1.0 PROJECT MANAGEMENT – ORGANIZATION & RESPONSIBILITIES

1.1 Title and Approval Page

Generic Secondary Data Quality Assurance Project Plan (QAPP) for the Connecticut (CT) Statewide Lake Nutrient Total Maximum Daily Load (TMDL)

Prepared by[1]:
Horsley Witten Group (HW)
90 Route 6A
Sandwich, MA 02563

FB Environmental Associates (FBE)
97A Exchange Street, Suite 305
Portland, ME 04101

Prepared for:
Region 1 - US Environmental Protection Agency (USEPA Region 1)
5 Post Office Square
Boston, MA 02109

FINAL | May 13, 2020

EPA RFA Number 20056

STEVEN WINNETT Digitally signed by STEVEN WINNETT Date: 2020.05.18 14:38:54 -04'00'
_____ _____
Steve Winnett, USEPA Region 1 Contracting Office Representative Date

MARY GARREN Digitally signed by MARY GARREN Date: 2020.05.18 11:24:45 -04'00'
_____ _____
Mary Garren, USEPA Region 1 CT TMDL Coordinator Date

TOBIAS STOVER Digitally signed by TOBIAS STOVER Date: 2020.05.18 14:21:43 -04'00'
_____ _____
Toby Stover, USEPA Region 1 NH TMDL/Nutrient Criteria Coordinator Date

Traci Iott 5/18/20
_____ _____
Traci Iott, CT DEEP Supervising Environmental Analyst Date

_____ __5/13/20_____
Richard Claytor, President, Horsley Witten Group Date

_____ __5/13/20_____
Forrest Bell, Principal, FB Environmental Associates Date

NORA CONLON Digitally signed by NORA CONLON Date: 2020.05.18 15:14:19 -04'00'
_____ _____
Nora Conlon, USEPA Region 1 Quality Assurance Officer Date

[1] Portions of content were drawn heavily from the *QAPP for Bantam Lake Nutrient TMDL Model* prepared by Comprehensive Environmental, Inc. and HydroAnalysis, Inc, November 28, 2018, as well as the *Bantam Lake Nutrient TMDL Model Modeling Report (DRAFT)* prepared by Comprehensive Environmental, Inc., December 2019.

1.2 Table of Contents

1.0 PROJECT MANAGEMENT – ORGANIZATION & RESPONSIBILITIES 1
1.1 Title and Approval Page 1
1.2 Table of Contents 2
1.3 Distribution List 4
1.4 Project Organization 4
1.5 Purpose of Study, Background Information, and Problem Definition 6
1.6 Overview of Project Tasks 8
1.7 Quality Objectives and Criteria 9
1.7.1 Measurement Data Acceptance Criteria 9
1.7.2 Model Performance and Acceptance Criteria 10
1.8 Special Training and Certification 11
1.9 Documentation and Records 11
2.0 DATA MANAGEMENT & ACQUISITION 11
2.1 Data Management 11
2.2 Data Acquisition 11
2.3 Intended Use of Existing Data 15
2.3.1 Period Selection 15
2.3.2 For LLRM Input 16
2.3.3 For BATHTUB Input 20
2.3.4 For Nutrient Load Reduction Analysis 23
2.4 Limitation on the Use of Existing Data 24
3.0 ASSESSMENTS AND OVERSIGHT 25
3.1 Project Oversight 25
3.2 Project Documentation 25
3.3 Corrective Actions 25
4.0 MODEL & DATA VERIFICATION, VALIDATION, AND EVALUATION 26
4.1 Data Verification and Validation 26
4.2 Data Evaluation 26
4.3 Model Parameterization (Calibration) 27
4.4 Model Corroboration (Validation and Simulation) 30
4.5 Reconciliation with User Requirements 31
5.0 PROJECT REPORTING 31
6.0 REFERENCES 31
7.0 APPENDIX A: Land Use Data Source Comparison 33
8.0 APPENDIX B: R script for filtering eBird data 38

List of Tables

Table 1. Generic Quality Assurance Project Plan (QAPP) distribution list. 4

Table 2. Data acceptance criteria for secondary data. 9

Table 3. Model calibration/validation targets (Donigian, 2002) 10

Table 4. Sources of existing data, sorted by the LLRM and BATHTUB model. If data were listed first under LLRM, then data were not repeated under BATHTUB. 12

Table 5. Runoff and baseflow export coefficients for precipitation, phosphorus, and nitrogen based on minimum, median, and maximum values from published scientific literature referenced in the LLRM documentation; default values used in the original LLRM spreadsheet; and final values used in the Bantam Lake LLRM (CEI, Inc., 2020). 18

Table 6. Parameters and defining ranges for the trophic state of lakes in Connecticut. Adapted from the State of Connecticut Department of Energy and Environmental Protection Water Quality Standards 2013 (Sec. 22a-426-6). 24

Table 7. Attenuation values for water, phosphorus, and nitrogen based on sub-basin characteristics. Gray shading indicates model default values. These attenuation values represent starting points for the calibration process but can be adjusted further based on other sub-basin characteristics such as slope grade. 28

List of Figures

Figure 1. Project organization chart. Personnel in gray shading may be subject to change, depending on CT DEEP staffing resources. 6

Figure 2. Workflow for achieving EPA-approved TMDLs and Watershed-Based Plans (WBPs) for all nutrient-impaired lakes and impoundments in Connecticut. 8

1.3 Distribution List

This generic QAPP, along with any amendments, will be distributed to the key personnel listed in Table 1, as well as to all federal, state, contractor, and subcontractor personnel involved in projects that employ this generic QAPP.

Table 1. Generic Quality Assurance Project Plan (QAPP) distribution list.

Name, Title, Organization	Contact Information	Mailing Address
Steven Winnett USEPA Region 1 New England TMDL Coordinator and Contracting Office Representative	617-918-1687 winnett.steven@epa.gov	5 Post Office Square, Suite 100 (OEP06-2) Boston, MA 02109
Mary Garren USEPA Region 1 CT TMDL Coordinator	617-918-1322 garren.mary@epa.gov	5 Post Office Square, Suite 100 (OEP06-2) Boston, MA 02109
Toby Stover USEPA Region 1 NH TMDL Coordinator, Numeric Nutrient Criteria Coordinator	617-918-1604 stover.toby@epa.gov	5 Post Office Square, Suite 100 (OEP06-2) Boston, MA 02109
Nora Conlon USEPA Region 1 Quality Assurance Officer	617-918-8335 Conlon.nora@epa.gov	EPA New England Regional Laboratory 11 Technology Drive (EQA) North Chelmsford, MA 01863-2431
Traci Iott CT Department of Energy & Environmental Protection Supervising Environmental Analyst	860-424-3082 traci.iott@ct.gov	CT Department of Energy & Environmental Protection 79 Elm Street Hartford, CT 06106-5127
Richard Claytor Horsley Witten Group President	508-367-8002 rclaytor@horsleywitten.com	90 Route 6A, Unit #1 Sandwich, MA 02563
Anne Kitchell Horsley Witten Group Associate Principal	508-833-6600 akitchell@horsleywitten.com	90 Route 6A, Unit #1 Sandwich, MA 02563
Gemma Kite Horsley Witten Group Senior Environmental Engineer	508-833-6600 gkite@horsleywitten.com	90 Route 6A, Unit #1 Sandwich, MA 02563
Forrest Bell FB Environmental Associates Principal	207-221-6699 info@fbenvironmental.com	97A Exchange Street, Suite 305 Portland, ME 04101
Laura Diemer FB Environmental Associates Environmental Monitoring Lead, Project Manager	603-828-1456 laurad@fbenvironmental.com	170 West Rd, Suite 6 Portsmouth, NH 03801
Jeffrey Walker Walker Environmental Research, LLC. Principal	978-985-5612 jeff@walkerenvres.com	Brunswick, ME

1.4 Project Organization

Project organization for this generic QAPP involves key personnel at the USEPA Region 1 and the CT Department of Energy & Environmental Protection (CT DEEP), as well as current and any future contractor personnel (Table 1, Figure 1). The first project phase includes development and application of this generic QAPP in modeling by contractor personnel from HWG and FBE. Subsequent project phases will include application of this generic QAPP in modeling by CT DEEP personnel. The principal users of the generic QAPP will be the USEPA Region 1 and CT DEEP, who will use the generic QAPP to assist with preparation and execution of the selected models for TMDL load reduction analyses applicable to nutrient impaired lakes

and impoundments in Connecticut. The model work outlined in this QAPP may also be applicable to unimpaired lakes and impoundments in Connecticut. The roles and responsibilities of key project personnel for the generic QAPP are summarized below.

- Steve Winnett is the New England TMDL Coordinator for the USEPA Region 1 (and the EPA Contracting Office Representative for the first project phase) and will be ultimately responsible for signing off on and maintaining the generic QAPP, as well as all contractual direction and necessary actions for the current project phase.
- Mary Garren is the CT TMDL Coordinator for the USEPA Region 1 (and the EPA Technical Lead) and will be responsible for reviewing the generic QAPP, as well as for projects employing the generic QAPP.
- Toby Stover is the NH TMDL Coordinator and Numeric Nutrient Criteria Coordinator for the USEPA Region 1 (and the EPA Technical Advisor) and will be responsible for reviewing the generic QAPP, as well as for projects employing the generic QAPP.
- Nora Conlon is the Quality Assurance Officer for the USEPA Region 1 and will be responsible for reviewing and approving the generic QAPP and all subsequent amendments.
- Traci Iott is the Supervising Environmental Analyst for the CT DEEP and will be responsible for overseeing the use of the generic QAPP with preparing and executing the selected models for lakes and impoundments in Connecticut. Traci will sign off on the generic QAPP on behalf of CT DEEP as the primary user of the document.
- Project Leader, Project QA Officer, and Project Support will be CT DEEP staff responsible for preparing and executing selected models for lakes and impoundments in Connecticut in subsequent project phases.
- Richard Claytor is President of HWG and will be responsible for ensuring overall completion of the first project phase.
- Anne Kitchell is an Associate Principal for HWG and serves as the primary contractor point of contact for the first phase of the project. She will be responsible for completing HWG project tasks and overseeing FBE project tasks.
- Forrest Bell is the Principal of FBE and will be responsible for ensuring completion of FBE tasks in the first phase of the project.
- Gemma Kite is a Senior Environmental Engineer for HWG and will be responsible for completing the BATHTUB portion of the modeling in the first phase of the project.
- Laura Diemer is the Environmental Monitoring Lead and Project Manager for FBE and serves as a secondary contractor point of contact for the first phase of the project. She will be responsible for completing the LLRM portion of the modeling in the first phase of the project.
- Jeffrey Walker is the Principal of Walker Environmental Research, LLC and is a consulting Water Resources Modeler for FBE. He will be responsible for reviewing the model and associated documentation in the first project phase.

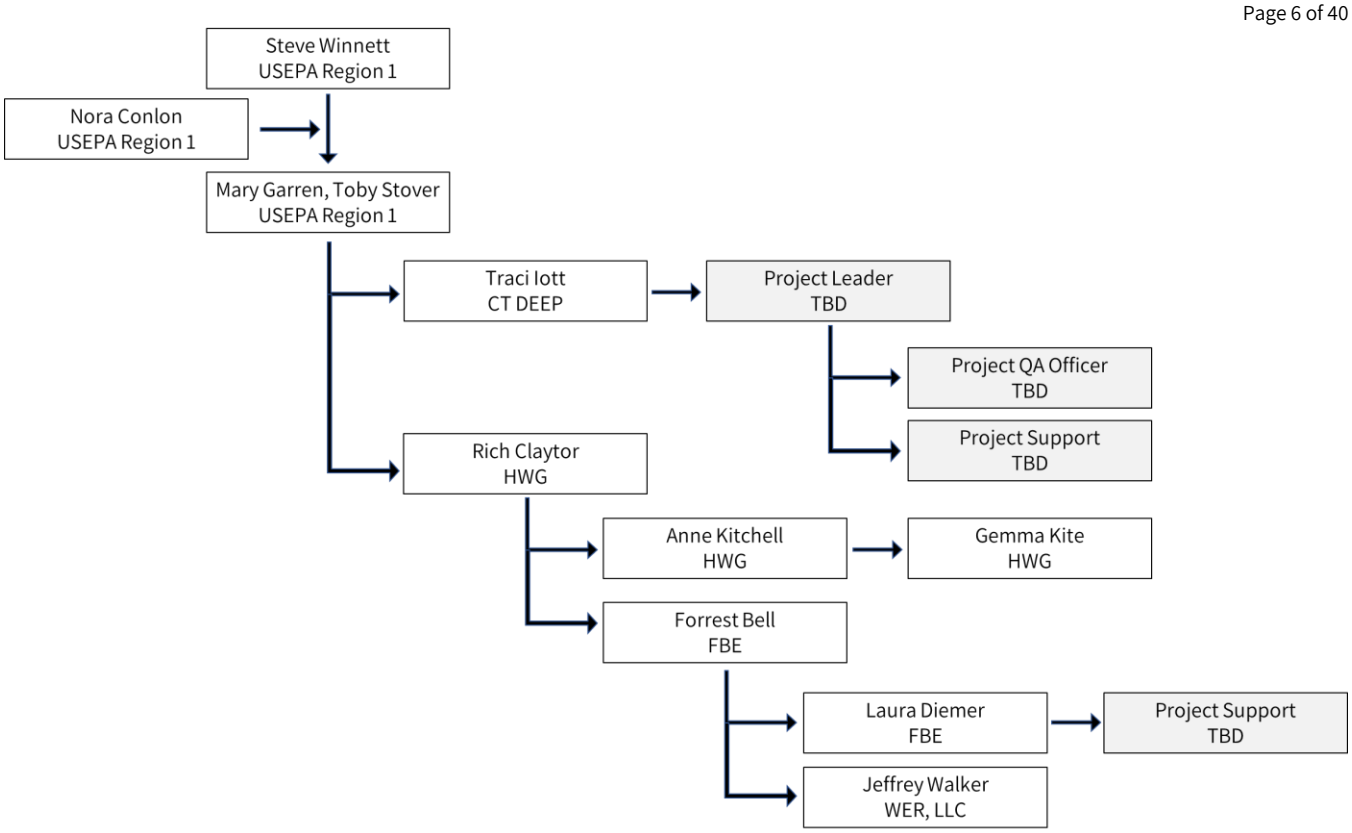


Figure 1. Project organization chart. Personnel in gray shading may be subject to change, depending on staffing resources.

1.5 Purpose of Study, Background Information, and Problem Definition

To limit the process of eutrophication and the formation of Harmful Algal Blooms (HABs) in lakes and impoundments, the CT DEEP has placed a high priority on addressing the impacts of excess nutrient loading on water quality. To this end, the CT DEEP will develop a statewide lake nutrient TMDL core document along with watershed-specific appendices to document and address waterbody-specific conditions. The core document will provide general information and resources applicable to nutrient and HAB management for all nutrient-impaired lakes and impoundments in Connecticut. Template watershed-specific appendices will be developed to help guide future development of all watershed-specific appendices for nutrient-impaired lakes and impoundments in Connecticut.

Along with being the first statewide TMDL written for nutrient-impaired lakes and impoundments in New England, this statewide TMDL will also incorporate the nine elements of an EPA watershed-based plan to the maximum extent possible. These elements will be incorporated more generally in the core document and more specifically (as available) in the appendices. A watershed-based plan addendum that provides additional, more specific information not included in the appendix for each impaired lake and impoundment will also be developed. Combining these two requirements (TMDL and watershed-based plan) will help to streamline environmental protection and restoration efforts so that the core document, watershed-specific appendices, and watershed-based plan addendums could be approvable as watershed-based plans.

The Lake Loading Response Model (LLRM) and BATHTUB model were selected for the evaluation of nutrient loading for each waterbody to attain its natural trophic status. Nutrient loads for lakes and impoundments will be evaluated against modeled changes in lake trophic status, as defined in Section 22a-426-6 of Connecticut's Water Quality Standards Regulations. For this project, watershed loading estimates from the LLRM will be used as inputs to BATHTUB for in-lake nutrient modeling.

The LLRM is an Excel-based model that uses environmental data to develop a water and nutrient loading budget for lakes/ponds and their tributaries (AECOM, 2009). Water and nutrient loads (in the form of mass and concentration) are derived from various sources in the watershed, through tributary and lake/pond sub-basins, to the outlet of the study waterbody. The model incorporates data about watershed and sub-basin boundaries, land cover, point sources, septic systems, waterfowl, rainfall, volume and surface area, and internal nutrient loading. These data are combined with coefficients, attenuation factors, and equations from scientific literature on lakes, rivers, and nutrient cycles. The model can be used to quantify loads from current and future pollution sources, estimate pollution level limits and water quality goals, and guide watershed improvement projects.

BATHTUB is a steady-state, empirical eutrophication model designed to evaluate the effect of nutrient loading on common eutrophication response parameters such as algal growth, transparency, and hypolimnetic oxygen depletion in lakes, ponds, and reservoirs. The model was originally developed by W. W. Walker, Jr., Ph.D. for the U.S. Army Corps of Engineers, and calibrated using a comprehensive dataset of reservoirs located across the U.S. The model computes water and nutrient mass balances of the target waterbody, which are then applied to empirical relationships to predict eutrophication response. The model can be configured to represent waterbodies as zero-dimensional (fully mixed) or one-dimensional (horizontally segmented) systems and can simulate multiple reservoir systems by routing flows and loads between waterbodies. The model can be used to both diagnose existing water quality impairments as well as to predict future conditions under various nutrient loading scenarios. By incorporating empirical relationships between nutrient loading and eutrophication response, the model can be applied to systems with a wide range of data availability. In addition to the eutrophication model, BATHTUB also comes with two utilities for: (1) estimating nutrient loads from tributary inflows based on direct observational data and water quality samples; and (2) processing in-situ water quality observations to estimate seasonal and long-term average conditions within the target water body. Finally, BATHTUB includes robust algorithms for performing uncertainty analyses, which are critical for informed decision making.

This generic QAPP describes the quality system that will be implemented for model development and output, including the data quality objectives for the LLRM and BATHTUB model and the quality control steps and techniques to be followed to achieve the data quality objectives. This generic QAPP addresses the use of secondary data (i.e., data collected for another purpose or collected by an organization not under the scope of this QAPP) to support model development and output. Model output for the TMDL will be used to conduct a watershed nutrient load reduction analysis to estimate watershed-based annual nutrient load reductions necessary to attain natural trophic status. These data will ultimately be used in the development of the watershed-specific appendices, part of the CT Statewide Lake Nutrient TMDL. See workflow in Figure 2.

The purpose of the overall project is to assist USEPA Region 1 and CT DEEP with developing a statewide lake nutrient TMDL with watershed-based plan addendums for specific waterbodies. Project objectives include:

- Setup, calibrate, and validate the LLRM and BATHTUB model for each lake and impoundment.
- Using the calibrated and validated models, calculate nutrient loading capacities and load reductions necessary to meet water quality targets for each lake, including total phosphorus (TP), total nitrogen (TN), chlorophyll-a concentration, transparency, and hypolimnetic oxygen depletion rate.

Results will be used by the USEPA Region 1, state agencies, and local municipalities or other key stakeholder groups to address excess nutrient loading and HAB formation in lakes and impoundments in Connecticut.

This generic QAPP was developed in accordance with the following guidance documents: *EPA Guidance for Quality Assurance Project Plans for Modeling* (EPA QA/G-5M) (EPA, 2002), *EPA New England Environmental Data Review Program Guidance* (EPA, 2018), *EPA New England QAPP Guidance for Projects Using Secondary Data* (EPA, 2009). Some of the language used in this generic QAPP was drawn directly from existing EPA-approved QAPPs or relevant documents, including the *QAPP for Bantam Lake Nutrient TMDL Model* (CEI, Inc. & HydroAnalysis, Inc., 2018), the LLRM User Guide (AECOM, 2009), and the *Bantam Lake Nutrient TMDL Model Modeling Report (FINAL)* (CEI, Inc., 2020).

1.6 Overview of Project Tasks

The following project tasks outline a series of steps to setup and execute the LLRM and BATHTUB model for individual lakes and impoundments. As part of the generic QAPP, these tasks are described at a general level to be applicable to any lake or impoundment in the state. More detailed descriptions of modeling inputs and assumptions, especially those that may differ from the generic QAPP, will be documented in a methodology report for each modeled lake and impoundment in Connecticut (refer to Figure 2 and definition in Task 4 below).

Task 1: Data Acquisition

Assemble, review, and format secondary data for the LLRM and BATHTUB model inputs for each lake. Refer to Section 2 for a general list of secondary data files necessary for inputs to each model.

Task 2: Model Setup & Execution

Setup and calibrate the LLRM model for each lake. Using the watershed loading results from the LLRM as inputs, calibrate and validate the BATHTUB in-lake water quality model. The latest model and documentation will be used for the LLRM (v. 2020, an expanded 2009 version from the Bantam Lake model version (CEI, Inc., 2020), which will be further refined under this project) and BATHTUB model (v. 6.1). Model execution will follow recommendations and procedures outlined in the user guides and/or QAPPs for both models (refer to AECOM, 2009; Walker, 1999; Walker, 2006).

Task 3: Nutrient Load Reduction Analysis

Use the calibrated BATHTUB model to calculate nutrient loading capacities and load reductions necessary to meet water quality targets for each lake, which may include TP, TN, chlorophyll a, and transparency. Water quality targets will be determined by CT DEEP.

Task 4: Documentation

Modeling methodology, approach, and/or decision-making processes and rationale, as well as a complete list of secondary data files used, will be documented in a methodology report (see CEI, Inc., 2019 or 2020 for examples). Deviations from the generic QAPP will be identified and justified. The final report will summarize the model setup, calibration, validation, and application to calculate nutrient load reductions needed to meet the water quality targets for each lake. Provide all data and related files used in the modeling.

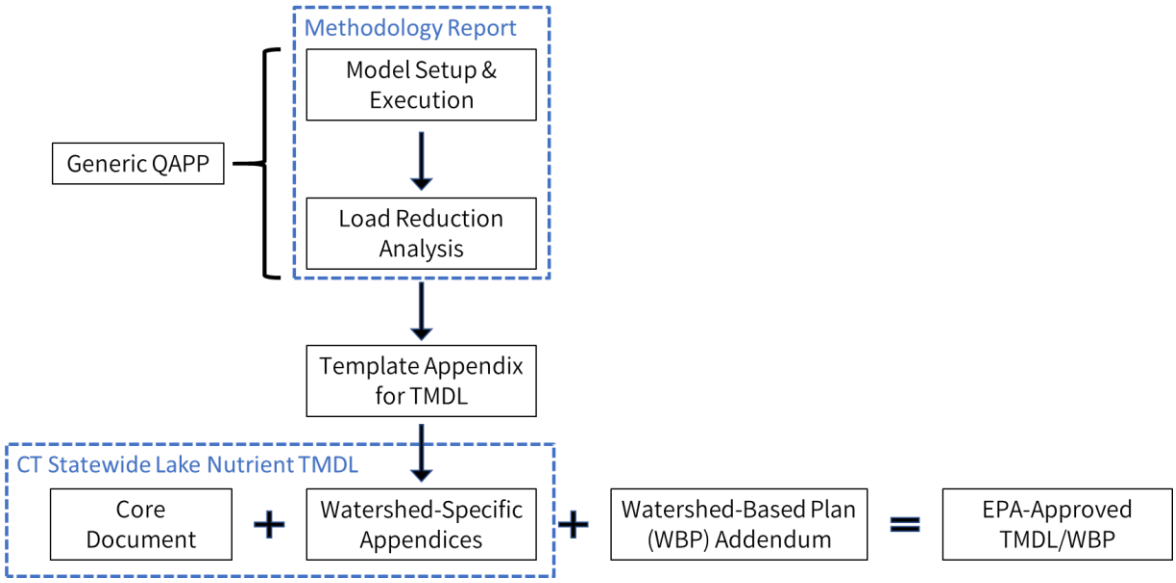


Figure 2. Workflow for achieving EPA-approved TMDLs and Watershed-Based Plans (WBPs) for lakes and impoundments in Connecticut.

1.7 Quality Objectives and Criteria

Data quality objectives and criteria for the project will ensure that data used to support modeling and TMDL determinations are scientifically valid and defensible, with a high level of transparency and data-sharing capabilities. The end users of the final products of the project are US EPA Region 1 and CT DEEP who will use the results to inform development of watershed-specific appendices, which will be combined with the CT Statewide Lake Nutrient TMDL core document and watershed-based plan addendum to achieve EPA-approvable plans for lakes and impoundments in Connecticut.

Data quality objectives and criteria for secondary data used in this project will be addressed by: (1) evaluating the quality of the data used (Section 1.7.1, Task 1), and (2) assessing the results or 'performance' of the model application (Section 1.7.2, Task 2). All other secondary data types will be gathered and assessed for inclusion in analyses based on the most recent, relevant files (see Table 4; see Section 4).

1.7.1 Measurement Data Acceptance Criteria

Model setup, calibration, validation, and application for this project will be accomplished using secondary data from qualified sources, such as governmental agencies. Data of known and documented quality are essential to the successful performance of the models because the model outputs will be used to support the TMDL decision-making process. Table 2 summarizes the acceptance criteria for secondary data that will be used in the setup and calibration of the model.

The organizations generating the secondary data that may be used in this project typically apply their own review and verification procedures to evaluate a dataset's integrity and conformance to data quality requirements. The quality of the data will be judged using information in source documents, from websites of origin, or directly from the authors. If the quality of the data can be adequately determined and the data meet quality criteria, the data will be used. If it is determined that no quality requirements exist or can be established for a dataset that must be used for this task, a case-by-case basis determination will be made regarding the use of the data. Data of unknown quality will not be used if the use of such data is believed to have a significant or disproportionate impact on the TMDL results.

Secondary data will be assembled, reviewed, and formatted in an Excel spreadsheet format ready for input to the LLRM and BATHTUB. Only data that meet the quality criteria in Table 2 will be marked as validated for use in the model. The final data used in the model, the period of record of the data, and the source of the data, along with justification for any excluded data, will be documented in the final documentation. Any use of secondary data of unknown quality, any data gaps, and the assumptions used in filling such gaps will also be documented.

Table 2. Data acceptance criteria for secondary data.

Quality Criterion	Description
Reasonableness	Datasets will be reviewed to identify anomalous values that may represent data entry or analytical errors. Such values will not be used without clarification from the agency providing the data.
Completeness	Datasets will be reviewed to determine the extent of gaps in space and time. It is likely that some data gaps will be evident. These gaps and the methods used to fill the gaps (if applicable) will be discussed in project documentation.
Comparability	Datasets from different sources will be compared by checking the methods used to collect the data and that the units of reporting are standardized.
Representativeness	Datasets will be evaluated to ensure that the reported variable and its spatial and temporal resolution are appropriate for the project. For example, datasets must be able to be reasonably aggregated (or disaggregated) to represent conditions in the model and must be representative of conditions during the simulation periods. The goal is for data and information to reflect typical variability (e.g., dry, wet, and average years).
Relevance	Data specific to the study site will be used. If needed, regional data and information that most closely represent the study site will be used.
Reliability	Sources of data and information will be considered reliable if they meet at least one of the following acceptance criteria:

Quality Criterion	Description
	• The information or data are from a peer-reviewed, government, industry-specific source. • The source is published through a credible source. • The author is engaged in a relevant field such that competent knowledge is expected (i.e., the author writes for an industry trade association publication versus a general newspaper). • The information was presented in a technical conference where it is subject to review by other industry experts. • The information or data are from a lake association / watershed group, deemed credible by CT DEEP. Sources of data that use unknown collection and data review procedures are considered less reliable, and will be used only if necessary to fill data gaps and following discussion with and approval by EPA.

1.7.2 Model Performance and Acceptance Criteria

EPA's *Guidance for Quality Assurance Project Plans for Modeling* (EPA QA/G-5M) discusses the importance of using performance criteria as the basis by which judgments are made on whether the model results are adequate to support the decisions required to address the project objectives.

A 'weight of evidence' approach that embodies the following principles will be adopted for model calibration in this project (Donigian, 2002):

- Given that models are approximations of natural systems, exact duplication of observed data is not a performance criterion. The model calibration process will measure, through comparability goals, the ability of the model to simulate observed data.
- No single procedure or statistic is widely accepted as measuring, nor capable of establishing, acceptable model performance. Thus, both quantitative (error statistics) and qualitative (graphical) comparisons of observed data and model results will be used to provide enough evidence to weight the decision of model acceptance or rejection.
- All model and observed data comparisons must recognize, either qualitatively or quantitatively, the inherent errors and uncertainty in both the model and the measurements of the observed data sets. These errors and uncertainties will be documented, where possible, in the final modeling methodology report.

Error! Reference source not found. lists the model calibration and validation guidelines that can be used to compare and evaluate the percent mean errors between model predictions and observed data. The ranges in **Error! Reference source not found.** are intended to be applied to spatially and/or temporally aggregated mean/median values; individual observations may show larger differences and still be deemed calibrated and validated for application so long as such excursions are limited (Donigian, 2002). If enough observed data exist, then model performance will be deemed acceptable where a performance evaluation of "good" or "very good" is attained. While the ranges in **Error! Reference source not found.** will be used as guidelines for model calibration and validation, they cannot be guaranteed to be met as they may not be achievable (e.g., if observed data are limited).

For model validation, a minimum of two years will be extracted from the calibration period, if at least 3 years of data are remaining for calibration. Validation of the model is not necessary if observed data are limited or if validation results do not meet guidelines but can be reasonably justified. An effort will be made to both calibrate and validate the models within the constraints of the available data set. See Section 4 for further discussion.

Table 3. Model calibration/validation guidelines (Donigian, 2002)

Variable	Percent Difference between Simulated and Observed Values		
	Very Good	Good	Fair
Water Quality / Nutrients	< 15	15 – 25	25 - 35

1.8 Special Training and Certification

To ensure that models are setup, calibrated, and validated and applied to TMDL determinations with sound scientific rigor, personnel performing and/or reviewing the modeling must hold advanced degrees in a related environmental field and have training in using the LLRM and BATHTUB model. It is highly recommended that an individual within or outside of CT DEEP, with extensive background in limnology and use of the LLRM and BATHTUB model, review model files.

1.9 Documentation and Records

All data and information collected and generated for the project will be stored to a backed-up network. A copy of all project files will be shared with and/or kept by the USEPA Region 1 and CT DEEP. Project files will be kept by the originator for a minimum of five years. Folder structures will be clearly labeled. Metadata associated with model and other files will be explicitly documented in accompanying text files or within the document itself (e.g., Excel spreadsheet).

2.0 DATA MANAGEMENT & ACQUISITION

Use of the LLRM and BATHTUB model will require the use of secondary data, also referred to as non-direct measurements. Secondary data are data that were collected under a different effort outside of the project. Secondary data to be used in the project will be collected from government publications and databases, scientific literature, industry published studies, lake associations/watershed groups, and other organizations. Table 2 summarizes the acceptance criteria for use of secondary data in the setup and calibration of the model.

2.1 Data Management

Consistent data management procedures will be used during pre-processing, model calibration, and post-processing stages of the project. Original data sources will be documented to identify the website or contact person that provided the data, data query parameters, and data request correspondence. The final documentation will include a summary of all final data (including complete citations) used in the setup, calibration, and validation of the model. Metadata will be housed in the spreadsheet files. Multiple versions of model files will be generated if significant changes are made or added, and all project files will be backed-up to a secure external hard-drive to protect from possible data loss or file corruption.

2.2 Data Acquisition

A variety of secondary data may be used for the project, including spatial files, water quality data, weather data, and literature references (if not associated with existing model metadata); see also the References section. A list of possible secondary data for use in modeling is provided in Table 4. Methodology reports will provide a more detailed list of data sources.

Table 4. Sources of existing data, sorted by the LLRM and BATHTUB model. If data were listed first under LLRM, then data were not repeated under BATHTUB.

Type of Data [File Name]	Description/Title	Format	Source (Date)	Intended Use	Possible Limitations	QA/QC	Analysis
LLRM							
Precipitation	Monthly precipitation data	.csv	NOAA NCEI, nearest quality-controlled land-based station for the critical period of interest	Model input	No weather stations in watershed; data gaps	Check for major gaps and use next nearest station to fill in	Sum monthly data to averaging period per year, then average annual data for critical period of interest
Aerial imagery [World_Imagery]	Satellite and aerial imagery	ArcGIS Layer	ESRI DigitalGlobe	Spatial reference, data verification	NA	NA	NA
Aerial imagery [Google Earth Pro Desktop]	Satellite imagery using Landsat 8	Google Earth Pro Desktop	Google DigitalGlobe	Spatial reference, data verification	Cannot use in ArcGIS for mapping	NA	NA
Topography [World_Topo_Map]	USGS/ESRI digitized topographic basemap	ArcGIS Layer	ESRI DigitalGlobe (updated 11/2018)	Spatial reference, data verification, manual sub-basin delineation checks	20-ft contour resolution coarse	NA	NA
Topography [Contour_2000_5ft]	5-ft contour lines for CT	.shp	CT DEEP GIS (Online)	Spatial reference, data verification, manual sub-basin delineation checks	NA	NA	NA
2018 Impaired Lakes [2018_Lake_Assessments]	Impaired lakes in CT based on 2018 assessment	.shp	CT DEEP (unpublished)	Sub-basin delineation determinations	NA	NA	NA
2016 Impaired Lakes [CT_Lakes_Nutrient_Impairments_2016]	Impaired lakes in CT based on 2016 assessment, subset to nutrient-impaired lakes	.shp	CT DEEP (unpublished)	Sub-basin delineation determinations	No changes made in 2018 assessment	NA	NA
Water Features [WATERBODY_POLY, WATERBODY_LINE]	Lakes and rivers	.shp	CT DEEP GIS (Online)	Sub-basin delineation determinations	NA	NA	NA
National Hydrography Dataset [NHDFlowlines, NHDWaterbody]	Lakes and rivers	.shp	USGS NHD (Online)	Basemap, spatial reference	NA	NA	NA
Watersheds and Drainage Basins [Watershed_gdb]	Regional, subregional, and local basins	.gdb with .shp	CT DEEP GIS (Online)	Sub-basin areas, routing, attenuation	NA	NA	May need to manually delineate to lake outflow
Wetlands [NWI Wetlands Mapper]	USFWS National Wetland Inventory (NWI)	Online mapper to extract .shp	USFWS (Online)	Sub-basin delineation and attenuation determinations	May not reflect site-specific wetlands	NA	NA
Land Cover [2016 National Land Cover Database]	2016 National Land Cover Database (NLCD)	Raster file	Multi-Resolution Land	Land use by sub-basin for model input	Coarse resolution	Convert grid codes to match	Not recommended for use in this project

Type of Data [File Name]	Description/Title	Format	Source (Date)	Intended Use	Possible Limitations	QA/QC	Analysis
			Characteristics Consortium (2016)			LLRM land use categories	
Land Cover [2015 UCONN CLEAR]	2015 UCONN CLEAR land use data for the State of Connecticut	Raster file	UCONN CLEAR (2015)	Land use by sub-basin for model input	Coarse resolution	Convert grid codes to match LLRM land use categories	Tabulate area of each land use type by sub-basin
Land Cover [TBD]	Land cover data for portions of the watershed outside of Connecticut	TBD	TBD	Land use by sub-basin for model input	TBD	Convert grid codes to match LLRM land use categories	Tabulate area of each land use type by sub-basin
Point Sources [Permits, Water Quality Data Portal online or other available sources]	Location and associated water quality data (nutrients, flow) for known point source discharges	.shp/.csv	CT DEEP; TBD	Model input	Data gaps; limited associated flow data	Check for major outliers; identify limitations with data gaps; calculate load (Q*C)	Aggregate according to Walker (1999) for averaging period during critical period of interest
Buildings [buildings2012]	Building polygons from 2012 statewide impervious surface layer for CT	.shp	UCONN/CTDEEP CTECO (2012)	Model input for septic system load estimate	Counts all buildings, including secondary structures	NA	Count number of buildings within 300 feet of lake; adjust count based on sewer coverage
Sewer Service Area [DRAFT_Sewer_Service]	Internal working draft of sewer service area for CT	.shp	CT DEEP (unpublished)	To filter building count	Internal working draft not yet complete; may not cover area completely so manual review for reasonableness may be necessary	NA	Eliminate buildings in the vicinity of sewer service area
Parcels [2010 Connecticut Parcels]	CT parcels used for protected open space mapping project	.shp	CT DEEP GIS Online	Alternative approach to septic system count	Parcel data not complete, caused gross underestimate of building count	NA	Not recommended for use in this project
Waterfowl [Migratory_Waterfowl]	Presence/absence data for waterfowl species in concentrated areas of migratory waterfowl in CT	.shp	CT DEEP GIS Online	Reference check for model input	NA	NA	Identify migratory waterfowl hotspots for possible use in determining bird count
eBird Online	Database of bird counts by site	.txt	eBird (2019)	Waterfowl estimate for model input	Inadequate site coverage for lake	Filter data to large waterfowl	Average weekly or monthly bird count

Type of Data [File Name]	Description/Title	Format	Source (Date)	Intended Use	Possible Limitations	QA/QC	Analysis
						in the averaging period for the critical period of interest	data to determine average birds per day in the averaging period for the critical period of interest
Water Quality Data	Tributary TP, TN	.csv or similar, database format	TBD (critical period of interest)	Model calibration and validation	TBD	TBD	Refer to Section 2.3.2 for details
USGS Flow Data	USGS daily or monthly flow data	.csv	USGS (critical period of interest)	Model input and calibration	No gages in watershed; data gaps	Check for major gaps and use next nearest gage to fill in; use drainage area ratio from gage for conversion	Sum to averaging period per year, multiply by the number of days or month in the period, then average annual flow data for the critical period of interest
BATHTUB							
Water Quality Data	Lake TP (epi/hypo), TN, Chl-a, SDT, DO/temp data	.csv	TBD (critical period of interest)	Model calibration; internal loading calculation for model input; estimate mixed layer depth and hypolimnetic depth/thickness; determine hydraulic segmentation	Data gaps	Check for major data gaps	Median by month for averaging period per year, then all data in the critical period of interest; see Segments, Morphometry, and Internal Loading for analysis discussion
Lake Bathymetry [Lake_Bathymetry_Poly]	Surface area by depth for a waterbody	.shp	CT DEEP GIS (Online)	Estimate volume and surface area; determine hydraulic segmentation; internal loading calculation for model input	Conflicting values from different sources	Default to this file unless otherwise justifiable	Calculate volume; review bathymetry for distinct areas
Diagnostic Reports or similar assessments	TBD	TBD	TBD	TBD	TBD	TBD	TBD
Water Level Data	Monthly average water level data	.csv	TBD (critical period of interest)	Model input related to change in storage during the averaging period	Data gaps	Check for major data gaps	Average difference between start and end of averaging period during the critical period of interest

2.3 Intended Use of Existing Data

The intended use of existing data for each data file is summarized in Table 4. More detailed discussion of the analysis and/or assumptions made when using these data files for model input are provided below in four sub-sections: Period Selection, For LLRM Input, For BATHTUB Input, and For Nutrient Load Reduction Analysis. For this project, the LLRM will only be used to estimate watershed phosphorus and nitrogen loading to a given lake. These external loading estimates will be input to the BATHTUB model to predict in-lake water quality for use in the nutrient load reduction analysis.

2.3.1 Period Selection

To make informed decisions about model input data (and how to best summarize those data), the project team will review available water quality data to set the project's averaging period and critical period of interest.

Critical Period of Interest

The critical period of interest is the time period (span of years) over which the model simulations (calibration, validation, and TMDL) will be performed. Unless otherwise justified, the critical period of interest for all lakes will represent the most recent ten years and will match the TMDL time period.

Within the recent ten-year critical period of interest, climatic outliers will be identified (e.g., very wet or dry year) by flagging years with Standardized Precipitation Index (SPI) values less than -1.5 or more than 1.5 (based on an average of recent 30 years of data from local weather stations; see full description in McKee, 1993[2]). These years will be flagged to use in interpreting the representativeness of calibration, validation, and TMDL periods, aiming for only average year conditions in all three periods with the understanding that the nutrient load reductions calculated for the TMDL from the model will likely not be protective of water quality in abnormal precipitation years when conditions are typically most limiting for lakes.

Calibration and validation time periods will depend on and correspond, as feasible, to available tributary and in-lake water quality monitoring data for the LLRM and BATHTUB model, respectively. If possible with available data, the validation period will be an independent time period. There may be instances when water quality data are limited so that only 1-2 years are used for calibration. In these cases, caution must be used in interpreting the model outputs for the simulated years and the annual weather conditions those years represent (e.g., very wet or dry year) to better estimate the possible error or data skewness (e.g., high or low) and determine whether or not the model results can be used for the TMDL. SPI values can provide a more nuanced definition of yearly precipitation conditions (e.g., moderately wet, severely dry, extremely wet, near-normal, etc.).

Averaging Period

Model analyses and inputs for both the LLRM and BATHTUB model will be generated for a specific, recurring time period within the critical period of interest, referred to as the averaging period. There are two averaging periods to consider: (1) the in-lake water quality calibration and validation in the BATHTUB model and (2) the nutrient load inputs to both the LLRM and the BATHTUB model.

The averaging period for the in-lake water quality calibration in the BATHTUB model will be set by the spring/summer or growing season (at the most April-October, but more practically May/June-September) for which the BATHTUB model is designed to simulate and for which the lake trophic levels are defined by in Section 22a-426-6 of Connecticut's Water Quality Standards. Available water quality data for the growing season will determine whether the model can be calibrated.

The averaging period for the nutrient load inputs to both the LLRM and the BATHTUB model will be set by a lake turnover ratio (which is the averaging period divided by the hydraulic residence time) of 2.0 or more (Walker, 1999; Walker, 2006).

[2] Obtain precipitation data from the nearest weather station for the prior 30 years. Calculate the average annual total precipitation and the standard deviation for the dataset. For each year in the 10-year critical period of interest, calculate the difference of total precipitation from the 30-year average and divide by the 30-year standard deviation. Match values with ranges associated with nuanced conditions defined as follows: 2.0 or more = extremely wet, 1.5 to 1.99 = severely wet, 1.0 to 1.49 = moderately wet, -0.99-0.99 = near normal, -1.0 to -1.49 = moderately dry, -1.5 to -1.99 = severely dry, -2.0 or less = extremely dry.

Achieving a lake turnover ratio of 2.0 or more ensures that the in-lake water quality during the growing season reflects the nutrient loading contributing to it (even if outside the growing season). The appropriate averaging period is typically the full year or more for reservoirs with relatively long nutrient residence times or a limited seasonal period (e.g., May-September) for reservoirs with relatively short nutrient residence times. Hydraulic residence time or flushing rate can be estimated as the total water inflow per year divided by the lake volume.

2.3.2 For LLRM Input

The following presents the analysis and/or assumptions made for significant inputs to the LLRM. The critical inputs to the LLRM are precipitation, sub-basins delineations and land use. Other potential model inputs include point source discharges (if applicable), septic systems, and waterfowl.

Precipitation

Obtain quality-controlled, land-based station Global Summary of the Month precipitation data from the NOAA National Centers for Environmental Information (NCEI) online at https://www.ncdc.noaa.gov/cdo-web/datatools/findstation. Find a station or multiple stations that have the most data (and watershed spatial) coverage for the most recent 30-years, including the critical period of interest, and download the data to .csv files for analysis. Stations will be ranked by representativeness (i.e., within the watershed and near the waterbody of interest are ranked highest). The primary station for model input will be selected from the ranked list if the station has at least 90% data completeness during the averaging period for the critical period of interest. Data gaps can be filled in with data from the next highest ranked station and so forth. If multiple stations have a near-complete dataset and the watershed is large, then it is recommended that the average of multiple stations covering the extent of the watershed be performed. The dataset will then be summed during the averaging period for each year and then averaged (mean) for all years in the critical period of interest. The resulting summary statistic (e.g., 1.25 meters) for model input represents the average annual precipitation for the averaging period during the critical period of interest.

Sub-Basins

Watershed and tributary drainage basin (sub-basin) boundaries are needed to determine both the amount of water flowing into a surface waterbody and the area of different land cover types contributing to nutrient loading. At a minimum, a given lake watershed will be broken out into sub-basins based on the major (mapped) tributary inflows to the lake plus the direct shoreline drainage. Additional sub-basins could be added if there are sufficient data for upstream tributaries (or point sources) or if an upstream nutrient impaired lake feeds into the given lake. Other notable ponds, wetlands, or river junctions could also have sub-basins delineated, depending on the project goals and available water quality data for calibration. Delineations will be based on regional, subregional, and local watershed basin areas available through CT DEEP but could also be spot-checked through StreamStats or local knowledge (refer to Table 4). Model defaults are no sub-basin routing (i.e., sub-basins flow through themselves directly to the lake and not through any other sub-basin) and some sub-basin attenuation of water and nutrients (i.e., a small amount of water and nutrients is expected to be retained within the sub-basin). Refer to Section 4.3 Model Calibration for calibration process.

Land Use

Accurately defining watershed land use types with regionally appropriate export coefficients is essential to generating reliable water and nutrient loading estimates. Unmanaged forested land, for example, tends to deliver very little nutrients downstream when it rains, while row crops and low to high density urban development export significantly more nutrients due to fertilizer use, soil erosion, car and factory exhaust, pet waste, and many other sources. Smaller amounts of nutrients are also exported to lakes and streams via groundwater under baseflow conditions. This nutrient load is delivered with groundwater directly to the lake or indirectly to tributary streams; however, much of the nutrients are retained in the soil as water infiltrates to the ground.

The LLRM includes 14 pre-defined land use categories (Table 5) with each assigned both runoff and baseflow export coefficients for precipitation, phosphorus, and nitrogen to calculate load generation. Export coefficients are applied to each land use type regardless of location in the watershed. Possible differences in export coefficients for similar land use types in different sub-basins can be accounted for in attenuation factors.

We recommend using the University of Connecticut (UConn) Center for Land Use Education & Research (CLEAR) 2015 land use data for this project (refer to Appendix A for comparison between two different land use data sources; Table 4). Convert land use codes to LLRM land use categories (Table A-1). Use GIS tools to calculate the area of each land use category by sub-basin (making sure to exclude the area of the given lake or impoundment from the direct shoreline drainage before input to the LLRM). It is expected that several watersheds will extend outside of the State of Connecticut, and thus, other state land use files will need to be used. These state land use files should be compared to the 2016 NLCD using the same approach as described in Appendix A to determine which file is more accurate for modeling purposes.

We recommend using default LLRM values for initial model inputs but consider tailoring the default LLRM values to a more Connecticut-specific reference list (AECOM, 2009; Table 5). The LLRM provides default coefficients, including an overall range of possible values, for each land use category, identified from published scientific literature that may be regional or national. These values may be adjusted during the calibration process but should not drive the calibration.

Table 5. Runoff and baseflow export coefficients for precipitation, phosphorus, and nitrogen based on minimum, median, and maximum values from published scientific literature referenced in the LLRM documentation; default values (DEF) used in the original LLRM spreadsheet; and final values (BAN) used in the Bantam Lake LLRM (CEI, Inc., 2020).

LLRM LU Classification	PRECIPITATION EXPORT (FRACTION)					PHOSPHORUS EXPORT (KG/HA/YR)					NITROGEN EXPORT (KG/HA/YR)				
	MIN	MED	MAX	DEF	BAN	MIN	MED	MAX	DEF	BAN	MIN	MED	MAX	DEF	BAN
RUNOFF															
Urban 1 (LDR)				0.30	0.30	0.19	1.10	6.23	0.65	0.55	1.48	5.50	38.47	5.50	4.95
Urban 2 (MDR/Hwy)				0.40	0.40	0.19	1.10	6.23	0.75	0.55	1.48	5.50	38.47	5.50	4.95
Urban 3 (HDR/Com)				0.60	0.60	0.19	1.10	6.23	0.80	0.55	1.48	5.50	38.47	5.50	4.95
Urban 4 (Ind)				0.50	0.50	0.19	1.10	6.23	0.70	0.55	1.48	5.50	38.47	5.50	4.95
Urban 5 (P/I/R/C)				0.10	0.10	0.19	1.10	6.23	0.80	0.55	1.48	5.50	38.47	5.50	4.95
Agric 1 (Cvr Crop)				0.15	0.15	0.10	0.80	2.90	0.80	0.40	0.97	6.08	7.82	6.08	5.47
Agric 2 (Row Crop)	0.10	0.40	0.95	0.30	0.30	0.26	2.20	18.60	1.00	1.10	2.10	9.00	79.60	9.00	8.10
Agric 3 (Grazing)				0.30	0.30	0.14	0.80	4.90	0.40	0.40	1.48	5.19	30.85	5.19	4.67
Agric 4 (Feedlot)				0.45	0.45	21.28	224.00	795.20	224.00	112.00	680.50	2923.20	7979.90	2923.20	2630.88
Forest 1 (Upland)				0.10	0.10	0.02	0.20	0.83	0.20	0.10	1.38	2.46	6.26	2.86	2.21
Forest 2 (Wetland)				0.05	0.05	0.02	0.20	0.83	0.10	0.10	1.38	2.46	6.26	2.86	2.21
Open 1 (Wetland/Lake)				0.05	0.05	0.02	0.20	0.83	0.10	0.10	1.38	2.46	6.26	2.46	2.21
Open 2 (Meadow)				0.05	0.05	0.02	0.20	0.83	0.10	0.10	1.38	2.46	6.26	2.46	2.21
Open 3 (Barren)				0.40	0.40	0.14	0.80	4.90	0.80	0.40	1.48	5.19	30.85	5.19	4.67
BASEFLOW															
Urban 1 (LDR)				0.150	0.150	0.001	0.010	0.050	0.010	0.010	1.00	5.00	20.00	5.00	5.00
Urban 2 (MDR/Hwy)				0.100	0.100	0.001	0.010	0.050	0.010	0.010	2.00	10.00	40.00	5.00	10.00
Urban 3 (HDR/Com)				0.050	0.050	0.001	0.010	0.050	0.010	0.010	4.00	20.00	80.00	5.00	20.00
Urban 4 (Ind)				0.050	0.050	0.001	0.010	0.050	0.010	0.010	1.00	5.00	20.00	5.00	5.00
Urban 5 (P/I/R/C)				0.050	0.050	0.001	0.010	0.050	0.010	0.010	1.00	5.00	20.00	5.00	5.00
Agric 1 (Cvr Crop)				0.300	0.300	0.001	0.010	0.050	0.010	0.010	0.50	2.50	10.00	2.50	2.50
Agric 2 (Row Crop)	0.01	0.20	0.40	0.300	0.300	0.001	0.010	0.050	0.010	0.010	0.50	2.50	10.00	2.50	2.50
Agric 3 (Grazing)				0.300	0.300	0.001	0.010	0.050	0.010	0.010	1.00	5.00	20.00	5.00	5.00
Agric 4 (Feedlot)				0.300	0.300	0.001	0.030	0.100	0.010	0.030	5.00	25.00	100.00	25.00	25.00
Forest 1 (Upland)				0.400	0.400	0.001	0.004	0.010	0.005	0.004	0.05	0.50	1.00	1.00	0.50
Forest 2 (Wetland)				0.400	0.400	0.001	0.004	0.010	0.005	0.004	0.05	0.50	1.00	1.00	0.50
Open 1 (Wetland/Lake)				0.400	0.400	0.001	0.004	0.010	0.005	0.004	0.05	0.50	1.00	0.50	0.50
Open 2 (Meadow)				0.300	0.300	0.001	0.004	0.010	0.005	0.004	0.05	0.50	1.00	0.50	0.50
Open 3 (Barren)				0.200	0.200	0.001	0.004	0.010	0.005	0.004	0.05	0.50	1.00	0.50	0.50

Point Sources

Any known point source discharges can be included in the LLRM. Point sources can be routed through a tributary sub-basin or the direct shoreline drainage. National Pollutant Discharge Elimination System (NPDES) permits for surface water discharges can be selected from the CT DEEP's permit files (internal version shared). Some of these permitted discharges likely have quarterly data for nutrients and flow, which can be summarized during the averaging period for the critical period of interest and input to the LLRM. Data review and aggregation for model input will follow methods described in Walker (1999). At least one year of monthly or more frequent nutrient concentration and flow data for the averaging period would be required to be considered for model input.

In addition, if an upstream nutrient impaired lake feeds into a given lake, then the upstream lake should be modeled first and the outputs from that model be fed into the LLRM as a point source to the given lake (through the appropriate sub-basin). Outputs from the model to be fed into the LLRM as a point source to the given lake include the modeled in-lake concentration and the annual volume of discharged water from the upstream lake.

Septic System Estimates

The nutrient load to groundwater baseflow is generally accounted for in urban land use export coefficients, except when the horizontal distance of a septic system to the lake is short (and thus has minimal time for retention). The following steps will be completed to generate a load estimate for nearshore septic systems:

- Determine the number of primary dwellings within <100 feet and within 100-300 feet of a given lake. UCONN/CTDEEP Environmental Conditions Online (CTECO) provides a buildings layer based on the 2012 statewide impervious surface layer. Note that the buildings layer accounts for all types of buildings, including secondary sheds, barns, and garages, and thus could inflate the estimate for primary dwellings.
- Subtract the number of primary dwellings on sewer. CT DEEP is currently working to revise the sewer service area layer. The most recent internal working version will be used for this project. Note that the defined sewer service area may not completely overlap all polygons in the buildings layer, so a brief manual check may be warranted.
    - As an exercise using the 2012 buildings layer, we counted the number of buildings (257) within 300 feet of Bantam Lake and excluded those buildings within the vicinity of the sewer service area (232). By comparison, the model for Bantam Lake estimated 149 primary dwellings around the lake based on a previous 2009 study, highlighting the fact that our count was inflated by secondary structures around the shoreline. As an alternative approach, we used the CT DEEP parcels layer (2010 Connecticut Parcels) to count the number of parcels with at least one building within 300 feet of the shoreline (to possibly help eliminate double counting of secondary structures). The parcels layer has not been updated, however, and resulted in a gross underestimate of the number of primary dwellings (55).
    - Assuming that Bantam Lake represents typical primary to secondary structure ratios for other Connecticut lakes, we could apply a 0.64 factor correction to the primary dwelling estimate from our initial approach. With this approach, we also assume that all possible primary dwellings are occupied at least part of the year, which is likely not realistic, but may help to account for possible additional effluent loading from leaking sewer infrastructure in proximity to the lake. Alternatively, if resources allow, a manual count of primary dwellings within 300 feet of the lake shore (outside of any sewer areas) can be performed using recent aerial images from ESRI and/or Google Earth.
- Sort the number of non-sewered primary dwellings by use: year-round (365 days) or seasonal (90 days). Use the US Census Bureau's American Fact Finder online search tool to locate General Housing Characteristics by town (Place) for the 2010 Census[3]. Use the seasonal, recreational, or occasional use vacancy housing unit rate out of the total housing units in one or more towns covered by the lakeshore area to derive a year-round versus seasonal housing estimate. Keep in mind that these are town-wide statistics that may underestimate the seasonal residency of the lake community. Because of the demand for lakefront property, we assume that nearly

[3] https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_GCTH2.ST13&prodType=table

all primary dwellings are occupied for some portion of the year on average (and are not completely vacant and unused).

- Determine the average number of people per primary dwelling. Unless local information is readily available for a given lake, the project will need to rely on the most recent US Census demographic information. The average number of persons per household (2014-2018) for Connecticut is estimated at 2.54[4].
- Use default LLRM values for water use per person per day (0.25 cubic meters), phosphorus concentration in effluent water (8 ppm), nitrogen concentration in effluent water (20 ppm), phosphorus attenuation factors (0.2 for dwellings within <100 feet from the lake, 0.1 for dwellings within 100-300 feet from the lake), and nitrogen attenuation factors (0.9 for dwellings within <100 feet from the lake, 0.8 for dwellings within 100-300 feet from the lake).
- Adjust the days of occupancy per year to reflect the averaging period. For example, using an averaging period of May-September, the seasonal input of 90 days would remain the same and the year-round input of 365 days would be changed to 154 days. All other factors are assumed to remain the same for the averaging period.

Waterfowl

Waterfowl can be a direct source of nutrients to lakes; however, if they are eating from the lake and their waste returns to the lake, the net change may be less than might otherwise be assumed; even so, the phosphorus excreted may be in a form that can be readily used by algae and plants. The LLRM provides default estimates of phosphorus and nitrogen loading from waterfowl on a per bird basis from the published scientific literature (0.20 kg/unit/yr for phosphorus; 0.95 kg/unit/yr for nitrogen; AECOM, 2009).

For this project, we will rely on waterfowl counts obtained from eBird online. The “hotspot map” or species map can be navigated to an area of interest for display of available bird observation sites. Clicking on these sites provides further information about the type and quantity of bird species present on a day of observation. These data are displayed in table form and could be copied to a spreadsheet with some formatting effort. However, we recommend that the entire eBird database be downloaded, cleaned, filtered, and aggregated (by taxonomy and duplicate group checklists) for the State of Connecticut. This is a moderate processing effort that could be repeated each year by CT DEEP (code available in Appendix B). From the filtered database, data can be further filtered to the area of interest (lake or impoundment) for the time period of interest (averaging period, critical period of interest) and for the species of interest (large congregating waterfowl such as ducks, geese, gulls, cormorants, herons, swans). Aggregation of data to determine the average number of waterfowl using the lake during the averaging period will depend on the frequency and distribution of available data. For example, if monthly, weekly, or more frequent data through the averaging period are available, then a monthly or weekly average count of waterfowl will be estimated for the number of birds using the waterbody per day during the averaging period. Best professional judgement and local knowledge will be relied upon for waterbodies with limited waterfowl observation data. Waterfowl counts will be made for each segment, if possible, unless otherwise justifiable (for example, if other segments are not well represented because of lack of access).

2.3.3 For BATHTUB Input

The following presents the analysis and/or assumptions made for significant inputs to the BATHTUB model. The critical inputs to the BATHTUB model are tributary loading (from the LLRM), hydrologic segmentation assumptions, morphometric features, atmospheric deposition, internal loading, evaporation, and storage.

Tributary Loading

Estimated sub-basin outflow, phosphorus loading, and nitrogen loading from the LLRM will be combined based on the assigned BATHTUB tributary (which represent direct inflows to the lake). Waterfowl and septic system loading estimates will be added to the direct shoreline drainage. Point source loading will also be added to the direct shoreline drainage if not already incorporated to sub-basin loading. Phosphorus and nitrogen concentrations for the BATHTUB tributaries will then be re-calculated. The compiled data ready for direct input to the BATHTUB model includes the following: (1) total

[4] https://www.census.gov/quickfacts/CT

area of BATHTUB tributaries; and (2) average inflow and nutrient concentration from each BATHTUB tributary (during the averaging period).

Segments

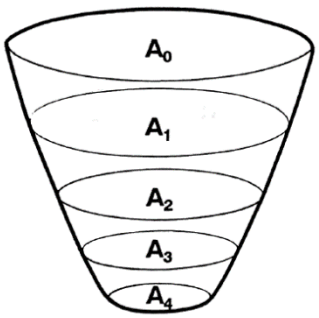
The project team will determine if the lake needs to be hydraulically segmented into distinct areas, which is best for lakes with complex morphometry (e.g., many bays or deep basins or large river impoundments that limit whole lake mixing). A spatially segmented hydraulic network can help account for advective and diffusive transport, as well as nutrient sedimentation within distinct areas. Even if there are geographically distinct areas of the lake, the project team will also consider water quality data if available for each of the distinct areas. If there are no statistically significant differences in water quality between distinct areas, then a single segment will adequately represent lake conditions in the model. A single segment assumes minimal spatial (horizontal) variation in nutrient concentrations and trophic state indicators. Water quality predictions are calculated for the entire lake (if one segment is used). If multiple segments are created for a lake, then tributary loads are routed to individual segments that are linked in a whole lake network. Each segment has its own morphometry (i.e., area, mean depth, length, mixed layer depth, hypolimnetic depth). Morphometric features reflect average conditions during the simulation period. For multiple segments, water quality predictions are made for each segment and then the area-weighted mean of all segments is calculated to represent the lake-wide average.

Morphometry

Morphometric features to be calculated for one or more segments of the lake include surface area, volume, mean depth, length, mixed layer depth, and hypolimnetic depth. The CT DEEP has available online a lake bathymetry file that includes water depth and surface area estimates for lakes. This file can be used to calculate several of these morphometric features. Any available reports can be used to cross-check these estimates for reasonableness.

- Surface area (km^2) can be obtained from the lake bathymetry file as the sum of the areas at all depths. The CT DEEP bathymetry file is set up as donut-shaped polygons for each depth, all of which need to be summed to obtain a total lake surface area estimate.
- Volume (m^3) can be calculated from the summation of the volumes of individual strata, determined by the lake surface area at successive depths. Volume (V) equation is provided in Wetzel (2001) and as follows:

$$V = \frac{h}{3}\left(A_1 + A_2 + \sqrt{A_1 A_2}\right)$$



With,

h = vertical depth of the stratum (m)
A_1 = area of the upper surface of the stratum (m^2)
A_2 = area of the lower surface of the stratum (m^2)

- Mean depth (m) can be calculated as volume divided by surface area.
- Length (km) is the average distance of major flow axes in the lake and is used to estimate the diffusive exchange rate (i.e., longitudinal dispersion).
- Mixed layer depth (m) generally represents the epilimnion and is used for chlorophyll-a computations. Review temperature profiles for the lake during summer thermal stratification. Estimate the epilimnion depth by determining the depth at which there is more than a 1 °C per meter change in temperature near the surface.
- Hypolimnetic depth (m) is used to calculate hypolimnetic and metalimnetic oxygen depletion rates in stratified lakes. Since the empirical models for predicting oxygen depletion rate were developed using data from near-dam stations, hypolimnetic depths will be specified only for near-dam (i.e., outlet) segments in the study area. As indicated in the "mixed layer depth" calculation methodology, review temperature profiles for the lake during summer thermal stratification. Estimate the hypolimnion depth by determining the depth at which there is more than a 1 °C per meter change in temperature near the bottom.

Atmospheric Deposition

The atmospheric deposition rate will be obtained from the published literature for phosphorus (0.11 kg/ha/yr, Schloss et al. 2013) and nitrogen (see map in Figure 9A in USGS, 2004, using results from Ollinger et al. 1993) and multiplied by the lake surface area (from the bathymetry file).

Internal Loading

Internal loading rates reflect nutrient recycling from bottom sediments. As indicated by the BATHTUB model documentation, rates are normally set to zero, since the pre-calibrated nutrient retention models already account for nutrient recycling that would normally occur (Walker, 1999; Walker, 2006). The BATHTUB model documentation further recommends that "*nonzero values will be specified with caution and only if independent estimates or measurements are available*" and "*in situations where monitoring data indicate relatively high internal recycling rates to the mixed layer during the growing season, a preferred approach would generally be to calibrate the phosphorus sedimentation rate (specify calibration factors< 1)".*

Based on the BATHTUB guidance cited above, internal loading will initially be entered into the model as zero. Internal loading would then be adjusted as needed during the model calibration process based on the best fit of both external and internal pollutant loads to the predicted and observed in-lake water quality concentrations. As noted in the BATHTUB guidance, any overestimation of internal load will result in an associated underestimation of external load. As a quick rule of thumb, if the internal loading (calculated as follows below) shows a rate of more than 0.5 mg/m^2/day (representing typical background), then it is likely that an internal loading estimate should be input to the model.

As a point of reference for calibration, a simple internal load calculation can be performed by comparing the difference in hypolimnion phosphorus concentration at the beginning of the season (i.e., pre-stratification, May) or conversely the end of the season (i.e., fall turnover, October) and the time of the highest observed hypolimnetic concentrations (summer). Alternatively, the average difference between epilimnion and hypolimnion phosphorus concentrations during summer thermal stratification can also be used. For both approaches, this difference is then multiplied by the estimated volume of the hypolimnion to estimate the mass of phosphorus assumed to be derived from internal loading.

The estimate generated by the first approach (using hypolimnion data in early spring or late fall when the water is not stratified) can be further adjusted to account for the fraction of total particulate phosphorus assumed to be exchanged with the epilimnion during summer stratification. The mass of phosphorus from internal loading for model input is adjusted by multiplying by the Nürnberg Retention Parameter (**R**) (i.e., fraction of sediment retained by lake) (Nurnberg, 1984), which is calculated as follows:

$$R = \frac{15}{18 + Hydraulic\ Overflow\ Rate\ (H)}$$

With,

$$H = \frac{Annual\ Average\ Discharge\ (\frac{m^3}{yr})}{Lake\ Surface\ Area\ (m^2)}$$

Evaporation

Unless lake-specific information is available, we will rely on the use of the regional monthly mean potential evaporation estimator, $E_{p,r}(\tau)$, developed by Fennessey & Vogel (1996)[5] for freshwater lakes in the northeast (see equations 2, 3a-e, and 4a-e). Data needed for the equation include longitudinal location in decimal degrees, lake surface elevation in meters, and mean monthly temperature in °C (to be obtained along with precipitation data, see Precipitation). Monthly pan evaporation data between 1950-2001 for NOAA COOP stations are also available, but there is only one station available for the State of Connecticut (Station 65445 in Norfolk, CT) (Hobbins et al. 2017).

[5] https://www.academia.edu/17826839/ESTIMATING_AVERAGE_MONTHLY_LAKE_EVAPORATION_IN_THE_NORTHEAST_UNITED_STATES

Storage

If water level data are available from a USGS gage or a locally controlled dam at the outlet of the lake or impoundment, the BATHTUB model can account for the increase in water level elevation between the start and end of the averaging period. If water level data are not available, then a "no increase" in storage can be assumed to be representative of steady-state conditions. According to BATHTUB user documentation (Walker, 1999; Walker, 2006), this value is only used for completeness in mass balance computations and does not influence predicted nutrient concentrations.

2.3.4 For Nutrient Load Reduction Analysis

This section provides a summary of the TMDL determination approach using outputs from the LLRM and BATHTUB model to estimate watershed based annual phosphorus and nitrogen load reductions needed to attain the natural trophic status of a given lake. First, the TMDL endpoints or water quality targets for the critical period of interest are determined for each lake based on its natural trophic status. Next, the BATHTUB model is used to define the relationship between nutrient loading and in-lake nutrient concentration for each lake system by performing five iterations of the model using different tributary concentration values. The nutrient loading and resulting in-lake nutrient concentration is plotted to derive a best-fit line, the equation of which can be used to determine the nutrient load required to meet the in-lake nutrient concentration or water quality target. Subsequently, the estimated reductions in watershed loading needed to meet the water quality targets can be calculated as the difference between current loading and the loading target divided by the current loading.

Water Quality Targets

TMDLs require one or more quantitative measures that can be used to assess the relationship between pollutant sources and their impact on water quality. These are often numeric water quality standards. For lake eutrophication, relevant Connecticut Water Quality Standards specify ranges of chlorophyll *a*, total phosphorus and total nitrogen concentrations, and Secchi disk transparency depth levels that are associated with different lake trophic levels (Table 6). In addition to water quality data, the trophic state of a lake is determined by the percentage of lake surface area affected by macrophytes.

The TMDL determination establishes load reductions needed to meet numeric water quality targets for the "natural" trophic state of each lake for total phosphorus, total nitrogen, chlorophyll *a*, transparency, and hypolimnetic oxygen depletion rate. The "natural" tropic status of the lake is defined in Connecticut's Water Quality Standards as the biological, chemical, and physical conditions and communities that occur within the environment which are unaffected or minimally affected by human influences. Water quality targets based on trophic state were established by Section 22a-426-6 of Connecticut's Water Quality Standards for these parameters except for the hypolimnetic oxygen depletion rate[6] (Table 6). These water quality targets are collectively used to determine the trophic status of a lake (i.e., characterization of biological productivity). Trophic status can range from high productivity ("Highly Eutrophic") to low productivity ("Oligotrophic"). Section 22a-426-6 of Connecticut's Water Quality Standards include dissolved oxygen criteria for Class AA, A, and B waterbodies. The standards indicate that dissolved oxygen must always exceed 5 mg/L.

CT DEEP will evaluate available data for each lake and set water quality targets for each lake that are intended to be conservatively protective of the lake's ecology and designated uses. It is CTDEEP's intention to develop an approach that identifies the natural trophic tendencies for lakes and thus define appropriate water quality-based goals for lake management. A defined procedure has not yet been developed, so water quality targets for this project will be set by CT DEEP based on professional judgement. As the process for determining appropriate water quality targets for lakes in Connecticut matures, the water quality targets for these lakes may be revisited.

[6] The hypolimnetic oxygen depletion rate can be estimated and documented during the BATHTUB modeling process and can be used as an additional parameter to qualitatively evaluate lake conditions. However, a specific water quality target for hypolimnetic oxygen depletion rate will not be evaluated for the following reasons:

1) Section 22a-426-6 of Connecticut's Water Quality Standards do not indicate a specific range or rate for hypolimnetic oxygen depletion.
2) Similar lakes with nearly identical hypolimnetic oxygen depletion rates can have very different in-lake dissolved oxygen concentrations due to natural conditions such as lake bathymetry. To set a reasonable "reference" hypolimnetic oxygen depletion rate for a lake, an empirical characterization of natural background levels of depletion would be required for various lake types and classifications across the state.

Table 6. Parameters and defining ranges for the trophic state of lakes in Connecticut. Adapted from the State of Connecticut Department of Energy and Environmental Protection Water Quality Standards 2013 (Sec. 22a-426-6)[1].

Trophic State	Description	Parameters	Defining Range
Oligotrophic	Low in plant nutrients. Low biological productivity characterized by the absence of macrophyte beds. High potential for water contact recreation.	Total Phosphorus Total Nitrogen Chlorophyll-a Secchi Disk Transparency	0-10 µg/l spring and summer 0-200 µg/l spring and summer 0-2 µg/l mid-summer 6 + meters mid-summer
Mesotrophic	Moderately enriched with plant nutrients. Moderate biological productivity characterized by intermittent blooms of algae and/or small areas of macrophyte beds. Good potential for water contact recreation.	Total Phosphorus Total Nitrogen Chlorophyll-a Secchi Disk Transparency	10-30 µg/l spring and summer 200-600 µg/l spring and summer 2-15 µg/l mid-summer 2-6 meters mid-summer
Eutrophic	Highly enriched with plant nutrients. High biological productivity characterized by occasional blooms of algae and/or extensive areas of dense macrophyte beds. Water contact recreation opportunities may be limited.	Total Phosphorus Total Nitrogen Chlorophyll-a Secchi Disk Transparency	30-50 µg/l spring and summer 600-1000 µg/l spring and summer 15-30- µg/l mid-summer 1-2 meters mid-summer
Highly Eutrophic	Excessive enrichment with plant nutrients. High biological productivity characterized by severe blooms of algae and/or extensive areas of dense macrophyte beds. Water contact recreation may be extremely limited.	Total Phosphorus Total Nitrogen Chlorophyll-a Secchi Disk Transparency	50 + µg/l spring and summer 1000 + µg/l spring and summer 30 + µg/l mid-summer 0-1 meters mid-summer

[1] *Standards also include additional indicator for aquatic macrophyte distribution and abundance (Eutrophic: extensive and dense growth 75-100% of water body area; Mesotrophic: extensive and dense growth 30-75% of water body area when water column indicators are Oligotrophic.*

Load Reduction Calculation

Run BATHTUB iteratively for five hypothetical loading scenarios using different tributary concentration values for both phosphorus and nitrogen at similar magnitude changes (suggest 20% change increments above and below current nutrient loading), though additional simulations adjusting one nutrient more than the other may be warranted in some circumstances. Hypothetical loading scenarios can be created by sequentially adjusting the tributary input concentrations for total phosphorus and total nitrogen at selected intervals to enable visualization and analysis of a wide range of potential conditions.

Multiply tributary nutrient concentrations by flow rates to calculate nutrient loading rates. Plot these loading rates against the simulated in-lake nutrient concentrations. Apply a best fit trend line to define the relationship and derive the lake's loading capacity or the maximum nutrient loads allowed to meet the water quality targets. The load reductions needed can then be calculated as the difference between estimated current loading and the modeled loading target. Chl-a will dictate the nutrient load reduction in order to meet the Chl-a water quality target.

CT DEEP may determine that additional model scenarios are warranted to aid in the TMDL development process. Additional model scenarios will follow the general guidelines and procedures detailed in this QAPP. Notes and justifications for changes made to the model to run a scenario will be documented fully.

2.4 Limitation on the Use of Existing Data

The limitations on and quality control needed for the use of existing data are summarized in Table 4. Major limitations and quality control actions are described in greater detail below. Acknowledging and understanding model limitations is critical to interpreting model results and applying any derived conclusions to management decisions.

Generally, the models will represent a static snapshot in time based on the best information available at the time of model execution. Factors that influence water quality are dynamic and constantly evolving; thus, the models will be regularly updated when significant changes occur within the watershed and as new water quality and physical data are collected. In this respect, the models reflect the best available information or knowledge at the time of release. Model results represent annual averages and are best used for planning level purposes and will only be used (such as to set regulatory limits) with full recognition of the model limitations and assumptions.

Water quality data will be compiled into a common database using MS Excel 2016. Validated data from federal and state quality-controlled databases will be selected for analysis. All validated data obtained from publicly available databases will likely meet data quality objectives and criteria for the project, as outlined below.

- Samples must be collected by trained personnel under an approved QAPP or similar document to meet representative data quality criteria.
- Samples must be analyzed in accordance with approved laboratory methods to meet similar precision, accuracy, and comparability data quality criteria.

Since validated data obtained from publicly available federal and state databases will likely meet quality acceptance criteria for field sampling and certified laboratory methods, any data exclusions will likely be due to reporting limit incompatibilities as a result of differing laboratory methods, dilution, or matrix interferences.

Land cover types are based on coarse resolution data that are not ground truthed. Literature values and best professional judgement will be used in evaluating and selecting appropriate land cover export coefficients for the watershed. While these coefficients may be accurate on a larger scale, they are likely not representative on a site-by-site basis.

Sub-basin delineations are based on CT DEEP data and topographic review, with no ground-truthing. Errors in sub-basin delineations are more likely in highly urban areas with significant stormwater networks.

Septic systems are based on coarse estimates. Default literature values for daily water usage per person, phosphorus concentration output per person, and system phosphorus attenuation factors are used and may not reflect local watershed conditions. The true count and functioning of individual septic systems in the watershed would likely be unknown. The sewer service area file (refer to Table 4) is also considered an Internal working draft that is not yet complete and may not cover the area completely so manual review for reasonableness may be necessary.

Waterfowl counts are based on coarse estimates from web-based resources.

3.0 ASSESSMENTS AND OVERSIGHT

3.1 Project Oversight

The hierarchy of reporting within the project team, along with the project roles of key personnel, are identified in Section 1.4. The Project Lead has primary responsibility for monitoring the activities of this project, identifying or confirming any quality problems, and ensuring that the quality requirements specified in this generic QAPP are followed. Significant problems will be brought to the attention of the CT DEEP and US EPA Region 1 lead staff, who will initiate corrective actions described above, document the nature of the problem, and ensure that the recommended corrective action is carried out. Data collection and assessment by Project Support will be reviewed weekly (or as needed) prior to analyses by the Project QA Officer. The Project QA Officer will consider the data's usability, quality, and consistency with other data sources and document any data limitations that do not meet data acceptance criteria.

3.2 Project Documentation

Data quality review procedures will be documented in the draft and final project products (including metadata associated with the tabular and spatial databases), which will be reviewed by the Project QA Officer and then reviewed by the Project Leader before sending to the CT DEEP and US EPA Region 1 lead staff for final review. Any feedback will be incorporated by the Project Support to the final project products. The project products are identified in Section 1.6 and Section 5. Project files will be stored on the hard drives of individual personnel (including all files by the Project Leader) and will be backed up to an external hard drive daily; alternatively, project files will be stored on CT DEEP's in-house client system.

3.3 Corrective Actions

If quality problems that require attention are identified, the Project Lead will determine whether attaining acceptable quality requires either short- or long-term corrective actions. 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 that cannot be resolved in this manner require more formalized, long-term corrective action.

Possible data problems that would require corrective actions include lack of appropriate or complete metadata, incomplete datasets, and data that conflict with other quality-assured data sources. Corrective actions could include:

- Contacting the data originator for more complete metadata or explanation.
- Researching alternative data sources or formats that could be added to the incomplete dataset.
- Filtering out or discarding highly conflicting data with justification based on the metadata.

In some cases, acceptance criteria may need to be lessened or altered to accommodate problematic data that is necessary for the project but are the only data source available. Any data limitations will be documented in project products. To avoid data loss or file corruption, working copies of each dataset will be created so that originals remain intact and the copies stored on primary and back-up hard drives or the CT DEEP's in-house client system.

4.0 MODEL & DATA VERIFICATION, VALIDATION, & EVALUATION

4.1 Data Verification and Validation

Data verification and validation processes provide a method for determining the usability and limitations of data and provide a standardized data quality assessment. Accurate and complete metadata are needed to ensure that the data source and collection/analysis methods are adequately defined and meet data quality objectives for comparability and representativeness.

All secondary data will be reviewed to assess whether the data meet data quality acceptance criteria for use in analyses. The review process will include thorough metadata review, documentation, and investigation (as necessary). The methods and reporting limits (if applicable) of data will be reviewed and validated for use in analyses if the data meet the data quality objectives and criteria set in Section 1.7. Any data not meeting the data quality criteria will be excluded from analyses or properly justified for use if certain approaches are appropriate.

Secondary data such as spatial files and written documents will be selected for use based on relevance, completeness, accuracy, quality, and age of data (i.e., most recent available source that meet criteria). Data may be rejected for use if metadata are incomplete, data are outdated or incomplete, or data are redundant. Low quality data will not be used for analysis (except possibly as a supporting reference) unless it is the only available data; justification for use of and limitations to the low-quality data will be noted in project products. Any files with draft indication will be followed-up with the originator for the final version, if available.

Metadata for all secondary data will include a data description, originator, source of access, publication date, time period and/or specific time and date collection information (for sampling data), and spatial domain information (such as projection/coordinate systems used; see Table 4). Additional metadata for sampling data sets will include the following: sampling and analysis plan, laboratory method, reporting limit, reporting units, field qualifiers or notes (e.g., missing values), and laboratory qualifiers.

4.2 Data Evaluation

Data evaluations will be made to ensure that QC is maintained throughout project. QC evaluations will include reviewing model setup and double-checking work, and other review to ensure that the standards set forth in the QAPP are met or exceeded. Raw (original) data will be entered into a standard database. All entries will be compared to the original data files to ensure no transcription errors. A screening process will be used to scan through the database and flag data that are outside typical ranges for a given parameter (outliers defined as more than 1.5 times the interquartile range). Values outside typical ranges will not be used to develop model calibration data sets or model kinetic parameters, unless otherwise justified. All data will be transformed to a common measurable unit by parameter. Duplicate data entries will be averaged. Some data may be manipulated using Microsoft Excel or R x64 3.5.1 / RStudio. Ten percent of the calculations will be recalculated to ensure that correct formula commands were entered into the program. If any of the data calculations are incorrect, all calculations will be rechecked after the correction is made to the database.

4.3 Model Parameterization (Calibration)

Calibration is the process by which model results are brought into agreement with observed data and is an essential part of environmental modeling. Calibration focuses on the parameters with the greatest uncertainty. Changes are made within a plausible range of values, and an effort is made to find a realistic explanation among environmental conditions for these changes. If the observed data are insufficient for calibration, default values and best professional judgement will be used, and any existing data points will serve only as guideposts in the calibration process. Observed nutrient concentrations will be given primacy during the calibration process, such that the ability of the models to accurately simulate nutrient concentrations will be used as a leading indicator of acceptable model performance (refer to Section 1.7.2).

Models are often calibrated through a subjective trial-and-error adjustment of model parameters because many interrelated factors influence model output. The model calibration goodness-of-fit measures should include both qualitative metrics and quantitative methods. Qualitative measures of calibration progress are commonly based on plots depicting observed and predicted data, while quantitative measures of calibration progress are based on error statistics, correlation test, or cumulative distribution tests. The models will be considered calibrated when they reproduce data within an acceptable level of accuracy (see **Error! Reference source not found.**). Calibration and validation activities and procedures, along with goodness-of-fit validation targets for specific parameters, will be documented in the final report.

The LLRM and BATHTUB model will be calibrated to the best available data, including literature values and interpolated or extrapolated existing field data. If multiple datasets are available, an appropriate period and corresponding dataset will be chosen based on factors characterizing the dataset, such as corresponding weather conditions, amount of data, and temporal and spatial variability of data. LLRM outputs will first be calibrated (as feasible) based on available data then formatted for input into BATHTUB. The following sub-sections discuss specific calibration approaches for the LLRM and BATHTUB model.

LLRM – Routing

Nested sub-basins or sub-basins that feed into another sub-basin before outflowing to the lake must be directed as such in the model's routing table. Attenuation factors will be applied to the downstream-most sub-basin to avoid the additive effect of attenuation factors through more than one sub-basin (since attenuation in a downstream sub-basin can affect inputs from an upstream sub-basin that flows through the downstream sub-basin). Attenuation factors can be applied to each sub-basin during the calibration process if adequate water quality data are available and possibly adjust the final attenuation factors that are only applied to the downstream-most sub-basins.

LLRM – Attenuation

Water can be lost through evapotranspiration, deep groundwater, and wetlands, while nutrients can be removed by infiltration, sedimentation, or uptake processes. Larger water and nutrient losses can be expected with lower gradient or wetland-dominated sub-basins. Additional infiltration, filtration, detention, and uptake of water and nutrients will decrease water and nutrient attenuation values[7], such as for sub-basins dominated by moderate/small ponds or wetlands or channel processes that favor uptake, depending on the grade. Headwater systems can be assumed to have a greater attenuation than the mainstem rivers since the flow of water is lower relative to the mainstem, giving more opportunity for infiltration, adsorption, and uptake.

It is important to note that attenuation processes for nitrogen and phosphorus can be different, and we will expect lower losses for nitrogen compared to phosphorus. Nitrogen moves more readily through soil, and while transformations occur in the stream, losses due to denitrification require slower flows and low oxygen levels not commonly encountered in steeper, rockier channels (AECOM, 2009). Losses from uptake and possibly denitrification are more likely in wetland areas.

[7] Attenuation values represent the fraction of water and nutrients passing through a sub-basin, ranging from 0 to 1 with 0 representing complete attenuation within the sub-basin and 1 representing no attenuation within the sub-basin.

We can generally expect at least a 5% loss (95% passed through, default) in water, a 10% loss (90% passed through, default) in phosphorus, and a 5% loss (95% passed through, default) in nitrogen for each sub-basin (Table 7). However, if a sub-basin has steep slopes and/or a short channel length to the lake, then it is reasonable to adjust the attenuation value up to 1.0.

Table 7. Attenuation values for water, phosphorus, and nitrogen based on sub-basin characteristics. Gray shading indicates model default values. These attenuation values represent starting points for the calibration process but can be adjusted further based on other sub-basin characteristics such as slope grade.

Water	Phosphorus	Nitrogen	Description
1.00	1.00	1.00	Steep slopes and/or short channels adjacent to lake
0.95	0.90	0.95	Default values reflecting some attenuation
0.95	0.85	0.90	Channel processes that favor uptake (i.e., long channel length)
0.90	0.80	0.85	Small-sized ponds or small wetlands with low-lying areas
0.85	0.75	0.80	Moderate-sized ponds or moderate wetlands with low-lying areas
0.80	0.70	0.75	Large-sized ponds or large wetlands with significant low-lying areas

LLRM – Standard Water Yield & Observed Flow Data

The LLRM uses a standard water yield in cubic feet per second per square mile (cfsm) as one possible check on flow values derived from water export from the watershed. The LLRM provides a range of default values for the northeast region (1.5-2.0 cfsm). Watersheds that generate more runoff per square mile of land area will have higher standard water yields, such as watersheds with steep slope topography or significant urban development.

As an initial estimate for model input, calculate the standard water yield from the nearest quality-controlled USGS stream gage station. Obtain quality-controlled monthly mean stream gage data from the USGS Current Water Data for Connecticut online at https://waterdata.usgs.gov/ct/nwis/rt[8]. Average the mean of monthly mean discharge for the months in the averaging period and divide by the drainage area to obtain the standard water yield. If there are multiple stations in a watershed, then calculate the standard water yield for each station and then average all stations (if reasonably justified as there may be stations more representative than others). If there appear to be significant discrepancies in predicted compared to observed water yield, we recommend investigating the potential for any permitted water diversions or withdrawals from the system.

The LLRM uses the standard water yield to generate a reality check estimate of flow output per sub-basin, which is compared to the modeled flow output derived from land use export coefficients. The ratio between the two estimates will be at or around 1.0.

A second check on flow values can be made using observed flow data from stream gage data in the watershed, available through the USGS (website link above). The average of the mean of monthly mean discharge calculated above in cubic feet per second can be converted to cubic feet per year (in the averaging period) as a total volume output estimate for the tributary sub-basin. If no flow data are available for the watershed, then we will rely on the default standard water yield as a check.

LLRM – Land Use Export Coefficients

It is not recommended that the land use export coefficients be adjusted to match observed data during the calibration process. The most appropriate export coefficients will be selected from published literature. Tributary sub-basin outputs will be checked against observed data, if available. If significant deviations between modeled and observed data exist for

[8] Under "Predefined displays," select Daily streamflow, which will direct to a new page. Click "show sites on a map." Zoom to the area of interest and select a station in or near the target watershed. Click Access Data from the pop-up Site Information box. From the drop-down menu, select Time-series: Monthly statistics. Check the box for Discharge. Under Choose Output Format, input the date range (YYYY-MM) to reflect the critical period of interest. Select Table of monthly mean and then Submit. The table depicts monthly mean discharge for the years specified. The bottom of the table calculates the mean of monthly mean discharge for all years in the table.

multiple sub-basins (and there are enough observed data to support it), then land use and associated export coefficients may be re-evaluated. Any adjustments to the land use export coefficients will be cited and justified.

LLRM – Observed Tributary Data

Observed tributary data can be used for calibration of tributary sub-basin nutrient load outputs. Model inputs and assumptions will be carefully reviewed when comparing observed versus predicted nutrient loads and any adjustments documented and justified. If at least one year (preferably three years) of monthly or more frequent nutrient concentration data are available for the averaging period, then the observed data can be used for model calibration. Nutrient concentrations are paired with at-collection flow measurements, if available, otherwise mean daily flow measurements collected from a nearby benchmark station with a continuous flow record and similar watershed (with flow adjusted for drainage area ratio). First, the log-transformed concentration-discharge relationship should be reviewed to determine nutrient response to flow regime and thus the appropriate method of calculating a load estimate (refer to page 2-7 in Walker, 1999). For example, an often-used method of calculating a load estimate is based on a flow-weighted mean concentration and average daily flow during the period of interest. If minimal observed data are available for a tributary sub-basin, then we can use the data as a reasonability check but not for calibration.

BATHTUB – Observed Lake Data

Observed lake data can be used in the model for calibration of in-lake water quality outputs. Model inputs and assumptions will be carefully reviewed when comparing observed versus predicted nutrient concentrations and any adjustments documented and justified. To properly calibrate the model to lake data, a minimum of one year (preferably three years) with a minimum of four temporally-representative nutrient samples (preferably monthly) during the growing season (May/June-September) is required for each segment; otherwise, the observed data will only be used as a guide and not a calibration point. Lake data will be subset for the upper mixed layer (i.e., the epilimnion or the depth at which there is more than a 1°C change in water temperature over one meter) for the growing season over the critical period of interest. If multiple stations exist within a segment, then professional judgement will be used on the spatial representativeness and data availability of those stations and whether a station should be used in the average. If multiple stations are used for one or more segments, then aggregation procedures will follow Walker (1999) on page 3-5. Data aggregation will be based on taking the median value.

BATHTUB – Water Balances

The segment inflows and outflows will be checked after measured flows for all inflow and outflow streams are specified in BATHTUB. Inflow, outflow, and increase-in-storage values can be adjusted until water balances are established. It is recommended to adjust only the values that are most likely causing the water balance error depending on knowledge of watershed characteristics and flow monitoring networks.

BATHTUB – Nutrient Turnover

Nutrient turnover ratios will be checked to be sure that the appropriate averaging period is selected based on recommendations by Walker (1999) and described in Section 2.3.1. There may be instances when the averaging period extends beyond a year, depending on the flushing rate of the lake system.

BATHTUB – Diffusive Transport

The diffusive transport terms should be checked and as needed, calibrated. When the numeric dispersion is greater than the estimated dispersion, the segmentation scheme can be revised until this condition is met. It is recommended to increase the segment numbers to decrease the segment lengths (unless predicated nutrient profiles to alternative segmentation schemes is shown to be minimal). Biologically conservative tracer data (such as chloride or conductivity) can be used to calibrate the diffusive transport terms where there is an error in more than one segment. To calibrate the diffusive transport terms, the calibration factor for dispersion needs to be adjusted to match observed tracer data. A less conservative method is to also adjust the segment calibration factors. BATHTUB has a sensitivity procedure to test the sensitivity of predicted tracer concentrations to variations in dispersion rates. When tracer data are not available, dispersion rates can be calibrated to match observed nutrient gradients.

BATHTUB – Nutrient Balances & Model Selection

BATHTUB includes multiple model options that can be selected to generate in-lake water quality predictions. For example, total phosphorus can be predicted using a second-order available phosphorus model (default, best for either natural lake system or impoundment), or a more simplistic model such as the Vollenweider Equation (best for natural lake systems and not impoundments).

Available models for each in-lake water quality variable will be evaluated and selected based on the best goodness of fit to observed water quality monitoring data. BATHTUB model output provides statistical comparisons of predicted and observed nutrient concentrations, including:

- Type (1): observed error only
- Type (2): error typical of model development data set
- Type (3): observed and predicted error

Type (2) and Type (3) are evaluated for model applicability. Type (1) can be used to determine if calibration is necessary once a sedimentation model is selected. When the absolute value of Type (1) exceeds 2.0, it is recommended to calibrate the model to match the predicted and observed concentrations. Refer to Walker (1999) for more information about two calibration methods specifically for phosphorus and nitrogen.

Internal loads, such as nutrient release from bottom sediments and fixation of atmospheric nitrogen, may be specified if the nutrient retention coefficients for phosphorus or nitrogen are negative. However, it is recommended that independent evidence and estimates are obtained before these internal loads are input to the model.

BATHTUB provides an application (Calibration Factors) to calibrate the model to account for site-specific conditions by modifying the reservoir response predictions: nutrient sedimentation rates or concentrations, chlorophyll *a* concentrations, Secchi depths, longitudinal dispersion rates, and oxygen depletion rates. The calibration factors can either be applied globally to all segments and/or individually to each segment. BATHTUB output includes statistical tests to help assess if the calibration factors used are appropriate. Calibration Factors should only be used for this project if sufficient observational data are available, and initial model runs suggest a systematic bias resulting from model error (see page 4-44 of Walker, 1999 for guidelines on when the use of calibration factors is appropriate and justified). If there are inadequate data for use in BATHTUB calibration, then the lake model will be put on hold until further data are collected. The model will be considered calibrated when it reproduces data within an acceptable level of performance (refer to Section 1.7.2).

BATHTUB - Chlorophyll *a* and Secchi

After nutrient balances have been established, the eutrophication responses are tested and calibrated as needed to select the most appropriate responses. Refer to Walker (1999) for more information about calibration methods for chlorophyll *a* and Secchi.

4.4 Model Corroboration (Validation and Simulation)

Model validation is an evaluation of the model goodness-of-fit using a dataset that is independent of that used for calibration. Validation is an important step in the process to guard against model over-fitting. The model will be considered validated if its accuracy and predictive capability have been proven to be within acceptable limits of error relative to the performance of the model during the calibration period. Since it is likely that tributary data will be minimal for project lakes, the LLRM will not be validated except to update tributary inflow loads for input to the BATHTUB model.

Model validation will be performed using a dataset that is independent of the calibration dataset and compared to model performance and acceptance criteria defined in Section 1.7.2. The LLRM and BATHTUB model will be independently run with inputs adjusted for the validation period only (i.e., precipitation, observed data) and all other inputs and calibration factors left the same.

4.5 Reconciliation with User Requirements

The value of the information generated by this project will be determined by evaluating data quality and by comparing methods and results with published data and scientific literature and the data quality objectives identified in this QAPP. Confidence in model predictions can be limited by several factors including representativeness of calibration data, knowledge of actual nutrient inputs (external loading and internal loading), and the inherent ability of the model to simulate the conditions in the lake. Data quality indicators will be calculated during model setup, calibration, and validation. Measurement quality requirements will be compared with the data quality objectives to confirm that the correct type, quality, and quantity of data are being used for the model setup and calibration. Computation and post-simulation analysis results will be reviewed for reasonableness. To ensure reproducibility of the work, the final report will identify sources of data, assumptions made during model setup, and calculations performed as part of input data pre- and post-processing.

5.0 PROJECT REPORTING

Final project reports will provide a complete and clear summary of the modeling methodology and all data and assumptions used in the LLRM and BATHTUB model such that the analysis can be easily reproduced. All model files, as well as all sources of data used in or generated by the project will be either provided as attachments to final project reports or made available upon request. Draft and final project products will be generated in common or publicly-available programs that are compatible with end user systems for ease of maintenance or updates in the future such as MS Office (e.g., Word for written reports, Excel for spreadsheets, CSVs for analysis), ArcMap Desktop (e.g., geodatabase of spatial files and/or map packages of project maps), and R / RStudio (e.g., R scripts or markdowns for statistical analyses, calculations, and data visualization). All MS Excel spreadsheets and/or model files will include metadata on data sources, corrections, and exclusions (by whom and on what date). All R scripts or markdowns will be annotated to ensure that the code for analysis can be easily reproduced and understood. All MS Word reports (as applicable) will document QA/QC procedures either within the report or as an attachment.

6.0 REFERENCES

AECOM (2009). LLRM Lake Loading Response Model Users Guide and Quality Assurance Project Plan. AECOM, Willington, CT. https://github.com/MattAtMassDEP/LLRM_model

CEI, Inc. & HydroAnalysis, Inc. (2018). QAPP for Bantam Lake Nutrient TMDL Model. Prepared by Comprehensive Environmental, Inc. and HydroAnalysis, Inc, November 28, 2018.

CEI, Inc. (2019). Bantam Lake Nutrient TMDL Model Modeling Methodology (FINAL). Prepared by Comprehensive Environmental, Inc for the United States Environmental Protection Agency Region 1 – New England, October 2019.

CEI, Inc. (2020). Bantam Lake Nutrient TMDL Model Modeling Report (FINAL). Prepared by Comprehensive Environmental, Inc. for the US Environmental Protection Agency Region 1 – New England, February 2020.

Donigian, A.S. Jr. (2002). Watershed Model Calibration and Validation: The HSPF Experience. WEF National TMDL Science and Policy 2002, November 13-16, 2002. Phoenix, AZ. WEF Specialty Conference Proceedings on CD-ROM.

EPA. (2002). Guidance for Quality Assurance Project Plans for Modeling (EPA QA/G-5M). EPA/240/R-02/007. U.S. Environmental Protection Agency. December 2002.

EPA. (2009). EPA New England Quality Assurance Project Plan Guidance for Environmental Projects Using Only Existing (Secondary) Data. U.S. Environmental Protection Agency New England. October 2009.

EPA. (2018). Region 1 - New England Environmental Data Review Program Guidance. U.S. Environmental Protection Agency Region 1 – New England. June 2018.

Fennessey, N.M., and Vogel, R.M. (1996). Regional models of potential evaporation and reference evapotranspiration for the northeast USA. Journal of Hydrology, 184 (3-4), 337-354. https://doi.org/10.1016/0022-1694(95)02980-X

Hobbins, M.T., Barsugli, J.J., Dewes, C.F., and Rangwala, I., 2017, Monthly Pan Evaporation Data across the Continental United States between 1950-2001: https://doi.org/10.21429/C9MW25.

Northeast Regional Climate Center (NRCC) (2019). Potential Evapotranspiration Data. http://www.nrcc.cornell.edu/wxstation/pet/pet.html

Nürnberg, G. K. (1984). The prediction of internal phosphorus load in lakes with anoxic hypolimnia. *Limnol. Oceanogr.,* 29: 111–124.

Ollinger, S.V., Aber, J.D., Lovett, G.M., Millham, S.E., Lathrop, R.G., and Ells, J.M. (1993). A spatial model of atmospheric deposition for the northeastern United States: *Ecological Applications*, v. 3, no. 3, p. 459-472.

Schloss, Jeffrey A. and Robert C. Craycraft. (2013). Final Report: Newfound Lake Water Quality Modeling: September 2013. UNH Center for Freshwater Biology, UNH Cooperative Extension and UNH Center for Freshwater Biology, University of New Hampshire. Durham, NH. CFB Report # 2013-09-DES-02.

State of Connecticut, Department of Energy and Environmental Protection, Water Quality Standards Regulations. Adopted October 10, 2013 (Sec. 22a-426-6 Lake trophic categories). Online at: https://eregulations.ct.gov/eRegsPortal/Browse/RCSA/Title_22aSubtitle_22a-426/

USGS (2004). Estimation of Total Nitrogen and Phosphorus in New England Streams Using Spatially Referenced Regression Models. USGS Scientific Investigations Report 2004-5012.

Walker, W.W. (1999). Simplified Procedures for Eutrophication Assessment & Prediction: User Manual. US Army Corps of Engineers Water Operations Technical Support Program, Instruction Report W-96-2, September 1996 (Updated April 1999). Online at: http://wwwalker.net/bathtub/Flux_Profile_Bathtub_DOS_1999.pdf

Walker, W.W. (2006). Simplified Procedures for Eutrophication Assessment & Prediction: BATHTUB Online user documentation, http://www.wwwalker.net/bathtub/help/bathtubWebMain.html

Wetzel, R. (2001). Limnology: Lake and River Ecosystems, Third Edition. Academic Press.

7.0 APPENDIX A: Land Use Data Source Comparison

Land use data are available through UCONN CLEAR for 2015 and through the National Land Cover Database (NLCD) for 2016. The 2015 UCONN CLEAR land use file provides satellite-derived state coverage at 30 m resolution for 12 land use classes, while the 2016 NLCD land use file provides nationwide data coverage at 30 m resolution for 15 land use classes.

As an exercise, we selected the portion of the Lake Zoar watershed in the state of Connecticut as our area of interest, clipping all relevant files to this extent. We matched land use codes for the 2015 UCONN CLEAR and 2016 NLCD land use files to the LLRM land use classifications (Table A-1) and tabulated total areas by land use for the 2015 UCONN CLEAR and 2016 NLCD files. The 2016 NLCD file had more specific types of developed land use compared to the 2015 UCONN CLEAR file.

Comparison between the two land use data sources showed that the 2015 UCONN CLEAR file better captured developed areas (especially roads) and agricultural fields (Figures A-1, A-2), which increased the estimated total phosphorus load by about 10% compared to the 2016 NLCD file (Table A-2). Thus, we recommend using the 2015 UCONN CLEAR land use file for use in this project.

We also further investigated whether additional updates to the 2015 UCONN CLEAR file could enhance the accuracy of land use coverage. We used the National Wetland Inventory for CT to add in any missing forested, emergent, or open wetlands (Open 1, Forest 2); we used the USGS NHDWaterbody file to add in any missing lakes/ponds (Open 1); we used the CT DEEP GIS WATERBODY_LINE file (with 15 ft buffer) to add in a stream network (Open 1); we used the CT DEEP GIS Road Master file (with 25 ft buffer) to add in any missing roads and distinguished between paved (Urban 2) and unpaved (Open 3) roads; we used the UCONN CLEAR 2012 buildings layer to add in any missing buildings (Urban 1); and we used the UCONN CLEAR 2012 other impervious layer (which doesn't include roads) to add in any missing impervious area (Urban 2). Comparison between the two land use data files showed that the 2015 UCONN CLEAR updated file better accounted for Urban 1 in lieu of Urban 5, as well as Forest 2 and Open 1 land use categories (Figure A-3), which increased the estimated total phosphorus load by about 2% compared to the original 2015 UCONN CLEAR file (Table A-2). Thus, we recommend that the original 2015 UCONN CLEAR land use file is sufficient for this project. It will be important to note during the model calibration process, however, of the possible underestimate in nutrient load from land use.

Table A-1. Land use (LU) codes and descriptions for the 2015 UCONN CLEAR file and the 2016 National Land Cover Database (NLCD) file matched to the LLRM land use classifications.

LLRM LU Classification	LLRM Description	UCONN CLEAR LU Gridcode	UCONN CLEAR LU Description	NLCD LU Gridcode	NLCD LU Notes
Urban 1 (LDR)	Low density residential (>1 ac lots)			22	Developed, Low Intensity
Urban 2 (MDR/Hwy)	Medium density residential (0.3-0.9 ac lots) + highway corridors	1	Developed	23	Developed, Medium Intensity
Urban 3 (HDR/Com)	High density residential (<0.3 ac lots) + commercial			24	Developed, High Intensity
Urban 4 (Ind)	Industrial				
Urban 5 (P/I/R/C)	Park, Institutional, Recreational or Cemetery	2	Turf & Grass	21	Developed, Open Space
Agric 1 (Cvr Crop)	Agricultural with cover crops (minimal bare soil)				
Agric 2 (Row Crop)	Agricultural with row crops (some bare soil)			82	Cultivated Crops
Agric 3 (Grazing)	Agricultural pasture with livestock	4	Agricultural Fields	81	Pasture/Hay
Agric 4 (Feedlot)	Concentrated livestock holding area				
Forest 1 (Upland)	Land with tree canopy over upland soils and vegetation	5, 6	Deciduous, Coniferous Forest	41, 42, 43	Deciduous, Evergreen, Mixed Forest
Forest 2 (Wetland)	Land with tree canopy over wetland soils and vegetation	9	Forested Wetland	90	Woody Wetlands
Open 1 (Wetland/Lake)	Open wetland or lake area (no substantial canopy)	7, 8, 10	Water, Non-Forested Wetland, Tidal Wetland	11, 95	Open Water, Emergent Herbaceous Wetlands
Open 2 (Meadow)	Open meadow area (not clearly wetland, but no canopy)	3, 12	Other Grasses, Utility Corridors	52, 71	Shrub/Scrub, Grasslands/Herbaceous
Open 3 (Barren)	Mining or construction areas, largely bare soils	11	Barren Land	31	Barren Land

Table A-2. Land area (hectares, ha) and phosphorus (P) runoff estimates by LLRM land use classification for the 2016 National Land Cover Database (NLCD) file and the 2015 UCONN CLEAR file (original and updated). See Appendix A text for explanation of updates.

		Land Area (ha)			P Runoff (kg/yr)		
LLRM LU Classification	Default P Runoff Export Coeff. (kg/ha/yr)	2016 NLCD	2015 UCONN	2015 UCONN Updated	2016 NLCD	2015 UCONN	2015 UCONN Updated
Urban 1 (LDR)	0.65	7,244	0	2,510	4,708	0	1,632
Urban 2 (MDR/Hwy)	0.75	3,550	25,575	28,212	2,663	19,182	21,159
Urban 3 (HDR/Com)	0.80	985	0	0	788	0	0
Urban 4 (Ind)	0.70	0	0	0	0	0	0
Urban 5 (P/I/R/C)	0.80	15,552	11,609	9,790	12,441	9,288	7,832
Agric 1 (Cvr Crop)	0.80	0	0	0	0	0	0
Agric 2 (Row Crop)	1.00	2,144	0	0	2,144	0	0
Agric 3 (Grazing)	0.40	13,377	23,641	21,947	5,351	9,457	8,779
Agric 4 (Feedlot)	224.00	0	0	0	0	0	0
Forest 1 (Upland)	0.20	146,175	133,442	123,342	29,235	26,688	24,668
Forest 2 (Wetland)	0.10	14,354	6,445	9,590	1,435	645	959
Open 1 (Wetland/Lake)	0.10	7,877	9,329	14,174	788	933	1,417
Open 2 (Meadow)	0.10	2,262	3,233	2,751	226	323	275
Open 3 (Barren)	0.80	386	631	1,632	309	505	1,305
			Total		**60,088**	**67,020**	**68,027**

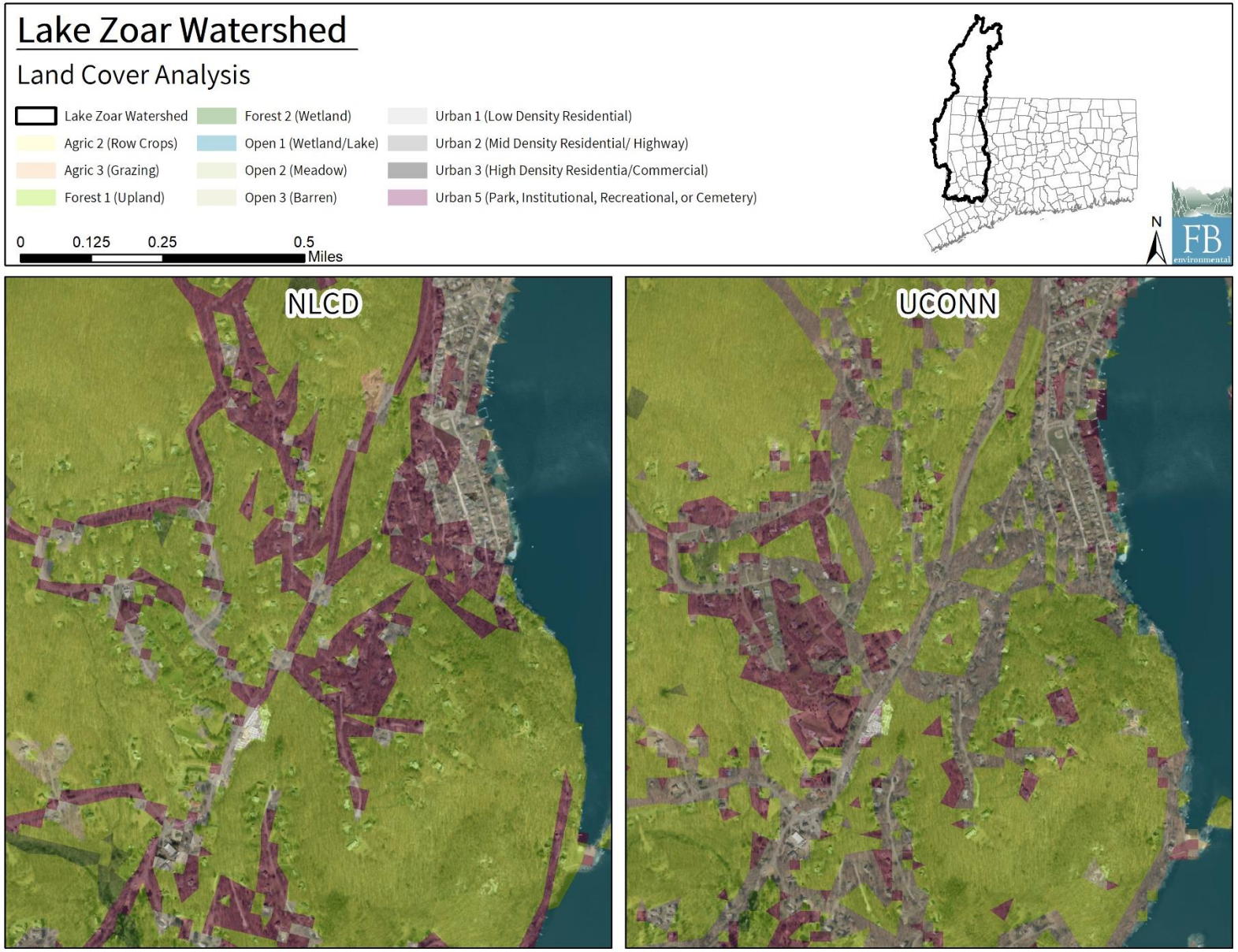


Figure A-1. Comparison of land use coverage between the 2016 National Land Cover Database (NLCD) file (left) and the 2015 UCONN CLEAR file (right).

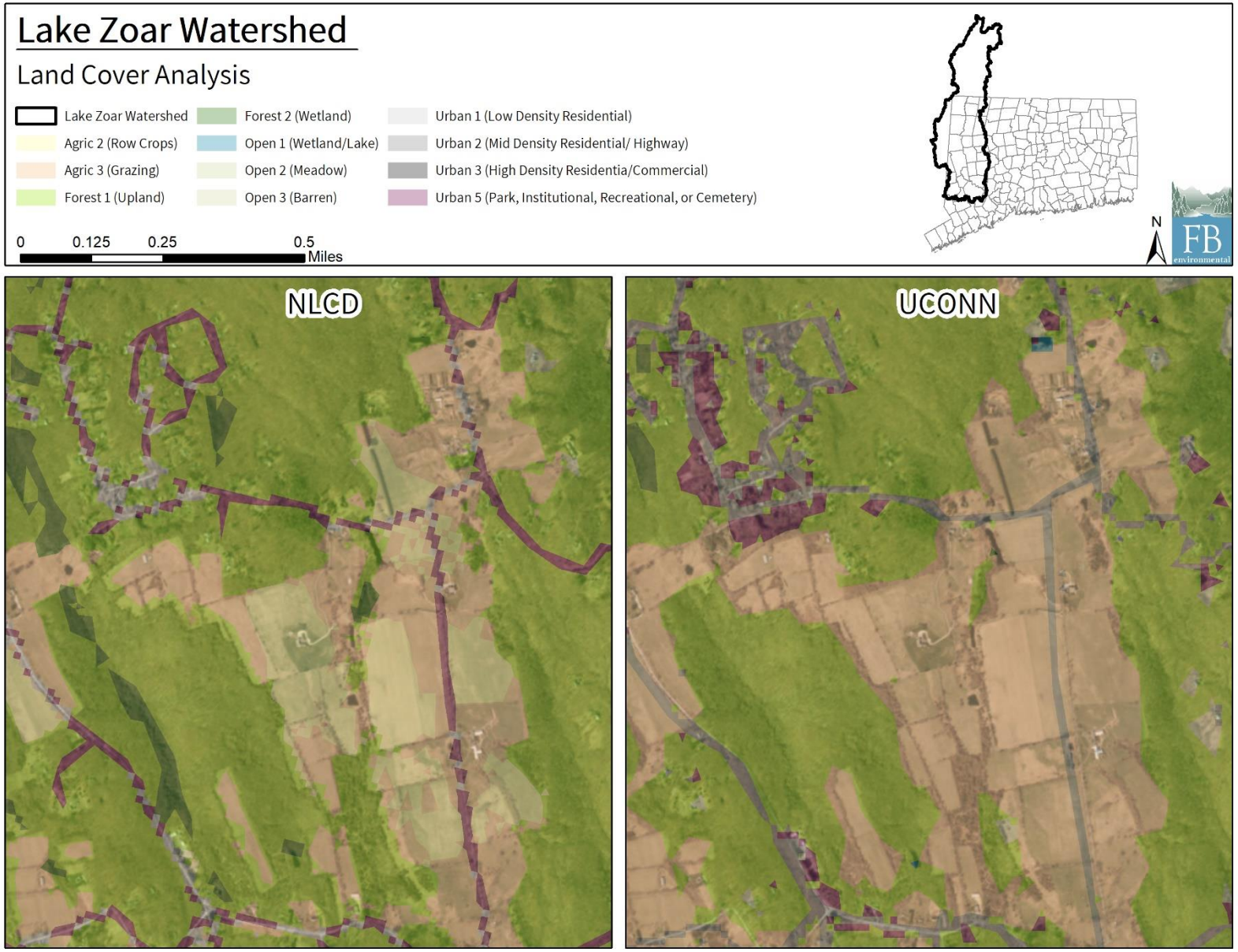


Figure A-2. Comparison of land use coverage between the 2016 National Land Cover Database (NLCD) file (left) and the 2015 UCONN CLEAR file (right).

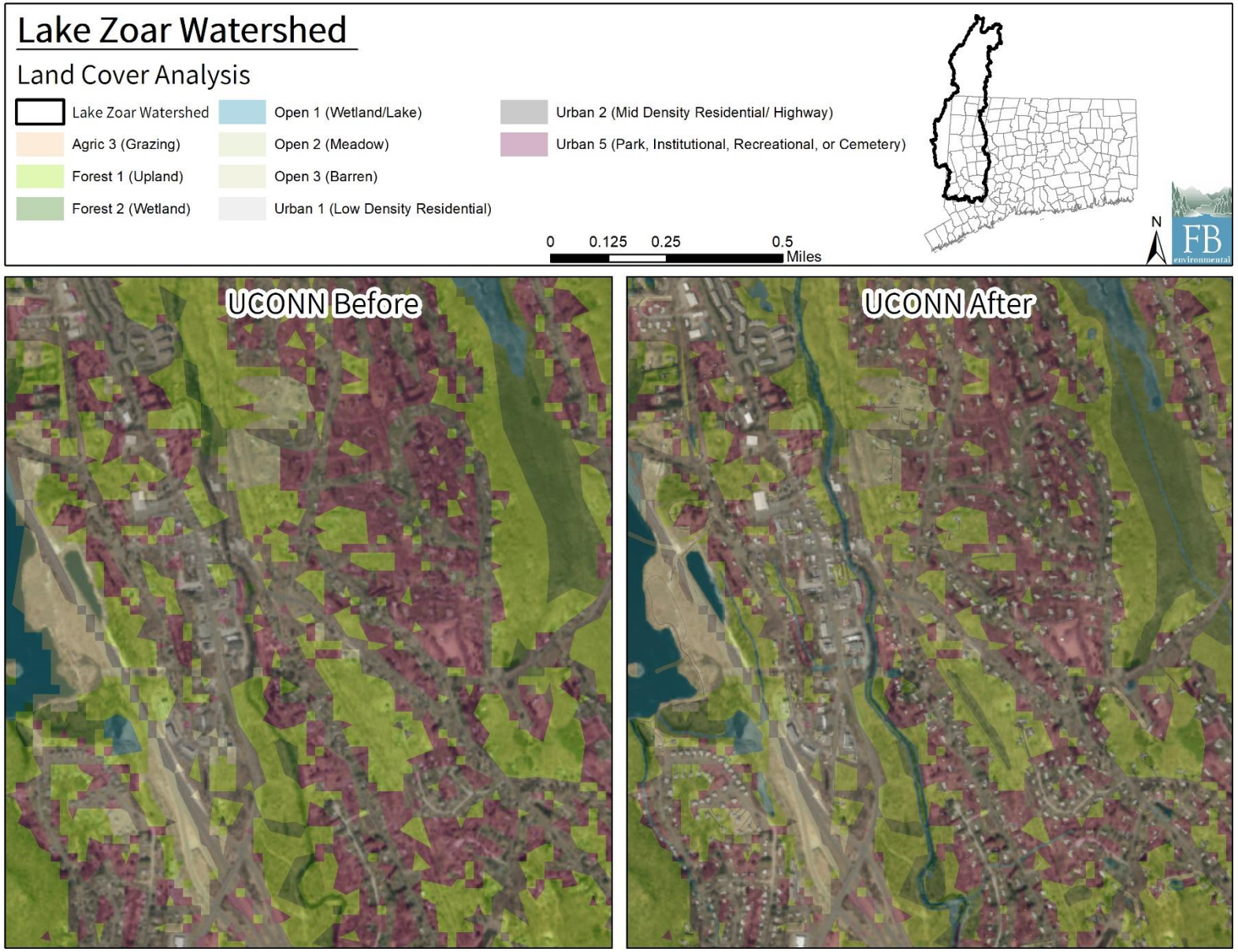


Figure A-3. Comparison of land use coverage between the 2015 UCONN CLEAR original (left) and updated (right) files.

8.0 APPENDIX B: R script for filtering eBird data

```
# extracting eBird data

# followed instructions from: https://ropensci.org/blog/2018/08/07/auk/

# development version through github better for Windows users (see below)
# for MacOS: install.packages("auk", dependencies = TRUE, repos='http://cran.rstudio.com/')

# downloaded Cygwin to C:

# Created an eBird account and requested access to data, once a download link is approved and sent, then
# a second request can be sent to filter the global database (we extracted data from CT for 2010-2019)
# A zipped folder was sent via email within a couple hours of request

### SETUP ####
setwd("~/R/projfiles/cttmdl/ebd")

install.packages("assertthat", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("countrycode", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("dplyr", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("httr", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("magrittr", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("rlang", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("stringi", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("stringr", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("tidyr", dependencies = TRUE, repos='http://cran.rstudio.com/')
library(assertthat)
library(countrycode)
library(dplyr)
library(httr)
library(magrittr)
library(rlang)
library(stringi)
library(stringr)
library(tidyr)

library(devtools)
devtools::install_github("CornellLabofOrnithology/auk")

# install packages
install.packages("raster", dependencies = TRUE, repos='http://cran.rstudio.com/')
```

```
install.packages("tidyverse", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("sf", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("rnaturalearth", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("tigris", dependencies = TRUE, repos='http://cran.rstudio.com/')
install.packages("viridisLite", dependencies = TRUE, repos='http://cran.rstudio.com/')

library(auk)
library(raster)
library(tidyverse)
library(sf)
library(rnaturalearth)
library(tigris)
library(viridisLite)

# path to ebird data
ebd_dir <- "~/R/projfiles/cttmdl/ebd"

#### CLEAN DATASET - STEP NOT NEEDED ####

# ran code but returned as "Deprecated" meaning that function is no longer required by current versions of EBD
# sampling event data not included in data download, may require special request, moved forward without it

#### IMPORT DATASET ####

# import ebd, de-duplicate group checklists (unique) and simplify taxonomy (rollup), lapsed time = 12 min
ebd <- read_ebd("ebd_US-CT_201001_201912_relFeb-2020.txt", unique = TRUE, rollup=TRUE)

# get complete list of species and subset those larger birds, import as vector file
unique <- unique(ebd$common_name)
write.table(unique, file="unique_species_ct.txt", sep=" ", row.names=FALSE)

# 393 unique bird species in the state of CT, subset to 97 species of large waterbirds, including
# geese, gulls, ducks, herons, swans, ibis, cranes, cormorants

subset = read.csv("~/R/projfiles/cttmdl/ebd/unique_species_ct_largebirds.csv", header = TRUE)
subset <- as.character(subset$SUBSET)

#### FILTER DATASET ####

# define the paths to ebd file
f_in_ebd <- file.path(ebd_dir, "ebd_US-CT_201001_201912_relFeb-2020.txt")
```

```
# create an object referencing these files
auk_ebd(file = f_in_ebd)

# set up filter query
ebd_filters <- auk_ebd(f_in_ebd) %>%
  auk_species(subset) %>%
  auk_complete()

# run filter
f_out_ebd <- "ebd_US-CT_201001_201912_relFeb-2020_subset.txt"
ebd_filtered <- auk_filter(ebd_filters, file = f_out_ebd)

# import filtered file, time lapse = 2 min
ebd_subset <- read_ebd("ebd_US-CT_201001_201912_relFeb-2020_subset.txt")

# select relevant columns, subset for counties of interest = Fairfield, New Haven, Litchfield
ebd_subset_co <- ebd_subset[,c("common_name", "observation_count", "county", "locality_id", "latitude",
                               "longitude", "observation_date")]
ebd_subset_co <- ebd_subset_co[ which(ebd_subset_co$county=="Fairfield" |
                                      ebd_subset_co$county=="New Haven" |
                                      ebd_subset_co$county=="Litchfield"), ]

write.table(ebd_subset_co, file="ebd_US-CT_201001_201912_relFeb-2020_subset2.txt", sep=" ", row.names=FALSE)

# get complete list of lat/long and subset sites in ArcMap, import as vector file
library(data.table)
coord <- setDT(ebd_subset_co)[, .(LAT = mean(latitude), LONG = mean(longitude)), by = .(LOCALID = locality_id)]
write.table(coord, file="unique_locations.txt", sep=" ", row.names=FALSE)

# identified 121 unique stations on Lake Zoar and Lake Lillinonah in ArcMap, reloaded here
aoi_sub = read.csv("~/R/projfiles/cttmdl/ebd/unique_locations_aoi.csv", header = TRUE)
aoi_sub$locality_id <- aoi_sub$LOCALID
aoi_sub$LOCALID <- NULL
aoi_sub$LAT <- NULL
aoi_sub$LONG <- NULL

# subset datatable for selected stations
ebd_subset_co_aoi <- merge(ebd_subset_co, aoi_sub, by="locality_id", all=FALSE)
length(unique(ebd_subset_co_aoi$locality_id)) # check that matches 121 stations - yes

# write to csv and perform analysis in Excel
write.csv(ebd_subset_co_aoi, file = "filtered_ebd_for_analysis.csv")
```