOL will submit written deliverables to a qualified technical editor for editorial review to ensure that the writing is clear and concise; that the written material conforms to predetermined requirements for format, style, and usage; and that the terms, resources, and format are used consistently throughout the document. The technical editor will sign and date the Technical/Editorial Review form (Figure 2) or equivalent documentation and equivalent documentation signed by the internal technical and editorial reviewers will be attached to the marked-up page(s) and filed in the project file.

TECHNICAL/EDITORIAL REVIEW

Contract (name) ______________________

Date ______________ DCN: ______________

Title (project or other) ______________________________

Project Leader ______________________ TC# ______________

Author(s) ______________________________

Version: Draft ____ or Final ____ AND Entire ____ or Part (Specify) ______________

Special Instructions for the Reviewer(s) (attach additional page if needed):

Return this form and the attached deliverable on which you have written comments to the Project Leader by the date indicated.

Reviewer ______________________ Type of Review (circle one): Technical Editorial

Date to Reviewer __________ Date to Be Returned __________ Date Comments Received __________

Reviewer (signature) ______________________ Date ______________

Overall Evaluation: ☐ Acceptable ☐ Acceptable With Minor Revision
☐ Needs Substantial Revision ☐ Needs Additional Review

Notes:

Reviewer ______________________ Type of Review (circle one): Technical Editorial

Date to Reviewer __________ Date to Be Returned __________ Date Comments Received __________

Reviewer (signature) ______________________ Date ______________

Overall Evaluation: ☐ Acceptable ☐ Acceptable With Minor Revision
☐ Needs Substantial Revision ☐ Needs Additional Review

Notes:

Reviewer ______________________ Type of Review (circle one): Technical Editorial

Date to Reviewer __________ Date to Be Returned __________ Date Comments Received __________

Reviewer (signature) ______________________ Date ______________

Overall Evaluation: ☐ Acceptable ☐ Acceptable With Minor Revision
☐ Needs Substantial Revision ☐ Needs Additional Review

Notes:

Original form and attached comments to be filed in Project File.

Figure 2. Example technical and editorial review form

The essential corrective action steps in the QA program for addressing any problems that could occur during the project (as presented in Figure 3) are as follows:

- Identify and define the problem.
- Assign responsibility for investigating the problem.
- Investigate and determine the cause of the problem.
- Identify the corrective action.
- Assign and accept responsibility for implementing appropriate corrective action.
- Establish the effectiveness of and implement the corrective action.
- Verify that the corrective action has eliminated the problem.

Many of the technical problems that might occur can be solved on the spot by the staff members involved; for example, by modifying the technical approach or correcting errors or deficiencies in documentation. Immediate corrective actions form part of normal operating procedures and are noted in records for the project. Problems not solved this way require more formalized, long-term corrective action.

If quality problems that require attention are identified, the project team will determine whether attaining acceptable quality requires short- or long-term actions. If a failure in an analytical system occurs (e.g., performance requirements are not met), the Tetra Tech QC Officer will be responsible for corrective action and will immediately inform the Tetra Tech Co-TOLs. Subsequent steps taken will depend on the nature and significance of the problem, as illustrated in Figure 3.

The Tetra Tech Co-TOLs have primary responsibility for monitoring the activities of this project and identifying or verifying correction of any quality problems. These problems will also be brought to the attention of the QA Officer, who will either initiate or verify the corrective action system components described above, document the nature of the problem (using a form such as that shown in Figure 4), and ensure that the recommended corrective action is carried out. The QA Officer has the authority to stop work on the project if problems that affect data quality and require extensive effort to resolve are identified. The Tetra Tech Co-TOLs and the EPA TOCOR will be notified of major corrective actions and stop work orders. The Tetra Tech Co-TOLs have primary responsibility for monitoring the activities of this project and identifying and verifying that any quality problems are sufficiently investigated, that appropriate solutions are evaluated, and that corrective actions are implemented, verified to be adequate to address the problem(s), and documented in project reports.

Failure to meet any QC requirements requires that appropriate corrective actions be taken. All major QC failures and associated corrective actions (and their effectiveness) will be documented on a Corrective Action Request and Response Verification form (Figure 4) and submitted to the EPA TOCOR. Data associated with QC problems will be clearly identified, along with an assessment as to the potential effects(s) of the QC failure on data quality. The Tt Co-TOLs or QA Officer will notify the EPA TOCOR of such problems/corrective actions as soon as possible after the actual occurrence.

Tetra Tech will provide to the EPA TOCOR any written reports generated on routine surveillance.

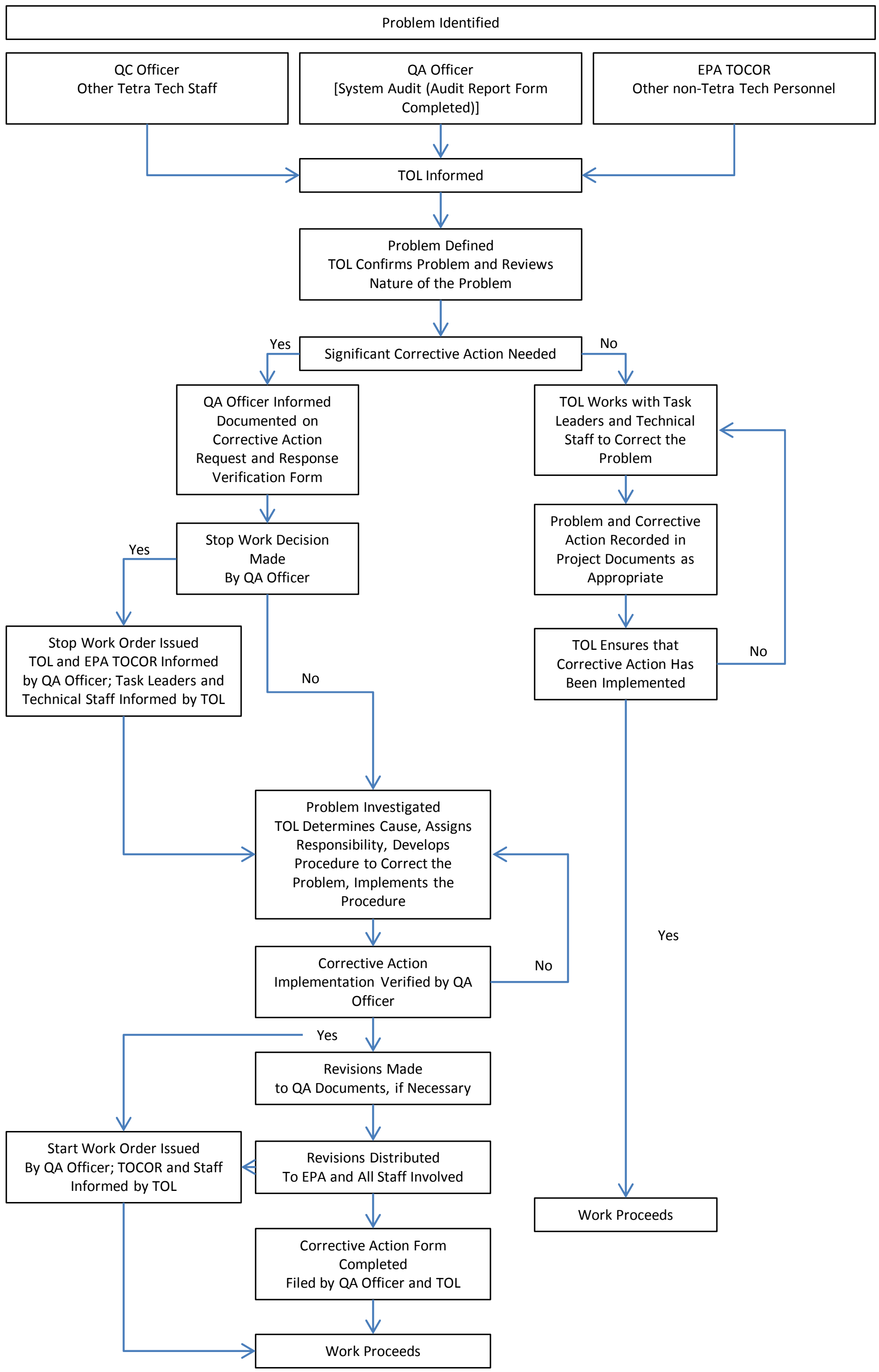

Figure 3. Corrective Action Process

CORRECTIVE ACTION REQUEST AND RESPONSE VERIFICATION

Contract (name) ____________________________

Date of Assessment __________________ Request No. _________________

Title (of project or other) ____

Project Leader ____ TC# ________________________

Other Responsible Personnel ____

Auditor or Initiator of This Corrective Action Request _____________________________________

Problem Description:

Recommended Action: **Date to Be Completed:**

_________________________________ _________________________
Quality Assurance Officer Date

_________________________________ _________________________
Principal-in-Charge or Program Manager Date

Action Taken: **Date:**

Verification of Completion of Corrective Action:

_________________________________ _________________________
Quality Assurance Officer Date

_________________________________ _________________________
Principal-in-Charge or Program Manager Date

Original form to be filed in QAO File; one copy to be filed in Project File and one copy in Contract File (if corrective action pertains to a project), or one copy to be filed in Contract File (if corrective action pertains to a contract).

Figure 4. Example Corrective Action Request and Response Verification form

4.0 DATA REVIEW – VERIFICATION, VALIDATION AND EVALUATION

4.1 How Project Members Will Review and Verify or Validate the Adequacy of Each Data Set Relative to Established Acceptance Criteria

Data quality is addressed, in part, by consistently performing valid procedures. It is enhanced by the training and experience of project staff (Section 1.4) and documentation of project activities (Section 6.0). This QAPP and other supporting materials will be distributed to all project personnel. The Tetra Tech QA Officer and QC Officers will ensure that tasks described in the approved Work Plan are carried out in accordance with the QAPP. The Tetra Tech QC Officers will evaluate the inclusiveness and comprehensiveness of the data sources collected for this project, and check electronic data transfers and data calculations (including conversions), as described further in this section of the QAPP. The Tetra Tech Co-TOLs will review staff performance throughout the project to ensure adherence to project protocols.

The principal QA/QC requirements for gathering existing data for this Task Order are (1) collecting inclusive and comprehensive information from data sources to support EPA in refining and applying a technical methodology for the numeric interpretation of narrative nitrogen criteria to inform permit writers for three watershed groupings (embayments, large riverine systems, and western LIS point source discharges to open water) in LIS, and (2) properly screening potential information/data sources for relevance to the project objectives and for quality. Tetra Tech, in consultation with the EPA TOCOR and Project Team, will determine what factors will be evaluated to determine whether the data provided in a secondary data source are acceptable for use in supporting this project. These evaluation factors and corresponding limits will be provided to the EPA TOCOR in the LRM.

Following the selection of literature and data sets for use in analyses, QC activities will include an independent review of data sets/literature selected for use in this Task Order. A Tetra Tech QC Officer will review 10 percent of the data sets/literature collected, screened and selected for use in the Task Order to confirm the data/information are relevant to the project objectives and the data/information meet project standards in terms of data quality. It is anticipated that immediate corrective actions will be implemented if the second review identifies multiple articles/reports that do not seem suitable for use in this analysis. Corrective actions will be taken to ensure that further information collection and screening are improved and to minimize the potential for rework.

Tetra Tech will document in the progress report bullets and in the LRM the quality requirements for data collection. Tetra Tech will include a description of all QC activities and analyses in the corresponding deliverables. Summary statistics and discussion will include the following:

- Quality of secondary data
    - Requirements will be determined in consultation with the EPA TOCOR and documented in the LRM.
- Geospatial data compiled for this project will conform to data exchange protocols and applicable data standards as defined and maintained by EPA OEI as described in Section 1.7.1 of this QAPP.

- Tetra Tech will ensure that metadata are available for all spatial data elements included in the geospatial data compilations for this project in accordance with the provisions of FGDC-STD-001-1998, *Content Standard for Digital Geospatial Metadata*.
- Geospatial data will be screened to ensure that they meet a minimum accuracy of Geospatial Accuracy Tier 5 (i.e., 101 – 200 meter precision and accuracy) to maximize the potential for secondary users. If Tetra Tech determines that a particular data set or data layer obtained from a secondary source is provided at a scale that does not fit into the overall dataset or area of interest, Tetra Tech (in consultation with the EPA TOCOR) will describe the extent to which this particular data set can be used or applied in the metadata file.
- Accuracy of data entries
  - 20 percent of scanned or hand-entered data will be checked (100 percent of discrepancies will be resolved)
- Accuracy of data transfers
  - The Tetra Tech QC Officer or his designee will independently check transferred data using a standard-level review, consisting of independently checking each different file type (i.e., a file with different structure or legacy), confirming the first, last, and a selected middle portion of the data were transferred correctly. More files (up to 10 percent) will be reviewed if files are processed individually while fewer checks (no less than 2 data files of each type) will be used for automated to semi-automated procedures. All identified data processing errors will be corrected and the Tetra Tech QC Officer will perform a follow-up review of the corrected components to ensure that the errors have been corrected.
- Accuracy of extracted/interpreted pertinent data from existing sources
  - 10 percent of data extractions/interpretations will be checked (100 percent of discrepancies will be resolved)
- Accuracy of data calculations (including conversions)
  - The Tetra Tech QC Officer or his designee will perform up to 10 percent independent recalculations of computations (including conversions) and graphs, but no less than 2 examples of each type of computation and 2 examples of each graphic type. More calculations (up to 10 percent) will be reviewed if data sets or points are processed individually while fewer checks (no less than 2 examples of each type of computation and 2 examples of each graphic type) are appropriate for automated to semi-automated procedures. All identified data calculation errors will be corrected and the Tetra Tech QC Officer or his designee will perform a follow-up review of the corrected components to ensure that the errors have been corrected.

If any raw data are received in hard-copy format, they will be entered into a spreadsheet and a Tetra Tech QC Officer will compare 20 percent of data entries to the original hard-copy data sheets. If the percentage of incorrect data entries exceeds 1 percent for a staff member, a Tetra Tech QC Officer will review an additional 20 percent of the data entries performed by that staff member to determine whether the staff member is performing acceptable data entries. If the percentage of incorrect data entries for the additional sources evaluated exceeds 1 percent, a Tetra Tech QC Officer will evaluate 100 percent of the data entries performed by that staff member to ensure the accuracy of the information provided in databases or documents prepared to support the project's deliverables. Any discrepancies in data entries discovered by the Tetra Tech QC Officer will be resolved with the technical staff member who originally performed the data entry during the review process to ensure 100 percent agreement in data entries from the sources.

The accuracy of the transfer of data from electronic sources to the project deliverable will be determined by checking whether the same number of decimal places after the decimal point in the original database have been used and whether the same units from the original database have been used. As described in Section 1.7.3 above, a Tetra Tech QC Officer will independently check transferred data using a standard-level review, consisting of independently checking each different file type (i.e., a file with different structure or legacy), confirming the first, last, and a selected middle portion of the data were transferred correctly. Evaluating the first and last portions of data helps confirm that no records were accidentally dropped. Selection of the middle portion of the data to check will be done by targeting unique and unusual record types that might stress the transfer process. More files (up to 10 percent) will be reviewed if files are processed individually while fewer checks (no less than 2 data files of each type) will be used for automated to semi-automated procedures. All identified data transfer errors will be corrected and the Tetra Tech QC Officer will perform a follow-up review of the corrected components to ensure that the errors have been corrected.

A Tetra Tech QC Officer will perform standard-level reviews of data calculations (including conversions) for Task Order 0023. A standard-level of review consists of up to 10 percent independent recalculations of computations and graphs, but no less than 2 examples of each type of computation and 2 examples of each graphic type. More calculations (up to 10 percent) will be reviewed if data sets or points are processed individually while fewer checks (no less than 2 examples of each type of computation and 2 examples of each graphic type) are appropriate for automated to semi-automated procedures. Selection of which calculation types and graphs to check will include targeting unique and unusual record types that might stress the calculation and graphing process. All identified data calculation errors will be corrected and a Tetra Tech QC Officer or his designee will perform a follow-up review of the corrected components to ensure that the errors have been corrected.

Written deliverables for this project will be written in plain English and reviewed by a technical editor to ensure that the information is understandable to managers.

4.2 How Gathered Data Will Be Evaluated to Ensure They Can Be Used for Project Purposes

Tetra Tech will review data compilation and data value distributions in the final compiled data through analysis of complete data transfer at the time of transfer and by enumerating records in the original data source and the final compilation. Data that are transferred among databases will be checked for completeness at the time of transfer by enumerating the numbers of records in the original and final data sets. Data transfers will be tagged with upload dates and times to accommodate completeness reviews. If data transfer is incomplete, the missing records will be sought and transferred individually if they are valid. A second round of completeness checks will ensue after successive transfers. Once data sets are compiled, the complete set of data value distributions will be analyzed to identify outliers or bimodal distributions that may result from data entry errors or erroneous unit conversions. This will be accomplished by plotting the distributions against expectations of a normal distribution. The Tetra Tech team data analysts are familiar with expected values for all data types that will be collected for this project. Because of this familiarity, they will be able to identify unusual and potentially erroneous values. Outliers and bimodal distributions will be identified and resolved. Valid outliers can occur and will not be eliminated if the experienced analyst thinks they are plausible. Outliers that are not plausible or show a pattern of potential error will be brought to the attention of the original data supplier and will be excluded from analysis until the original data supplier can confirm their validity.

Tetra Tech will summarize its literature review in a detailed LRM, including the limitations on the use of existing data collected for this project. In the rare case that a paper or study represents a technical landmark and is considered seminal within a topic/subtopic but has not been peer reviewed or some aspect of study design, methods, or support of results or conclusions is reviewed and found to be poor, that will be documented and discussed (as a caveat) in the draft deliverables when the associated results are presented.

The Tetra Tech Co-TOLs will perform the final technical review of materials generated under this project. Appropriate senior level staff may also review these documents. The Tetra Tech Co-TOLs will ensure that the type of data, quantity, and quality of sources of data used in these deliverables have been documented, codes reviewed and tested, data transfers checked, and any data quality limitations or uncertainties clearly identified in the deliverables. If any problems are encountered in meeting the data acceptance criteria, the Tetra Tech Co-TOLs will discuss observations with the TOCOR to reconcile the problems, if possible. The EPA TOCOR will also review all deliverables prepared under this Task Order for content accuracy. If EPA determines that Tetra Tech’s analyses are factually inaccurate or if significant technical errors are found in any documents, Tetra Tech will initiate appropriate corrective action investigation, isolate the cause of error, implement appropriate corrective action, and redo the work, if required.

5.0 PROJECT SCHEDULE

The project schedule is presented in Table 2. The final results of the project will likely be used, in part, for EPA decision making. Although the requested deliverables may be presented in a letter report format with appendices, it is presumed the bulk of the deliverables will be data and the results of data analyses. The deliverables might also include Tetra Tech’s recommendations. In any event, the deliverables must be in the form of a clear and comprehensive presentation to facilitate review by the EPA Project Team and stakeholders, and to support an administrative record for decision making.

Table 2. Schedule of Products, Deliverables, and Milestones

Task	Deliverable	Schedule
Task 0. Work Plan and Budget Development		
0	Work Plan and Budget	Work Plan within 30 days of receipt of Task Order (TO)
	Progress and financial reports	Monthly
Task 1. Project Management and Administration		
1A	Kickoff meeting between EPA and Tetra Tech	Within 1 month of TO issuance
	Meeting summary	Within 1 week of Kickoff Meeting
1B	Conference Calls	Monthly
	Meetings	4 in-person meetings including presentation by Tetra Tech
	EPA Project Team Support	As needed Provision
Task 2. Development of a Quality Assurance Project Plan (QAPP)		
2	Draft QAPP	Within 3 months of Task 1A Kickoff Meeting
	EPA Project Team support for developing final QAPP submittal to QAU. See Section 2.6.3.1 of Attachment 1 (PWS) to TSAWP	As Needed Support
Task 3. Application and Refinement of Technical Approach		
3A	Literature Review Memorandum	Within 3 months of Task 1A Kickoff Meeting
3B	Draft summary of embayment loading	2 months of QAPP approval
	Draft summary of regulated point source discharges	2 months of QAPP approval
	Draft summary of tributary loading	2 months of QAPP approval
	Draft summary of water quality data	3 months of QAPP approval
	Draft results for relative tributary and WLIS load	4 months of QAPP approval
	Draft results for comparing both nitrogen loading rates and nitrogen concentrations to LIS chlorophyll-a	5 months of QAPP approval
	Draft results for nitrogen thresholds for all prioritized embayments and watershed groupings	6 months of QAPP approval
	Draft necessary reduction levels for all prioritized embayments and watershed groupings	6 months of QAPP approval
	Draft proposed allocations for all prioritized embayments	7 months of QAPP approval
	Finalize results for all priority tier watersheds	12 to 14 months from TO initiation

6.0 PROJECT REPORTING

The Tetra Tech Co-TOLs will distribute the QAPP to the technical staff working on this project. If any additional QAPP appendices are needed (as determined in consultation with the EPA TOCOR), Tetra Tech will submit these to the EPA TOCOR as they become available. Tetra Tech will provide updates to the appendix (or appendices) as they become available and as any major changes are made to it (them) during the project.

If any changes in the body of the QAPP or attached or cited documents are required during the project, a memo will be sent to each person on the distribution list as soon as possible describing the change(s), following approval by the appropriate persons. The memo(s) will be attached to the QAPP. Also, Tetra Tech will provide a description about QA procedures conducted for each project task in the monthly progress report.

The Tetra Tech Co-TOLs will submit a monthly report to the EPA TOCOR describing the status of the project and deliverables, including a discussion of QA/QC activities, the status of data collection and evaluation efforts, and any significant quality problems encountered and how they were addressed. The Tetra Tech Co-TOLs will communicate the level of QA/QC review associated with each deliverable in the monthly progress reports. In addition, Tetra Tech will document in the LRM and final deliverables the quality requirements for data collection and how Tetra Tech ensured that information collected for this project was as inclusive and comprehensive as possible. If Tetra Tech uses literature sources that provide useful information but do not fully meet all of the criteria described in Section 1.7 of this QAPP, Tetra Tech will cite these sources and fully document the limitations in their use or interpretation in the corresponding project deliverable. The final search criteria and all sources of existing data that were used in the project will be documented in the LRM prepared for this project. The Tetra Tech Co-TOLs will store these sources in project subdirectory (subject to regular system backups) in the Fairfax, Virginia, and Research Triangle Park, North Carolina, offices, and will copy the files to disk for archive for 3 years subsequent to project completion (unless otherwise directed by EPA). Tetra Tech will provide these sources to the EPA TOCOR upon request.

Tetra Tech will prepare geospatial metadata documentation for all spatial data elements included in the geospatial data compilations for this project in accordance with the provisions of FGDC-STD-001-1998, *Content Standard for Digital Geospatial Metadata*. Geospatial data compilations prepared for this project will be stored in conformance with the EPA's Enterprise Architecture and system protocols to maximize accessibility.

Tetra Tech will use project titles for electronic folders to maintain an index of electronic of data. The Tetra Tech Data Geospatial Data Steward will manage the geospatial data folders. Data may be transferred to electronic spreadsheets and software systems for analysis and presentation. The following naming convention of electronic files must be used to facilitate distinction between versions:

> "YearMonth Day" followed by or preceding the file name, example, 20121128 = November 28, 2012.

The Tetra Tech Co-TOLs will maintain a central project files in Tetra Tech's Fairfax, Virginia, and Research Triangle, North Carolina, offices to contain all related documents, reports, communications, data compilations, checklists or other records, and deliverables (electronic files and hard copies [when prepared]); the Co-TOLs will submit draft and final deliverables to the EPA TOCOR. Draft and final reports to be submitted will contain descriptions of the work performed, data sets used in analyses, and output data sets, and analyses outputs. Interim and final electronic files used in the development of project deliverables will be stored on the secure local network server in the project office. In addition, a description of geospatial data management is provided in Section 1.7.1 and a more broad description of overall data management is provided in Section 2.1 of this QAPP. Backup copies of all files are created and stored off-site for business continuity purposes.

REFERENCES

Boslaugh, S. 2007. Secondary Data Sources for Public Health: A Practical Guide. Cambridge University Press.

FGDC (Federal Geographic Data Committee). 1998. *Content Standard for Digital Geospatial Metadata. FGDC-STD-001-1998. Federal Geographic Data Committee*, Metadata Ad Hoc Working Group. Revised June 1998; accessed August 19, 2013. http://www.fgdc.gov/metadata/csdgm/

Mullaney, J.R. 2016. Nutrient, organic carbon, and chloride concentrations and loads in selected Long Island Sound tributaries—Four decades of change following the passage of the Federal Clean Water Act: U.S. Geological Survey Scientific Investigations Report 2015–5189, 47 p., http://dx.doi.org/10.3133/sir20155189.

USEPA (U.S. Environmental Protection Agency). 2001 (Reissued May 2006). *EPA Requirements for Quality Assurance Project Plans*. EPA QA/R-5. EPA/240/B-01/003. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC. http://www.epa.gov/quality/qs-docs/r5-final.pdf

USEPA (U.S. Environmental Protection Agency). 2002a. *Guidance for Quality Assurance Project Plans.* EPA QA/G-5. EPA 240/R-02/009. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC. http://www.epa.gov/sites/production/files/2015-06/documents/g5-final.pdf

USEPA (U.S. Environmental Protection Agency). 2002b. *Guidance for Quality Assurance Project Plans for Modeling*. EPA QA/G-5M. EPA/240/R-02/007. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC. http://www.epa.gov/sites/production/files/2015-06/documents/g5m-final.pdf

USEPA (U.S. Environmental Protection Agency). 2003. *Guidance for Geospatial Data Quality Assurance Project Plans (QA/G-5G*). EPA/240/R-03/003. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC. http://www.epa.gov/sites/production/files/documents/guidance_geospatial_data_qapp.pdf

USEPA (U.S. Environmental Protection Agency). 2005. CIO Policy Transmittal 05-002, Classification No. 2121, Policy Title: EPA National Geospatial Data Policy. U.S. Environmental Protection Agency. http://www.epa.gov/geospatial/docs/National_Geospatial_Data_Policy.pdf

USEPA (U.S. Environmental Protection Agency). 2007. CIO Policy Transmittal 08-004, Classification No. CIO 2131-P-01-0, Policy Title: EPA National Geospatial Data Policy Procedure for Geospatial Metadata Management. U.S. Environmental Protection Agency. http://www.epa.gov/geospatial/docs/2131.pdf

USEPA (U.S. Environmental Protection Agency). 2009a. *New England QAPP Guidance for Projects Using Secondary Data, Revision 2*. U.S. Environmental Protection Agency, New England, Quality Assurance Unit, Office of Environmental Measurement and Evaluation, Boston, MA. https://www.epa.gov/sites/production/files/2015-06/documents/EPANESecondaryDataGuidance.pdf

USEPA (U.S. Environmental Protection Agency). 2009b. *Guidance on the Development, Evaluation, and Application of Environmental Models*. EPA/100/K-09/003. U.S. Environmental Protection Agency, Council for Regulatory Environmental Modeling, Washington, DC. https://www.epa.gov/sites/production/files/2015-04/documents/cred_guidance_0309.pdf

USEPA (U.S. Environmental Protection Agency). 2013a. *EPA New England Environmental Data Review Program Guidance*. U.S. Environmental Protection Agency, New England, Quality Assurance Unit, Office of Environmental Measurement and Evaluation, Boston, MA. https://www.epa.gov/sites/production/files/2016-04/documents/epa-ne-edr-program-guidance.pdf

USEPA (U.S. Environmental Protection Agency). 2013b. *EPA New England Environmental Data Review Supplement*. U.S. Environmental Protection Agency, New England, Quality Assurance Unit, Office of Environmental Measurement and Evaluation, Boston, MA. https://www.epa.gov/sites/production/files/2016-04/documents/epa-ne-edr-supplement.pdf

USEPA (U.S. Environmental Protection Agency). 2016. *Design and Implementation of New Tools for Quality Assurance in Modeling*. U.S. Environmental Protection Agency, U.S. Environmental Protection Agency, New England Office, Boston, MA. https://www.epa.gov/quality/design-and-implementation-new-tools-quality-assurance-modeling

Vaudrey et al. (in prep). Comparative Analysis and Model Development for Determining the Susceptibility to Eutrophication of Long Island Sound Embayments - Final Technical Report. Appendices include embayment portfolios. Report to Connecticut Sea Grant. Long Island Sound Study Research Program.

(This page is intentionally blank.)

Appendix A.

Tetra Tech's Quality Assurance and Quality Control Considerations for Secondary Data Management

Tetra Tech's Quality Assurance and Quality Control Considerations for Secondary Data Management

April 6, 2016

Prepared by

Tetra Tech, Inc.

10306 Eaton Place, Suite 340

Fairfax VA 22030

Contents

Introduction

This guidance provides an overview of secondary data processing and management techniques. The Tetra Tech Project Manager, Data Manager, Quality Assurance (QA) Officer, and Quality Control (QC) Officer should refer to this guidance to ensure that the QA/QC requirements set by our EPA clients are met. The Tetra Tech Project Manager supervises the overall project and is responsible for coordinating project assignments, establishing priorities and schedules; ensuring completion of high-quality projects within established budgets and schedules; providing guidance, technical advice, and evaluating the performance of those assigned to the project; implementing corrective actions; preparing or overseeing preparation and review of project deliverables; and providing support to EPA in interacting with the project team, technical reviewers, and others to ensure that technical quality requirements are met in accordance with EPA's objectives. The Tetra Tech Data Manager is responsible for performing the data processing and management activities and the Tetra Tech QC Officer is responsible for checking those activities. The Tetra Tech QA Officer, with the assistance of the assigned QC Officer, will monitor QC activities to determine conformance with project QA/QC requirements.

Secondary data are data that were collected under a separate effort for some other purpose, whereas primary data are original data collected for a specific project. Secondary data analyses are becoming increasingly common because technological advances have made it possible to store and remotely access large amounts of data. Secondary data processing can be used to further refine and process data compiled from existing data sources. Information on evaluating secondary data sources for quality is provided in the quality assurance project plan (QAPP) or equivalent documentation prepared for a particular project.

Secondary Data Management

This guidance acknowledges that standard practices and protocols vary temporally and differ among various monitoring groups, states, and agencies. Secondary data processing techniques aim to detect and account for inconsistency in a data set compiled from multiple sources. The goal is to improve the comparability and consistency of secondary environmental monitoring data used for a particular project.

Relevant QA/QC practices for secondary data management include ensuring that the data processing steps are correct, documented, well organized, reproducible, and transparent. To ensure that we meet the QA/QC requirements set by our EPA clients, data processing steps must undergo QC reviews and those reviews must be documented in the project files. The Tetra Tech Project Manager and QA Officer will communicate to the Data Manager whether specific documentation of QC reviews is required for a particular task.

The appropriate level of secondary data management and corresponding level of QC review will vary with project goals, available data, resources, and the decisions to be made. At the beginning of a particular project, the Data Manager will provide recommendations to the Tetra Tech Project Manager regarding methods to be used for processing and managing the data. At this time, the Data Manager and Project Manager should discuss the approximate level of effort needed for the various processing steps and corresponding level of QC review required. Follow-up discussions should be held throughout the duration of the project, as needed, to clarify the analyses to be performed, level of QC review needed, and level-of-effort required.

- It should be noted that cursory level compilations of data that are used to inform whether more robust data compilations can be prepared, can be developed with minimum QC review so long as intermediate work products are identified as such in their transmittal to the client and in progress reports.
- Data compilations that directly inform decisions/actions (e.g., remediation, compliance decisions, regulatory action, source control, capital investment) require a higher level of QC review.

This document describes the following topics as related to ensuring the quality of Tetra Tech's secondary data management: data acquisition and documentation, data quality considerations, data organization, and data transformation. A quick reference list of common steps used for data management and processing developed specifically for water quality data, is also included as Attachment 1.

Data Acquisition and Documentation

Data acquisition involves the process of obtaining and documenting data of various types (e.g., water quality sampling data, spatial data, remote sensing imagery, survey results, 303(d) impairment or 305(b) assessment data, TMDLs, discharger data) using search criteria for the project determined in consultation with the client. Data acquisition must be a repeatable and transparent process. At the beginning of a project, the Tetra Tech Project Manager will consult with the Tetra Tech QA Officer to determine applicable documentation requirements. Automate and document each aspect of data acquisition. Avoid manual transcription (non-automated data processing) because of the potential to introduce error into the data set. However, automated processes must be properly checked and verified to ensure error-free results.

The important aspects of data documentation include keeping records of the data source (e.g., URL, agency providing the data, version), the access date, and the access procedure. At the beginning of the project, the Tetra Tech Project Manager will consult with the Tetra Tech QA Officer to determine applicable documentation requirements. Screen captures of search results (refer to Figure 1) can be a quick and effective way to document aspects of the download procedure. Figure 1 is an example of a screen capture of selection criteria entered in the Water

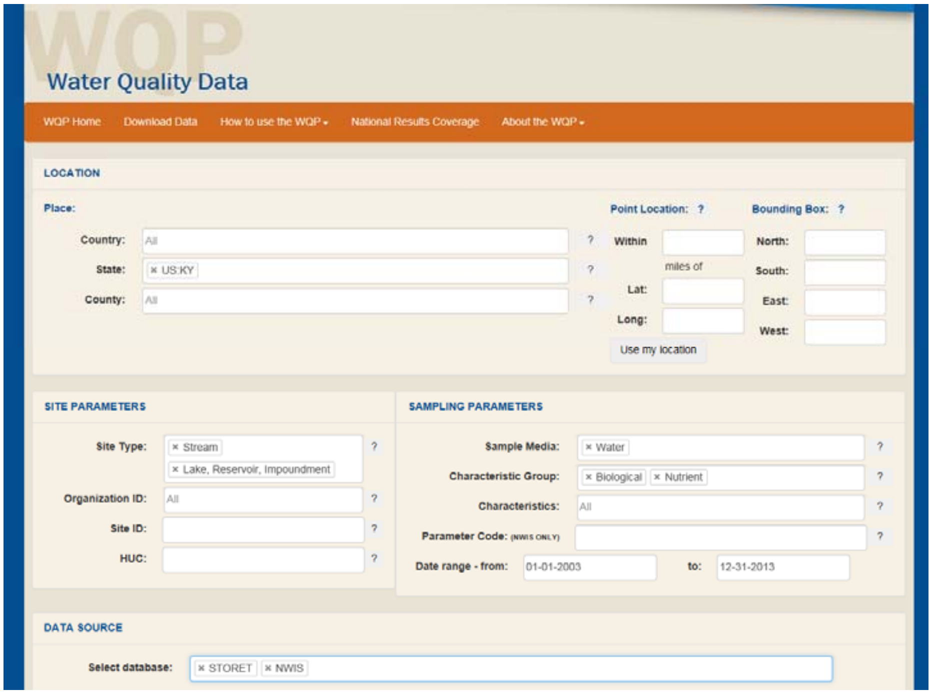

Figure 1. Example screen capture of search criteria selections in the Water Quality Portal[1]

Quality Portal: State = “Kentucky”; Site Type = ”Stream” and “Lake, Reservoir, Impoundment”; Sample Media = “Water”; Characteristic Group = “Biological” and “Nutrient”; Date Range = “01-01-2003” to “12-31-2013”; and Database = “STORET” and “NWIS”. Alternatively a README text file or word document can be saved with the original data to document this information. If data are acquired via e-mail or file transfer protocol (ftp), save a copy of the original e-mail or ftp access instructions.

For replicability and QA, maintain a copy of the raw, unaltered downloaded data and related metadata, including variable names/definitions. These raw data can also be important in troubleshooting processing errors introduced during the analysis and in maintaining version control. Data are increasingly dynamic with real-time data uploads and can be updated by data owners at any time. Also, maintain the ‘ready-to-analyze’ data sets. A ‘ready-to-analyze’ dataset refers to the dataset after all processing and transformations have been completed, prior to analysis. At a minimum, the original data, the ‘ready-to-analyze’ data, and all project deliverables should be electronically stored where automated backups are made on at least a daily basis for the purposes of catastrophic recovery. This can include office servers or cloud-based solutions. Test analyses and temporary files do not require this type of storage or backup.

[1] http://www.waterqualitydata.us/portal.jsp

Data Quality Considerations

At the beginning of a project, the Tetra Tech Data Manager will consult with the Tetra Tech Project Manager and QA Officer for applicable data quality considerations. The advantages of using secondary data include cost and time savings, more extensive data availability, and the potential for analysis by experts not available at smaller scales. However, secondary data have inherent disadvantages because the data were not collected by those conducting the analysis and were often not collected to answer the specific question(s) of the current analysis. For example, data might have been collected for different variables, geographic regions, or sampling frequencies. In addition, because the analyst did not participate in the sampling design or sampling process, the methods and quality of analysis might be unknown. Data might have been collected using different sampling techniques (grab sampling versus composite sampling or random sampling versus targeted sampling). The laboratory or sampling processing methods might also have differed. Differences in technique or documentation can contribute to variability in the data set when multiple secondary data sources are combined for an analysis. Errors in spatial position and taxonomic identification are particularly common in environmental data (Chapman 2005).

The amount of documentation associated with a particular source often varies widely. Documentation of the source, including metadata documented in project reports, validation reports, and any database information, should be maintained along with the data. Research into the origin and documentation of a data source might be necessary to properly evaluate the data source. Potential sources for this documentation might include the website for the agency or group that collected the data, published reports, research articles, and personal communication with the original researcher or monitoring group staff.

Consider this series of general questions when evaluating the quality of any secondary data source and the applicability of the data to the current project (Boslaugh 2007):

- What was the original purpose for which the data were collected?
- What kind of data are they, and when and how were the data collected?
- What data processing and/or recording procedures have been applied to the data?

Also consider the following questions, which are more specific to water quality data, when evaluating a water quality data source (USEPA 2009):

- Were the data generated under an approved QAPP or other documented sampling procedure?
- If multiple data sets are being combined, were the data sets generated using comparable sampling and analytical methods?
- Were the analytical methods sensitive enough (detection limits) to meet project needs?
- Is the sampling method indicated (e.g., grab, composite, calculated)?

- Was the sampling effort representative of the waterbodies of interest in a random way, or could bias have been introduced by targeted sampling?
- Are the data qualified? Are sampling and laboratory qualification codes or comments included? Are the qualification codes defined?
- Is sufficient metadata available about variables like sampling station location, date, time, depth, rainfall, or other confounding variables?

Specific evaluation criteria for each parameter being considered should also be applied across all sources. Although many water quality data sets include QC samples labeled as duplicate, split, spiked, blank, and so forth; re-checking QC samples is beyond normal practices for secondary data analyses. Rather, it is expected that project-specific QAPPs or similar documentation describing the performance criteria evaluated and met are available for data obtained from peer reviewed sources or from federal, state, or local government reports or data compilations. If this documentation is not readily available, Tetra Tech will consult with the client to determine how much effort should be expended to find reports or metadata that might contain that information. Nevertheless, establishing minimum data requirements for secondary data analyses is often valuable. For example, water chemistry data might require locational information, date, time (optional), depth (optional), chemical name, units, numerical result, and data qualifiers. Specific requirements would depend on project specific needs. For example, it might be necessary to identify outliers or changes in analytical methods. In those cases where requested by the client, QC samples can be used to double-check sample accuracy (e.g., whether duplicate samples are within 15 percent of the corresponding sample).

The National Environmental Methods Index (NEMI)[2] provides a searchable compendium of environmental methods. Different scientific methods can be compared using the method summaries, which also include literature citations. Generally, parameters monitored using different methods should not be combined unless the techniques are documented to be scientifically comparable. EPA also has compiled training materials to detect improper laboratory practices when working with monitoring data.[3]

Other QC checks could leverage the spatial aspect of the data. Stations should be mapped to verify that the data fall in the correct political boundary, ecoregion, waterbody type, or other descriptive spatial factor. Data that reportedly reflect sampling of a lake in Kansas but have coordinates in the Pacific Ocean should call the accuracy of the data and/or the coordinates into question, as should the occurrence of a fish species in a lake in Kansas, not found in inland lakes. Continuous data have a different set of quality concerns such as time stamps, drift in measurements over time, and trimming of the period of record to eliminate records that are out of water, choked in sediment, or exhibiting drift. These concerns are not addressed in this Tetra

[2] www.nemi.gov

[3] http://www.epa.gov/quality/trcourse.html#monitoring

Tech QA/QC document. For further information on continuous data quality concerns, refer to draft guidance from USEPA and USGS on this topic (USEPA 2014; Wagner et al. 2006).

Data Organization

After acquisition, data should be organized and stored. The original unaltered data and “as analyzed” data files should be archived to ensure replicability of the work. Data sets are constantly being updated, so without the original data, replicating an analysis is often impossible. If you are combining data from multiple sources, include information documenting the source of the data in spreadsheets or databases. For water quality data, generally seek to organize data into one of the following hierarchical structures: *(1) source → station → sample* or *(2) source → station → sample → result* so the data are ready for a variety of analyses.

A relational database, such as Microsoft Access or an Oracle-based system, is an efficient method used to organize multiple related tables. For water quality data, these tables can include station-level tables, sample-level tables, and lookup tables. A primary key or unique identifier, such as a numerical field or a composite primary key made up of multiple fields (e.g., station-sample-date-time-depth), should be assigned to each record. Each table should have a primary key. Foreign keys are fields in one table that uniquely identify a row in another table, often called a lookup table. Figure 2 provides an example of sample-level and lookup relational tables with the primary keys and foreign keys identified. Referential integrity should be maintained such that each foreign key corresponds to the value of a primary key or a null value in a lookup table.

Table 1 (Sample-level table)

Sample ID	Parameter	Result	Remark
1	Total Nitrogen	1.4	DQ
2	Total Nitrogen	1.4	
3	Nitrate	0.5	
4	Total Nitrogen	2.4	DQ
5	Total Nitrogen	2.4	
6	Total Phosphorus	0.08	T

Primary Key　　Foreign Key

Table 2 (Lookup table)

Remark	Description	Action
W	cloudy	NA
DQ	Duplicate Quality Assurance Sample	REMOVE
T	Sample exceeding holding time	REMOVE
?	Data should be rejected	REMOVE
*	Exceeded MDL	Flag

Primary Key

Figure 2. Example of relational tables with primary and foreign keys

A disciplined file structure and file naming convention can improve version control management. Label files with unique identifiers such as dates or other indicators of version control. Include a documentation table which identifies the database objects (tables, queries, reports, etc.). Maintain the original data in a read-only database and ‘link in’ to the analysis database to prevent accidental changes to original data.

Large (e.g., multiple-gigabyte) files sizes are increasingly common, especially with remote sensing imagery, spatial data, or large databases. Consider the storage and backup requirements of these large files. For example, you might need a separate server to accommodate the data needs for a project. If you are working with multiple people, consider the implications of file storage choices for file transfer. Spatial data management has some unique considerations discussed in a separate Geospatial and Data Management QA/QC Procedures document.

Sample-level water quality data are often stored in a vertical format with a column for parameter or characteristic name and a column for result value, as shown in Figure 3. After data transformations, but before statistical analyses, it is often more convenient and space-efficient to convert the data to a horizontal format, in which each parameter of interest has its own column and results for that parameter are reported in the parameter column. This approach allows for simpler identification of paired sampling data (samples taken from the same station-date-time) for multiple parameters, which in turn makes identifying relationships among parameters possible.

Vertical Format

Date/Time	Characteristic	Result
1/1/2000 15:00	Nitrogen, Total	1.5
1/1/2000 15:00	Nitrate	0.8
1/1/2000 15:00	Phosphorus, Total	0.5
1/1/2000 15:00	SRP	0.1
1/1/2000 15:00	Orthophosphate	0.2
6/1/2000 9:00	Nitrogen, Total	2.2
6/1/2000 9:00	Nitrate	1.3
6/1/2000 9:00	Phosphorus, Total	0.2
6/1/2000 9:00	SRP	0.1
6/1/2000 9:00	Orthophosphate	0.1

Horizontal Format

Date/Time	Nitrogen, Total	Nitrate	Phosphorus, Total	SRP	Ortho-phosphate
1/1/2000 15:00	1.5	0.8	0.5	0.1	0.2
6/1/2000 9:00	2.2	1.3	0.2	0.1	0.1

Figure 3. Example of water quality data in vertical (left) and horizontal (right) format

Effective data organization can improve the efficiency with which data can be checked for errors, processed, transformed, and documented. Sorting by location, source, parameter, or other column allows error-checking and transformation to be automated, which improves not only efficiency but also QA.

Aligning matching records can be arduous if not already performed. For example, StationID might differ among sampling visits and would need to be checked using latitude/longitude information (which should be associated with each station). Also to be matched when combining data sets or if the original data set is lacking could include chemical species, taxonomic names, and dates for nearly concurrent sample types. For example, if habitat was collected on one day and fish were collected 2 days later, there should be an indicator that those are (or are not) comparable for analysis.

Data Transformation

After acquiring the data, archiving the original unaltered data, performing QC checks, and organizing the data, the data often need to be transformed or processed to put them in a comparable format. Data transformation should be organized, systematic, repeatable, and automated as much as possible to reduce the chance of error and minimize the level of effort common to manual transformation.

This task often involves manipulating the data from the original data source to a 'ready-to-analyze' data set. The original data source can be one to multiple files with the same or different data structure. Additionally, different data sets can have different naming conventions, units, etc., that need to be unified. The principal QC questions include the following:

- Was the process documented?
- Were all data files processed?
- Were all data records processed (e.g., no dropped records)? If not, were excluded data justified?
- Were transformation and reshaping steps implemented correctly?

As described earlier in this document, cursory level compilations of data that are used to inform whether more robust data compilations can be prepared, can be developed with a cursory-level QC review so long as work products are identified as such in their transmittal to the client and in progress reports. With the exception of these cursory-level data compilations, independent checks of data compilations should be performed to ensure we meet EPA's QA requirements. Applicable QC checks for data reshaping and transformations tasks are summarized in Table 3.

Table 3. Applicable QC checks for data reshaping and transformations.

QC #	Description	Cursory Level Review	Standard Level Review
1.1	Confirm that the reshaping and transformation steps are documented with the data.	X	X
1.2	Confirm that the files processed and record counts of the end product meet expectations.	X	X
1.3	Review meta information prepared by the original analyst that documents transformations and reshaping.	X	X
1.4	For each different file type (i.e., a file with different structure or legacy), confirm the first, last, and a selected middle portion of the data were transformed and reshaped correctly.*		Up to 10% of processed data files, but no less than 2 data files of each type**
*Evaluating the first and last portions of data helps confirm that no records were accidentally dropped during processing. Selection of the middle portion of the data to check should be done by targeting unique and unusual record types that might stress the transformation and reshaping processing. ** More files should be reviewed (up to 10%) if files are processed individually while fewer files are appropriate for automated to semi-automated procedures.			

This section describes cursory- and standard-level QC checks that should be performed. Some projects might specify complete independent checking of an entire data compilation. This specification, or even standard-level QC, could cause a significant and, perhaps unnecessary, resource burden in projects that involve multiple iterations and modifications; thus the Data Manager should confer with the Project Manager to confirm the most cost-effective and efficient process for QC checks.

For data compilations that will be used to directly inform decisions/actions (e.g., remediation, compliance decisions, regulatory action, source control, capital investment), processed data will be independently checked using a standard level review. This standard review consists of independently checking each different file type (i.e., a file with different structure or legacy), confirming the first, last, and a selected middle portion of the data were transformed and reshaped correctly. Evaluating the first and last portions of data helps confirm that no records were accidentally dropped during processing. Selection of the middle portion of the data to check should be done by targeting unique and unusual record types that might stress the transformation and reshaping processing. More files (up to 10 percent) should be reviewed if files are processed individually while fewer checks (no less than 2 data files of each type) are appropriate for automated to semi-automated procedures.

All identified data processing errors will be corrected and the Tetra Tech QC Officer will perform a follow-up review of the corrected components to ensure that the errors have been corrected. Where changes are made to previously checked compilation or changes are made to address the results of QC checks, it is normally expected that only the changed/corrected components of the compilation and the dependent, follow-on components would be subject to checking/re-checking. For example, if a change or correction is made to an analysis (e.g., substituting a maximum likelihood technique for a least squares estimation method) then it would not be normally expected that data transformation steps that led to creating the ‘ready-to-analyze’ data set would need to be re-checked.

Frequently, data column names as well as values (e.g., parameter names, comment fields, and result values) are not consistent between different data sources or even within a single source. A more detailed description of data source fields common to water quality data is provided in Attachment 2. To combine data while maintaining the original data, it is good practice to create additional user-specified fields to represent common parameters, standardized comments, and comparable values. Creating user-specified fields allows for correcting errors and performing transformation while retaining the original data in separate fields. Thus, the opportunity to go back to the original data is maintained. Maintaining documentation of data transformation and error correction is especially important when the processes are being performed by people other than the primary data collector.

Creating user-specified fields provides an opportunity to convert units to like units, standardize parameter names, interpret comment fields, convert non-detect values, or institute other data transformations. For instance, a user-specified data qualifier field might be used to flag or exclude blank samples or samples with non-numeric characters in the value field. Figure 4 provides an example of how user-specified fields might be used to convert field names and units and interpret comment fields. Another important use for user-specified fields is creating a column that documents the original source and the row ID of the original source when merging data, so that if systemic issues are found in a source, they can be resolved and processed more effectively. A quick reference guide of procedures to process water quality data is provided in Attachment 1.

Original Data					User-specified fields		
Sample ID	Parameter	Result	Units	Comment	PARM	RESULTVALUE (mg/L)	REMARK
1	Total Nitrogen	1.4	mg/L		TN	1.4	KEEP
2	Nitrogen	19000	µg/L	Sampler Error	TN	19	REMOVE
3	Nitrate	0.8	mg/L	Estimated	NO3	0.8	REMOVE
4	Nitrogen as N	2.4	mg/L		TN	2.4	KEEP
5	Total Nitrogen	2400	µg/L		TN	2.4	KEEP
6	Nitrate as N	500	µg/L		NO3	0.5	KEEP

Figure 4. Example user-specified fields

Unintended data duplication is frequently present in water quality data sets. It might be the result of obtaining the same data from different sources, or simply data entry error. This phenomena, should not be confused with field or laboratory duplicate samples which are commonly performed for QA/QC purposes, including evaluating data precision. Unintended duplication can be present within a single data source or among different data sources. Merging two data sets sometimes creates new inconsistencies and duplication. Unintended duplication can skew and bias data. It should be flagged and screened from the analysis as much as possible.

Some samples might resemble duplicate entries but actually have different depths, times, or other distinguishing features. If the only fields that are different are descriptive fields, such as comment fields, that might be an indicator of duplication. The organization ID and sampling name can be good indicators that duplication is present, but also look for duplicate values in the data over the same time frame. For example, several identical numerical values on the same day might indicate duplicate data. Sorting the data chronologically and looking for duplicate sample results is one way to begin to identify duplication. Excel has features to identify and highlight duplicate values in a field; when the data are sorted chronologically, Excel can identify potential duplicates. Duplicate records should be flagged using a user-specified field but generally not deleted. Simply deleting unintended duplicate data (i.e., not field or laboratory duplicate samples), rather than flagging and excluding the data, creates a potential for error and data loss that is difficult to identify.

References

Boslaugh, S. 2007. *Secondary Data Sources for Public Health: A Practical Guide*. Cambridge University Press.

Chapman, A.D. 2005. *Principles of Data Quality*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=f2cb9b7aff6fce23e2e4d8774bb0ad.

Edwards, P.J. 1986. *Conversion Factors and Constants Used in Forestry, with Emphasis on Water and Soil Resources*. U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. NE-GTR-113. November 1986. 12 p. http://www.treesearch.fs.fed.us/pubs/4159.

USEPA. 2009. *EPA New England Quality Assurance Project Plan Guidance for Environmental Projects Using Only Existing (Secondary) Data.* EPA-820-S-10-001. U.S. Environmental Protection Agency New England, Quality Assurance Unit, Office of Environmental Measurement and Evaluation. http://www.epa.gov/region1/lab/qa/pdfs/EPANESecondaryDataGuidance.pdf.

USEPA. 2012. *Guidance for Evaluating and Documenting the Quality of Existing Scientific and Technical Information Addendum to: A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*. Prepared for the U.S. Environmental Protection Agency, Science and Technology Policy Council, Washington, DC. http://www2.epa.gov/sites/production/files/2015-05/documents/assess3.pdf

USEPA. 2014. Best Practices for Continuous Monitoring of Temperature and Flow in Wadeable Streams. Global Change Research Program, National Center for Environmental Assessment, Washington, DC; EPA/600/R-13/170F. Available from the National Technical Information Service, Springfield, VA, and online at http://www.epa.gov/ncea.

Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A. 2006. Guidelines and standard procedures for continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1–D3, 51 p. + 8 attachments. http://pubs.water.usgs.gov/tm1d3.

(This page is intentionally blank.)

Attachment 1. Quick Reference Guide of Procedures to Process Water Quality Data

These procedures include examples of the types of checks that are performed—not every check to possibly perform. Site-specific steps will apply to many datasets. These steps do not necessarily need to be performed in sequential order and may be iterative.

✓	**Data Acquisition/Organization**
	Acquire data and companion metadata. Maintain a copy of all original files. Document the source data source, access date, and the download procedure.
	Start a recording sheet to record decisions, selections to review for quality control and data archive.
	Organize data in a spreadsheet or relational database. Organize data using a hierarchical structure (e.g., source→station→sample or source→station→sample→result).
	Data formatting • Convert "as text" values to numbers. Check for non-numeric characters in numeric fields. • Label all blank cells as blanks to avoid conversion to zero, remove all inappropriate zeros (e.g., chemistry methods rarely measure a true 0, if they have an MDL)
	Review data dictionaries and field names before combining data from multiple sources into a spreadsheet or database format—do not assume that field names are equivalent.
	Utilize exploratory data analysis techniques such as summary statistics or graphical techniques.
	Data Processing
	Generally – do not delete data. Add a screening column to track decision-making and remove records. Maintain removed records in separate file with justification.
	Compare the geographic/temporal scope of the data to the project objectives—it might not be necessary to process all data from a given data set. Map stations in a GIS to further refine and select data based on analysis selection criteria. Conduct quality assurance checks based on spatial location.
	Check for unintended duplicate entries (i.e., not field or laboratory duplicate samples). Identify and screen those samples that are duplicates. Check for samples or results that do not have stations.
	Interpret data qualifiers and comments (e.g., spikes, blanks, duplicates, holding time, errors.) Screen samples based on an interpretation of the codes.
	Check each field for inconsistencies. Screen undesired components. Examples include: • Coordinates – Are lat/long coordinates in comparable form? Negative values? • Date/Time – standard format should be used (XX/XX/XXXX). All in same time zone. • Depth – filled out? in like units? • Sample Media/Type – water, groundwater, air, effluent, stormwater, process water
	Add user-specified fields to interpret, standardize, and clean up existing fields: • Waterbody types – interpret and simplify • Analytes/taxonomy – consistent use of analyte and taxa names • Analytical method/Sample Fraction – consider accuracy and comparability of methods • Units – standardize units and convert values as appropriate.
	Censored Data – Data that are reported as not detected or below detection limit should be utilized but accounted for statistically. Several methods are available to interpret censored data depending on the analysis. At this stage, maintaining MDLs and PQLs is likely appropriate to provide later analysis flexibility.
	Data Transformation
	Calculate metrics or new parameters based on the data available • Calculate parameter sums or products (e.g., TN as sum of nitrate+nitrite and TKN). • Calculate TSI, M-IBI, F-IBI, other biological indices.

	Outliers – Analyze the data for potential outliers and consider screening those data which are clearly outliers and may introduce bias or error into the dataset.
	Document the process to ensure quality assurance and reproducibility.

Attachment 2. Data Source Field-specific Water Quality Data Tips

Several fields that provide more information about the sampling process or sampling location are often included with water quality data. These fields might include sample media, sample type, sampling type or location, and waterbody type. These fields might need to be interpreted or transformed to select the data that are of interest to the analysis. Descriptions of common fields and transformations that should be considered follow:

- **Sample media:** A field or two for sample media (e.g., water, soil, groundwater) are sometimes included. They can be used to verify that the correct query selections were made for the sample media of interest. Sometimes sample subdivisions identify distinctions that should not be included in an ambient analysis, e.g., effluent, process water.
- **Sample type:** A field that identifies routine samples versus duplicate or quality control samples such as spike samples, field replicates, laboratory replicates, or other duplicates is sometimes included. Checking routine values against duplicate values can be a valuable quality control check, but also ensure that duplicate values are not included in the data set used for analysis.
- **Sampling type or location:** Fields indicating the type of sampling, such as effluent, ambient, stormwater, baseflow, pipes, finished water, or process water, are sometimes available. Consider the location of the sampling effort. Sampling focused on effluent outfalls or on pristine waters could introduce bias into an analysis, depending on what the purpose of the project is. Sampling type or location can be an important indicator of sampling bias or spatial bias inherent in the data set resulting from opportunistic sampling rather than random sampling.
- **Waterbody type:** An indication of the type of waterbody where the sampling occurred might be included (e.g., stream/river, lake/reservoir, estuary, ocean, wetland, canal, stormwater). This field can be used to further subset sampling data to the data of interest.

Descriptive fields such as temporal indicators (e.g., date, year, time), sample depth, latitude/longitude, or units are often included in varying formats. A description of common fields and transformations that should be considered is provided below:

- **Temporal:** Ensure all date and time fields are in the same format (e.g., DD-MM-YYYY, YYYY-MM-DD). It is recommended that you use military time and account for time zones. It might be helpful to have one field with "Date" and separate fields for "Year," "Month," "Day," and "Time." If a measurement of diurnal fluctuations is not needed in a parameter, averaging data by day might remove some inconsistencies resulting from data without time information or with slightly different times due to different processing labs or data entry error. Searching for dates outside the range of interest or outside reasonable date or time values (e.g., month <1 or >12, day <1 or >31, year <1900, time <0 or >24) can be a helpful screening tool. Having a sampling date is a reasonable minimum requirement for data.

- **Depth:** Depth should generally be a numeric field. Sometimes a surface or bottom indicator is included as well as a numeric depth field (e.g., S, B). It can be helpful, especially in lakes and estuaries, to add a separate text depth column for profile data that indicate surface, depth, or bottom measurements for some parameters (e.g., dissolved oxygen). Depth units should be standardized to a consistent format (feet or meters).
- **Latitude/Longitude:** Ensure that latitude and longitude are reported in a consistent format. Latitude and longitude units are most often reported in degrees, minutes, and seconds (DMS) (e.g., 39°59’56.055”N, 102°3’5.452”W), decimal degrees (DD) (e.g., 39.999012, -102.052062), or sometimes UTM coordinates (e.g., 17T 630084 4833438). The examples provided are all roughly from the same point on the border of Kansas, Nebraska, and Colorado. To convert from DMS to DD, use the formula: (degrees) + (minutes/60) + (seconds/3600) = decimal degrees. If values are missing, consider digitizing from GIS or geocoding from an address if provided. One of the most frequent errors is omitting the negative sign (-) in decimal degree coordinates from the southern or eastern hemispheres. If all the records are from North America, all the longitude values should include a negative sign. Consider spatial accuracy. With today's standards, be wary of decimal degree data with less than six digits of precision accuracy or seconds reported with less than two digits of precision (although for larger waterbodies less precision might be acceptable). A typical minimum data requirement for station-level data is that the station must have a latitude and longitude measurement as well as the reported datum. Look for extreme values: Latitude should never be outside the range of 90 to -90 degrees; longtiude, 180 to -180.
- **Units:** Units should be standardized by parameter and among parameters. Check for systematic incorrect reporting of units when converting all values for a parameter to one unit of measurement. Note that laboratories often report results on a weight-per-weight basis, such as parts per million (ppm) or part per billion (ppb). In water samples, 1 ppm is essentially equivalent to 1 mg/L and 1 ppb is equivalent to 1 µg/L unless concentrations are very high (>7,000 mg/L) (Edwards 1986). In addition, µg/L and mg/m^3 can be considered identical in most cases in water samples. Outliers for a parameter might be an indication that data are reported in varying units.

Appendix B.

Quality Assurance and Quality Control Considerations for Statistical Analyses

(This page is intentionally blank.)

Tetra Tech's Quality Assurance and Quality Control Considerations for Statistical Analyses

April 7, 2016

Prepared by

Tetra Tech, Inc.

10306 Eaton Place, Suite 340

Fairfax VA 22030

Contents

Introduction

This guidance provides an overview of quality assurance/quality control (QA/QC) practices for statistical analyses. The Tetra Tech Project Manager, Statistical Analyst, QA Officer, and QC Officer should refer to this guidance to ensure that the QA/QC requirements set by our EPA clients are met. The Tetra Tech Project Manager supervises the overall project and is responsible for coordinating project assignments, establishing priorities and schedules; ensuring completion of high-quality projects within established budgets and schedules; providing guidance, technical advice, and evaluating the performance of those assigned to the project; implementing corrective actions; preparing or overseeing preparation and review of project deliverables; and providing support to EPA in interacting with the project team, technical reviewers, and others to ensure that technical quality requirements are met in accordance with EPA's objectives. The Tetra Tech Statistical Analyst is responsible for performing the statistical calculations and analyses and the Tetra Tech QC Officer is responsible for checking those activities. The Tetra Tech QA Officer, with the assistance of the assigned QC Officer, will monitor QC activities to determine conformance with project QA/QC requirements.

Statistical analysis of data covers a wide range of calculations and graphical visualization techniques. Relevant quality assurance/quality control (QA/QC) practices for statistical analyses include ensuring that the analyses are correct, reproducible, and transparent. To ensure that we meet the QA/QC requirements set by our EPA clients, statistical calculations must undergo QC reviews and those reviews must be documented in the project files. The Tetra Tech Project Manager and QA Officer will communicate to the Statistical Analyst whether specific documentation of QC reviews is required for a particular task.

The appropriate level of statistical analysis and corresponding level of QC review will vary with project goals, available data, resources, and the decisions to be made. At the beginning of a particular project, the Statistical Analyst will provide recommendations to the Tetra Tech Project Manager regarding statistical methods to be used for analyzing the data. At this time, the Statistical Analyst and Project Manager should discuss the approximate level of effort needed for the various analyses and corresponding level of QC review required. Follow-up discussions should be held throughout the duration of the project, as needed, to clarify the analyses to be performed, level of QC review needed, and level-of-effort required.

- It should be noted that analyses that are expected to be used to inform future, more detailed analyses can be performed with a cursory-level QC review so long as work products are identified as such in their transmittal to the client and in progress reports.
- Analyses that directly inform decisions/actions (e.g., remediation, compliance decisions, regulatory action, source control, capital investment) require a higher, standard-level QC review.

This document describes the following topics as related to ensuring the quality of Tetra Tech's statistical analyses: method selection, best practices, and quality control.

Method Selection

Based on the characteristics of available data and the project's needs, the Tetra Tech Project Manager, in consultation with the client and the statistical analyst, will determine whether common exploratory summary statistics and/or standard graphical presentations will be needed for a particular project, or whether more advanced predictive procedures (e.g., applying a range of hypothesis tests, applying multivariate tools, developing empirical models) will be required.

- Common summary statistics include counts of observations and distribution characteristics (e.g., mean, standard deviation, coefficient of variation, variance, median, percentiles).
- Standard graphical presentations (e.g., distribution plots, scatter plots, boxplots, time series).
- Parametric and non-parametric hypothesis tests (e.g., t-test, analysis of variances (ANOVA, Kruskall-Wallis).
- Multivariate tools (e.g., principal components analysis, clustering analysis, canonical correspondence analysis, discriminant analysis, non-metric mulitdimensional scaling).
- Models (e.g., linear and non-linear regression, general additive models, general linear models, Bayesian hierarchical models).

When deciding which statistical procedure to apply to any data set, it is essential to consider the characteristics of the data, which will help determine the appropriate statistical analysis. Some common characteristics of data include one or more of the following:

- Presence of outliers, extreme low or high values that occur infrequently, but usually somewhere in the data set (outliers on the high side are common) resulting in skewed distributions.
- Variance heterogeneity.
- Non-normal distribution.
- Small sample size.
- Censored data – concentration data reported above or below one or multiple detection limits or reporting values.
- A lower bound of zero (e.g., no negative concentrations are possible).
- Missing values.
- Irregular sampling.
- Strong seasonal patterns.
- Autocorrelation – consecutive observations strongly correlated with each other.
- Dependence on other uncontrolled or unmeasured variables – values strongly co-vary with such variables as streamflow, precipitation, or sediment grain size.
- Measurement uncertainty.

Common Tools/Software

There are a wide variety of computer tools/software available to support statistical analyses including spreadsheets (e.g., Excel), databases (e.g., Access, SQL), commercial statistical packages (e.g., SAS, Minitab, Systat), customized software (software created by a state/federal agency or 3rd party vendor designed for a particular analysis, e.g., ProUCL, EPIWEB), and programming code (e.g., FORTRAN, C++, Python, R). Hand calculations can also be used.

The functionality of these tools overlaps, yet different numerical results are sometimes computed when using different tools. For example, a key part in estimating percentiles is to assign ranks to the observed data. Some spreadsheet software programs assign the minimum rank to tied values rather than assigning a rank that is equal to the median of the ranks if the observations had not been tied. Other commercial software may include multiple formulas for computing percentiles, which the user can select. The outcome is that different percentiles might be computed among different software packages. Similarly, different analysts can compute different numerical results when applying similar steps, but simply in a different order (e.g., the logarithm of the average is not equal to the average of the logarithms). It is important that the original analyst and person performing QC checks be aware of these potential differences and their impact on the analyses and independent checking of results.

Overall Justification and Documentation of Methods Used

Common summary statistics and standard graphical presentations that follow normal practices for the type of data being evaluated require little or no justification for their usage. Method selection for hypothesis testing, multivariate procedures, model development, or more advanced procedures should be made by an experienced analyst with justification included in the corresponding report. Citing similar analyses available from applicable guidance/methods documents or refereed literature is sufficient. Methods selected from the internet, gray literature, software literature, or presentations require additional narrative to document why a particular method is, or might be expected to be, appropriate.

Best Practices

This section provides a list of best practices that can be implemented to reduce errors in statistical analyses and improve the overall work product. It is the responsibility of the Tetra Tech Project Manager and delegated statistical analyst to identify which practices are appropriate for a particular task.

Overall

- Maintain original copies of source data, related metadata, and the 'ready-to-analyze' data sets. See the Secondary Data Management SOP for more information on data organization and management. Use a naming convention for files that is understandable to you and others, and is designed in way that helps ensure that version control is maintained throughout the project (e.g., use of dates, version numbers, draft, final).
- Develop a written technical description of the analysis. This description can be written before beginning analyses and/or developed as a living document through the course of the project.
- Identify analysis milestones where data should be exported/saved to improve transparency and reproducibility as well as for QC analyses and record keeping.
- Perform statistical analyses in a similar fashion throughout the project. Document deviations in the technical description of the analyses.
- Document the name, version, and, where applicable, the source code of the software used to perform analyses. This is applicable for commercial and open source software.
- Give titles to objects in the spreadsheet, database, or software that lend an understanding to the purpose of the object. For example, a database query entitled 'selectData_v02' might be a useful object title for the 2nd version of a query that selects data from a primary source table.

Hand Calculations

- Hand calculations should be legible and document their purpose.
- Scan hand calculations so they can be maintained as electronic documents with other documentation.

Spreadsheets

- Include a documentation tab which includes information about the spreadsheet as a whole and a description of the other tabs.
- Organize tabs from left to right in the same order as the analysis steps.
- Organize calculations within a tab from left to right and/or top to bottom.
- Make judicious use of named cells and relative/absolute cell addresses to allow maximum use of ‘fill-down’ and ‘fill-right’ options.
- Limit cell and font styles for highlighting information that could be derived from examining the data. For example, it is an acceptable practice to set a cell color to “yellow” to help visualize all p-values less than 0.05. It is not a typically accepted practice to highlight statistically significant regression slopes but not show/include the actual p-values.

Commercial statistical packages

- Document the name and versions of the software used.
- Document the steps and settings used to implement calculations that are menu/interactively implemented.
- Develop macros to implement repeated tasks.

Customized software

- Document the name and versions of the software used.
- Document the steps and settings used to implement calculations that are menu/interactively implemented. (Note that it is a common practice for software packages to be developed by a third party on behalf of a state or federal agency to perform a very specific set of analyses that are not directly available in commercial software. While these software packages may be well tested for the primary work flow, they may not be as well tested or error proof, if used in a non-conventional manner. Therefore it important that the analyst have an understanding of the basic work flow of the software package and document its usage.)

Programming code (e.g., FORTRAN, C++, Python, R)

- Maintain all source, and if applicable compiled, code used to perform all analyses for documentation and future use. This allows for transparency and repeatability of the analysis.
- Where practicable, repeat the analyses with a separate tool to verify the results or code and/or independently unit test the source code.

Quality Control

The appropriate level of QC will vary with project goals, available data, resources, technical approach, and the decisions to be made. The principal QC questions include the following:

- Was an appropriate method chosen and applied?
- Were the statistics computed and graphics created correctly?
- Were the statistics and graphics representative of the data?
- Were method assumptions met?
- Were the results presented correctly?

Selection of a particular method depends on the data and the analysis objectives. Calculating summary statistics and developing basic graphics can normally be performed by any basic environmental consultant/staff member. Exceptions might include calculations with censored data or other non-standard data. Advanced statistical calculations and related output (tabular, graphic, etc.), including, but not limited to, hypothesis testing, multivariate tools, empirical models, and statistical simulations will generally benefit from oversight by an experienced analyst. However, it should be noted that multiple methods might be applicable for a given project and set of data (see *Overall Justification and Documentation of Methods Used*).

As described in the introductory section of this document, analyses that are expected to be used to inform future, more detailed analyses can be performed with a cursory-level QC review so long as work products are identified as such in their transmittal to the client and in progress reports. While a cursory-level QC review could include some independent checking of calculations, a cursory-level review may also be limited to reviewing selected sections of a technical report that focus on the data summary, technical approach, and results sections.

For statistical calculations performed using analysis software for which the results will be used to directly inform decisions/actions (e.g., remediation, compliance decisions, regulatory action, source control, capital investment), calculations will be independently checked using a standard-level review. As used here, independent calculations can refer to a different analyst performing the same analysis; or may refer to the same analyst performing the same analysis using a different software tool. Some projects might require complete independent checking of all calculations. This requirement, or even standard-level QC, could cause a significant resource burden in projects that involve multiple iterations and modifications; thus the Statistical Analysts should confer with the Project Manager to confirm the best timing for QC checks to best use the available budget.

With today's computer technologies, it is more appropriate in some instances to perform targeted checking rather than rely on a fixed "10 percent of all calculations" rule when performing independent calculations. A standard-level review consists of up to 10 percent independent recalculations of computations and graphs, but no less than 2 examples of each computed statistic and 2 examples of each graphic type. More calculations (up to 10 percent) should be reviewed if data sets or points are processed individually while fewer checks (no less than 2 examples of each computed statistic and 2 examples of each graphic type) are appropriate for automated to semi-automated procedures. Selection of which statistics and graphs to check should include targeting unique and unusual record types that might stress the calculation and graphing process. All identified calculation errors will be corrected and the Tetra Tech QC Officer will perform a follow-up review of the corrected components to ensure that the errors have been corrected. Where changes are made to previously checked analyses or changes are made to address the results of QC checks, it is normally expected that only the changed/corrected components of the analysis and the dependent, follow-on components would be subject to checking/re-checking. For example, if a change or correction is made to an analysis (e.g., substituting a maximum likelihood technique for a least squares estimation method) then it would not be normally expected to re-check data transformation steps the led to creating the 'ready-to-analyze' data set. In cases where codes are developed to perform statistical calculations, codes and changes to codes should be checked and tested for reproducibility by a qualified QC Officer, and if possible, run on independent software.

In the majority of instances, statistical calculations will be performed using analysis software. In (relatively uncommon) circumstances where statistical calculations are primarily performed by-hand, a Tetra Tech QC Officer will independently recalculate 10 percent of these calculations to ensure they were performed correctly. If more than 1 percent of the data calculations are incorrect, the Tetra Tech QC Officer will check independently check the remaining calculations to ensure they are correct. All identified errors will be corrected.

Appendix C.

Tetra Tech's Geospatial and Data Management QA/QC Procedures

(This page is intentionally blank.)

Tetra Tech's Geospatial and Data Management QA/QC Procedures

April 6, 2016

Prepared by
Tetra Tech, Inc.
10306 Eaton Place, Suite 340
Fairfax VA 22030

Contents

1 Introduction

The purpose of creating this document is the aggregate QA/QC procedures regarding geospatial and data management projects and tasks. This internal Tetra Tech documentation is designed for analysts and project managers to have QA/QC information readily available during project start up to aid in developing QAPPs, as well as in closing out projects, to document QC tasks performed for each project. Specific procedures are project-specific and require the input of analysts and project managers to determine the best course of QC measures to apply. In most cases, all of the information and procedures described in this document will not apply to each project, but rather project managers can pick and choose which apply to their project. The information described in this document is designed to provide general QA/QC background material related to geospatial and data management tasks.

2 Task Order Setup Procedures

The keys to successful contract start-up for Tetra Tech are effective communication, familiarity with the program goals of the client, key personnel who have worked on similar projects, and sufficient numbers of staff who can be tasked to work on initial Work Assignments or Task Orders. Upon receipt of a Work Assignment, the Tetra Tech Program Manager will contact the client Program Officer and Work Assignment Contracting Officer Representative to ensure clear, mutual understanding of the Work Assignment's objectives and the desired output, as well as the type, level, and location of expertise required to meet the objectives, schedule, and budget.

The Tetra Tech Program Manager will circulate copies of the Work Assignment to the Program Management Team, including the QA Officer and key personnel, for their input on staffing, QA requirements, and logistical issues identified in the statement of work. While managing a Work Assignment, the appropriate Tetra Tech Work Assignment Leader will have the primary day-to-day contact with the EPA Work Assignment Contracting Officer Representative. This approach allows EPA to work directly with the person conducting or supervising the Work Assignment.

3 Data Check-In

Input Data Integrity
Data are spot-checked to detect potential data entry errors. In addition, Tetra Tech may use a customized user input interface that performs certain appropriate checks on data as they are being manually entered when a project involves the input of large quantities of data, thereby reducing the potential for incorrect data entry. In any project with automated processing it is important to visually inspect the GIS data to check for adherence to database design, attribute accuracy, logical consistency and referential integrity.

Assessments of Processed Data
The ability of a desktop geospatial product to accurately characterize the conditions in the project area are dependent on the quality of data entering the process and imported into a geographic information system (GIS). QC procedures are implemented during data processing activities, and technical reviews of processed data are conducted by qualified personnel. Tetra Tech follows guidance on data management, information security, record management, and data processing provided or referenced by the client, including *Data Standards* (EPA CIO 2133.0), *Information Resources Management Policy Manual* (EPA CIO 2100.0), *Records*

Management Manual (EPA CIO 2155.0), and *Records Management Policy* (EPA CIO 2155.1) available on the Internet at http://www.epa.gov/irmpoli8.

4 Automation Plan

Large datasets require automated processes to ensure efficiency and accuracy. Macros can be a way to automate multiple processes in sequence. When using a macro in a database-related software, the macro must be coded in a way that the result can be independently followed and replicated. In these cases it is important to be able to trace an error back to the step it was introduced.

For GIS related processes, Earth Systems Research Institute's (ESRI's) ModelBuilder tool can help do the same task by linking multistep processes together and producing a visual flow diagram to track automated processing. In any project with automated processing it is important to visually inspect the GIS data to check for adherence to database design, attribute accuracy, logical consistency and referential integrity. Any visual inspection will be coupled with automated QA to ensure formulas and GIS algorithms have worked to their desired effect.

5 Data Organization

All information that is received by the project will be tracked and maintained from the moment of receipt, even though it may not be used in the final products for various reasons. Submitted and retrieved information, including suggested data sources and citations, will be immediately recorded to allow traceability throughout the entire lifecycle. Collected data will be stored via a directory structure that will allow Tetra Tech to work on and analyze copies of the data, while preserving the original versions. This will be accomplished by creating a 'RAW' and 'WORKING' directory structure that Tetra Tech has successfully used in the past.

Throughout the actual GIS data processing, analysis, and layout, a GIS practitioner will generate many versions of a shapefile. This includes all the edits needed to suit the particular function of the project. The final shapefiles will reflect the final .mxd (project file) as well as all the final versions of shapefiles used to create the .mxd. A "Test_Shapefile" folder may house all the separate versions of shapefiles. This includes all the spatial joins, clips, projections, or anything else that was not used in the final product. Additionally, a "Draft_Shapefile" folder may house edited versions that needed to be updated with more current data and shapefiles that were used for a portion of the project but not the final output.

6 Spatial Data Quality Assurance (QA)/QC

There are many considerations for spatial data QA/QC that must be adapted for each geospatial project. These considerations include the following, which were adapted from the ESRI GIS software developer:

- GIS Data completeness, consistency, accuracy, and resolution (including projection).
- Identifying errors visually in ArcMap.
- Creating methods (data workflow) for project processes, including QC workflow for associated processes.
- Noting and tracking data errors either within attribute table fields or in associated project documentation.

- Checking schema (names, fields, and coordinate systems); Checking attributes (missing or bad values).
- Visual review techniques: Performing visual QC; Setting symbols and labels; labeling techniques for points, lines, and polygons.

Tetra Tech will determine in consultation with the EPA Project Manager how spatial data QA/QC will be implemented for a particular project.

7 Attribute Data QA/QC

All geospatial data (shapefiles) downloaded from publically available online data sources will have associated attribute data contained within their respective database files. These attribute data quantify and occasionally narratively describe the spatial data within tabular fields. These data should be evaluated under the same measurement performance criteria that traditional data sources (spreadsheets and databases) are evaluated.

Measurement performance criteria that will be used for data handling for any given project will include accuracy and completeness. Tetra Tech will also evaluate GIS metadata against the FGDC metadata standard to determine whether the GIS data are suitable for use for a given project. Tetra Tech will provide a description of the data evaluation factors and limits (as determined in consultation with the client) in the report of data collected. Whenever possible, data will be downloaded electronically from various electronic sources to reduce scanning of hard copy data.

8 Metadata QA/QC

Many projects will rely on secondary data. Geospatial metadata is used throughout the project lifecycle. All personnel that download geospatial secondary data become Metadata Stewards.

Tetra Tech Metadata Stewards will evaluate GIS metadata against the Federal Geographic Data Committee (FGDC) (www.fgdc.gov) metadata standard to determine whether the GIS data are suitable for use for any given project. The FGDC has developed a metadata standard for geospatial data generated for and by all federal agencies which all federal agencies are to follow according to Executive Order 12906, “Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure.” Detailed metadata indicating the source, scale, resolution, accuracy, and completeness provide a basis to assess the adequacy of existing data for use (USEPA 2000).

If requested by the client through written technical direction, additional GIS QA/QC requirements can be addressed; examples include:

a. Full FGDC compliant metadata in XML format.

i. Use the appropriate metadata profile described in the FGDC Content Standard for Digital Geospatial Metadata (CSDGM), such as Biological Profile, Shoreline Profile, and Remote Sensing Profile. Metadata profiles can be obtained from http://www.fgdc.gov/metadata.

b. A single file represents the entire dataset (layer).

c. Each field that is mandatory and/or applicable must be described in the metadata.

d. The EPA Metadata Editor (EME) is used to create metadata (https://edg.epa.gov/EME/) and export to XML if using ESRI software.

e. Secondary data is accompanied by a metadata validation file. If a metadata validation file does not exist, metadata validation is performed prior to including the dataset in the project. This is to ensure and document that the dataset meets the needs of the intended use.

f. Where possible, extramural organizations are encouraged to use the EME. This facilitates subsequent review, collation, and verification of metadata validation.

g. The appropriate Geospatial Accuracy Tier noted in Appendix A of the EPA National Geospatial Data Policy is included as Supplemental Information. This facilitates collation of data and information related to scale.

h. Where practical, transition to the ISO 19115 metadata standards (North American Profile) is encouraged. At this moment, ISO metadata is optional.

9 Product Review

Tetra Tech will document the data collected in the final report of each project, as well as, a description of all QC activities and analyses where data analysis assumptions or procedures were not obvious. Summary statistics and discussion will include the following:

- Quality of secondary data (requirements will be determined in consultation with the client)
- Accuracy of extraction/interpretation of pertinent data from secondary data sources for use in deliverables.
- 10 percent of extractions/interpretations will be checked (100 percent of discrepancies will be resolved).
- Accuracy of data transfers, data entries 10 percent of data transfers will be checked (100 percent of discrepancies will be resolved).
- Hand-entered data will be checked (100 percent of discrepancies will be resolved).
- Accuracy of data conversions, including reformatting, will be checked (100 percent of discrepancies will be resolved).

10 Data Maintenance

Most work that Tetra Tech conducts involves acquiring and processing data, and generating reports and documents, all of which require the maintenance of computer resources. Tetra Tech's computers are either covered by on-site service agreements or serviced by in-house specialists. When a problem with a microcomputer occurs, in-house computer specialists diagnose the trouble and correct it if possible.

When outside assistance is necessary, the computer specialists will call the appropriate vendor. For other computer equipment requiring outside repair and not covered by a service contract, local computer service companies are used on a time-and-materials basis. Routine maintenance of microcomputers is performed by in-house computer specialists. Electric power to each microcomputer flows through a surge suppressor to protect electronic components from potentially damaging voltage spikes. All computer users have been instructed on the importance of routinely archiving work assignment data files from hard drive to compact disc storage. The Fairfax office network server is backed up on tape nightly during the week. Screening for viruses on electronic files loaded on microcomputers or the network is standard company policy and practice. Automated screening systems have been placed on all Tetra Tech systems and are updated regularly to ensure that viruses are identified and destroyed. Annual maintenance of software is performed to keep up with evolutionary changes in computer storage, media, and programs.

11 Version Control

Data can be managed in a number of different platforms. GIS versioning can be managed through ESRI's ArcCatalog via folder and file naming conventions. Date of creation, ArcMap processing tool, and project name should all be reflected in the file name. Including spaces and non-traditional characters in file names is required for GIS processing and management.

12 Data Transfer/Transmittal

Data that are transferred among databases will be checked for completeness at the time of transfer by enumerating the numbers of records in the original and final data sets. Data transfers will be tagged with upload dates and times to accommodate completeness reviews. If data transfer is incomplete, the missing records will be sought and transferred individually if they are valid. A second round of completeness checks will ensue after successive transfers. Once data sets are compiled, the complete set of data value distributions will be analyzed to identify outliers that may result from data entry errors or erroneous unit conversions. Outliers will be identified and resolved. Valid outliers can occur and will not be eliminated if the experienced analyst thinks they are plausible. Outliers that are not plausible or show a pattern of potential error will be brought to the attention of the original data supplier (if possible) and will be excluded from analysis until the original data supplier can confirm their validity.

The accuracy of the transfer of data from electronic databases to the project database(s) will be determined by checking whether data from the original database have been transferred to appropriate rows and columns, whether the same number of decimal places after the decimal point in the original database has been used, and whether the same units from the original database have been used. Such QC checks will be performed on approximately every 10th line of data in each electronic data set transferred into the project database(s). This procedure will aid the evaluation process by improving consistency in data transfers.

Spatial data such as shapefiles and model input files are often composed of a family of files that need to be stored together to function. When transferring spatial data, consider that all of these files should be transferred together and that project files such as .mxds will need to be relinked after the files have been moved. Geodatabases are also available, and they are becoming more common for storing multiple spatial data sets for a project while maintaining data set relationships, behaviors, annotations, and metadata.

Data generated within a GIS platform will likely be too large to deliver over email. In these cases setting up an FTP site may be necessary. Ensuring the file size that is in the product posted to

FTP is the same size as the project downloaded by the end user is a good way to ensure all data has been successfully transmitted.

13 Data Projections

All spatial data should have the same coordinate system for comparison; therefore, transformations are often necessary. Coordinate systems include both a geodetic datum and a projection type. A geodetic datum describes the model that was used to match the location of features on the earth's surface to coordinates on the map. Common datums include the World Geodetic System 1984 (WGS84) for a good representation for the world as a whole and the North American Datum 1983 (NAD83) or 1927 (NAD27) for a representation for North America. A projection type (e.g., the Universal Transverse Mercator [UTM] or state plane) is a visual representation of the earth's curved surface on a flat computer screen or paper. Often, if available, a state plane coordinate system or other state system is the most accurate system for a particular project area. Spatial data sets can be in the same projection but be referenced to different datums and therefore have different coordinate values (e.g., latitude and longitude or UTM). To fully represent a location spatially and avoid errors or confusion, coordinates are needed along with the datum. The difference between WGS84 and NAD83 is basically negligible (about 1 meter); however, the difference between NAD27 and NAD83 (and WGS 1984) varies from about 10 meters in the Great Lakes area to 100 meters on the west coast and up to 400 meters in Hawaii. NAD27 is still used as the basis for most USGS topographic maps, but NAD83 was created to provide a more accurate representation of the earth's ellipsoid shape. By default, most GPS units export points in WGS84, but settings can be changed to display points using different systems. Significant error can be introduced when data with different or unknown datums are introduced, including errors in distance or area measurement and errors in relating the spatial location of features between data sets. GIS software or mathematical algorithms allow for the conversion of spatial data from one coordinate system to another.

14 Storage and Archives

Data storage involves keeping the data in such a way that they are not degraded or compromised and that any datum desired can be retrieved. At every stage of data processing at which a permanent collection of data is stored, a separate copy is maintained for purposes of integrity and security. Data are securely archived in a suitable manner. Aspects such as storage media, conditions, location, access by authorized personnel, and retention time are addressed in consultation with the client. Before archiving, the project manager ensures that all data sets are complete, with all of the client-required data standards honored.

Tetra Tech will store all computer files associated with the project in a project subdirectory (subject to regular system backups). Tetra Tech will maintain version control of draft and final deliverables by indicating the preparation date or revision number in the file name. The length of archival will be decided upon consultation with Client specifications.

15 Training Requirements

Work assignments or task orders and quality assurance documents will be distributed to all project participants for review and reference. All relevant project personnel will have expertise in collecting and evaluating and analyzing GIS data. In addition, all relevant project personnel will have working knowledge of any additional software necessary to complete the project requirements.

GIS Analysts should have access to ArcGIS software no earlier than version 10.0 for file compatibility purposes. All project personnel will have expertise in environmental sciences, as well as knowledge of the quality system for the project and this knowledge and expertise will be enumerated in project documentation.

(This page is intentionally blank.)